Gradable Team Assignments in Large Scale Learning Environments

Collaborative Learning, Teamwork, and Peer Assessment in MOOCs

Thomas Staubitz

A Dissertation Presented in Partial Fulfillment
of the Requirement for the Degree Doctor Rerum Naturalium

Submitted to the
Digital Engineering Faculty
of the University of Potsdam, Germany

Submitted: September 13, 2019
Disputation: June 10, 2020

Supervisor:
Prof. Dr. Christoph Meinel (Hasso Plattner Institute, Potsdam)

Assessors:
Prof. Dr. Carlos Delgado Kloos (Universidad Carlos III de Madrid)
Priv.-Doz. Dr. Martin Ebner (TU Graz)

Date of Disputation: 10 June 2020

Published online in the
Institutional Repository of the University of Potsdam:
https://doi.org/10.25932/publishup-47183
https://nbn-resolving.org/urn:nbn:de:kobv:517-opus4-471830

Abstract

Lifelong learning plays an increasingly important role in many societies. Technology is changing faster than ever and what has been important to learn today, may be obsolete tomorrow. The role of informal programs is becoming increasingly important. Particularly, Massive Open Online Courses have become popular among learners and instructors. In 2008, a group of Canadian education enthusiasts started the first Massive Open Online Courses or MOOCs to prove their cognitive theory of *Connectivism*. Around 2012, a variety of American start-ups redefined the concept of MOOCs. Instead of following the connectivist doctrine they returned to a more traditional approach. They focussed on video lecturing and combined this with a course forum that allowed the participants to discuss with each other and the teaching team. While this new version of the concept was enormously successful in terms of massiveness—hundreds of thousands of participants from all over the world joined the first of these courses—many educators criticized the re-lapse to the cognitivist model. In the early days, the evolving platforms often did not have more features than a video player, simple multiple-choice quizzes, and the course forum. It soon became a major interest of research to allow the scaling of more modern approaches of learning and teaching for the massiveness of these courses. Hands-on exercises, alternative forms of assessment, collaboration, and teamwork are some of the topics on the agenda. The insights provided by cognitive and pedagogical theories, however, do not necessarily always run in sync with the needs and the preferences of the majority of participants. While the former promote action-learning, hands-on-learning, competence-based-learning, project-based-learning, team-based-learning as the holy grail, many of the latter often rather prefer a more laid-back style of learning, sometimes referred to as *edutainment*. Obviously, given the large numbers of participants in these courses, there is not just one type of learners. Participants are not a homogeneous mass but a potpourri of individuals with a wildly heterogeneous mix of backgrounds, previous knowledge, familial and professional circumstances, countries of origin, gender, age, and so on. For the majority of participants, a full-time job and/or a family often just does not leave enough room for more time intensive tasks, such as practical exercises or teamwork. Others, however, particularly enjoy these hands-on or collaborative aspects of MOOCs. Furthermore, many subjects particularly require these possibilities and simply cannot be taught or learned in courses that lack collaborative or hands-on features. In this context, the thesis discusses how team assignments have been implemented on the HPI MOOC platform. During the recent years, several experiments have been conducted and a great amount of experience has been gained by employing team assignments in courses in areas, such as Object-Oriented Programming, Design Thinking, and Business Innovation on various instances of this platform: openHPI, openSAP, and mooc.house.

This thesis is dedicated to
Charlotte and Niki Floyd Staubitz

List of Papers

The following is a list of the author's publications in reverse chronological order.

2019—First Author

PAPER I: **Graded Team Assignments in MOOCs - Effects of Team Composition and Further Factors on Team Dropout Rates and Performance**
Thomas Staubitz, Christoph Meinel
In Proceedings of the Sixth Annual ACM Conference on Learning at Scale (L@S) (ACM), (2019), Chicago, IL, USA.
DOI: 10.1145/3330430.3333619

PAPER II: **MOOCs in Secondary Education - Experiments and Observations from German Classrooms**
Thomas Staubitz, Ralf Teusner, Christoph Meinel
In Proceedings of IEEE Global Engineering Education Conference (EDUCON)(IEEE), 173-182 (2019), Dubai, UAE.
DOI: 10.1109/EDUCON.2019.8725138

2019—Editor

PAPER III: **EMOOCs 2019 Work in Progress Papers of Research, Experience and Business Tracks**
Eds.: Mauro Calise, Carlos Delgado Kloos, Catherine Mongenet, Justin Reich, José A. Ruipérez-Valiente, Gideon Shimshon, Thomas Staubitz, Martin Wirsing
(EMOOCs)(CEUR-WS), (2019), Naples, Italy.
URL: http://ceur-ws.org/Vol-2356/

2019—Second/Third Author

PAPER IV: **Performance of Men and Women in Graded Team Assignments in MOOCs**
Catrina Tamara Grella, Thomas Staubitz, Christoph Meinel
IEEE Learning with MOOCs Conference (LWMOOCs)(IEEE), (2019).
Accepted for publication.

PAPER V: **Improving learner experience and participation in MOOCs: A Design Thinking approach**
Traifeh Hanadi, Thomas Staubitz, Christoph Meinel
IEEE Learning with MOOCs Conference (LWMOOCs)(IEEE), (2019).
Accepted for publication.

PAPER VI: **Took a MOOC. Got a Certificate. What now?**
Catrina Tamara Grella, Thomas Staubitz, Christoph Meinel
IEEE Frontiers in Education Conference (FIE)(IEEE), (2019).
Accepted for publication.

PAPER VII: **Skill Confidence Ratings in a MOOC: Examining the Link between Skill Confidence and Learner Development**
Karen von Schmieden, Thomas Staubitz, Lena Mayer, Christoph Meinel
Proceedings of the 11th International Conference on Computer Supported Education - Volume 1: CSEDU (CSEDU2019) (SciTePress), 533-540 (2019), Heraklion, Greece.
DOI: 10.5220/0007655405330540

PAPER VIII: **MOOCs for business use: Six hands-on recommendations**
Christian Friedl, Agnieszka Żur, Thomas Staubitz
In the 2019 OpenupEd trend report on MOOCs. (OpenupEd)(EADTU), 10–13 (2019).
URL: https://tinyurl.com/2019OpenupEdtrendreport

2018—First Author

PAPER IX: **Team-Based Assignments in MOOCs - User Feedback**
Thomas Staubitz, Hanadi Traifeh, Christoph Meinel

In Proceedings of the 2018 IEEE Learning With MOOCs Conference (LWMOOCS) (IEEE), 39-42 (2018), Madrid, Spain.
DOI: 10.1109/LWMOOCS.2018.8534607

PAPER X: **Collaborative Learning in MOOCs - Approaches and Experiments**
Thomas Staubitz, Christoph Meinel
In Proceedings of the 48th IEEE Frontiers in Education Conference (FIE) (IEEE), 1–9 (2018), San Jose, CA, USA.
DOI: 10.1109/FIE.2018.8659340

PAPER XI: **Team Based Assignments in MOOCs: Results and Observations**
Thomas Staubitz, Christoph Meinel
In Proceedings of the Fifth Annual ACM Conference on Learning at Scale (L@S) (ACM), 47:1–47:4 (2018), London, UK.
DOI: 10.1145/3231644.3231705

2018—Second/Third Author

PAPER XII: **Effects of automated interventions in programming assignments: evidence from a field experiment**
Ralf Teusner, Thomas Hille, Thomas Staubitz
In Proceedings of the Fifth Annual ACM Conference on Learning at Scale (L@S) (ACM), 60:1–60:10 (2018), London, UK.
DOI: 10.1145/3231644.3231650

PAPER XIII: **What Stays in Mind? - Retention Rates in Programming MOOCs**
Ralf Teusner, Christoph Matthies, Thomas Staubitz
In Proceedings of the 48th Frontiers in Education Conference (FIE) (IEEE), 1–9 (2018), San Jose, CA, USA.
DOI: 10.1109/FIE.2018.8658890

PAPER XIV: **Flexible, self-directed and bottom-up: Are employees overtaking their Human Resource departments with MOOCs?**
Christian Friedl, Thomas Staubitz, Darco Jansen
In Proceedings of the 2018 IEEE Learning With MOOCS Conference (LWMOOCS) (IEEE), 66–69 (2018), Madrid, Spain.
DOI: 10.1109/LWMOOCS.2018.8534616

PAPER XV: **Coorporate MOOC Trends**
Christian Friedl, Thomas Staubitz
The 2018 OpenupEd Trend Report on MOOCs (OpenupEd)(EADTU), 35–38 (2018).
URL: https://tinyurl.com/2018OpenupEdtrendreport

PAPER XVI: **Digitale Lern-Communities**
Christoph Meinel, Stefanie Schweiger, Thomas Staubitz
Handbuch E-Learning 76. Erg.-Lfg. (Deutscher Wirtschaftsdienst), 13–25 (2018).
ISBN: 978-3-87156-298-3

2017—First Author

PAPER XVII: **Towards a repository for open auto-gradable programming exercises**
Thomas Staubitz, Ralf Teusner, Christoph Meinel
In Proceedings of 6th IEEE International Conference on Teaching, Assessment, and Learning for Engineering (TALE)(IEEE), 66-73 (2017), Hong Kong, SAR, China.
(Best Paper Award)
DOI: 10.1109/TALE.2017.8252306

PAPER XVIII: **Cellular Automata as an Example for Advanced Beginners' Level Coding Exercises in a MOOC on Test Driven Development**
Thomas Staubitz, Ralf Teusner,
Nishanth Prakash, Christoph Meinel
International Journal of Engineering Pedagogy (iJEP), Vol 7, No 2, 125–141 (2017)
eISSN: 2192-4880

PAPER XIX: **openHPI's Coding Tool Family: CodeOcean, CodeHarbor, CodePilot**
Thomas Staubitz, Ralf Teusner, Christoph Meinel
In Proceedings of the Third Workshop "Automatische Bewertung von Programmieraufgaben" (ABP)(CEUR-WS) (2017), Potsdam, Germany.
Online: http://ceur-ws.org/Vol-2015/

PAPER XX: **The Gamification of a MOOC Platform**
Thomas Staubitz, Christian Willems,
Christiane Hagedorn, Christoph Meinel
In Proceedings of IEEE Global Engineering Education Conference (EDUCON)(IEEE), 883–892 (2017), Athens, Greece.
DOI: 10.1109/EDUCON.2017.7942952

PAPER XXI: **Collaboration and Teamwork on a MOOC Platform - A Toolset**
Thomas Staubitz, Christoph Meinel
In Proceedings of ACM Learning at Scale Conference (L@S)(ACM), 165–168 (2017), Cambridge, MA, USA.
DOI: 10.1145/3051457.3053975

2017—Second/Third Author

PAPER XXII: **Video Conferencing as a Peephole to MOOC Participants**
Ralf Teusner, Nicholas Wittstruck, Thomas Staubitz
In Proceedings of IEEE International Conference on Teaching, Assessment, and Learning for Engineering (TALE)(IEEE), 100–107(2017), Hong Kong, SAR, China.
DOI: 10.1109/TALE.2017.8252312

PAPER XXIII: **Peer Assessment**
Christoph Meinel, Stefanie Schweiger, Thomas Staubitz
Handbuch E-Learning 73. Erg.-Lfg. (Deutscher Wirtschaftsdienst), 163–180 (2017).
ISBN: 978-3-87156-298-3

2016—First Author

PAPER XXIV: **Cellular Automata as Basis for Programming Exercises in a MOOC on Test-driven Development**
Thomas Staubitz, Ralf Teusner,
Nishanth Prakash, Christoph Meinel
In Proceedings of IEEE International Conference on Teaching, Assessment, and Learning for Engineering (TALE2016)(IEEE), 374–380 (2016), Bangkok, Thailand.
(Best Paper Award)
DOI: 10.1109/TALE.2016.7851824

PAPER XXV: **Automatisierte Online-Aufsicht im Kontext der Wertigkeit von Zertifikaten einer MOOC Plattform**
Thomas Staubitz, Ralf Teusner, Jan Renz, and Christoph Meinel
Lecture Notes in Informatics (LNI) - Proceedings Series of the Gesellschaft für Informatik (GI), (DeLFI)(Springer), 125–136 (2016), Potsdam, Germany.
ISBN: 978-3-88579-656-5

PAPER XXVI: **Vagrant Virtual Machines for Hands-On Exercises in Massive Open Online Courses**
Thomas Staubitz, Maximilian Brehm, Johannes Jasper,
Thomas Werkmeister, Ralf Teusner, Christian Willems,
Jan Renz, and Christoph Meinel
In Proceedings of International KES Conference on SMART EDUCATION AND E-LEARNING (KES-SEEL)(Springer), 363–373 (2016), Puerto de la Cruz, Spain.
DOI: 10.1007/978-3-319-39690-3_32

PAPER XXVII: **Improving the Peer Assessment Experience on MOOC Platforms**
Thomas Staubitz, Dominic Petrick, Matthias Bauer,
Jan Renz and Christoph Meinel
In Proceedings of ACM Learning at Scale Conference (L@S)(ACM), 389–398 (2016), Edinburgh, UK.
DOI: 10.1145/2876034.2876043

PAPER XXVIII: **First Steps in Automated Proctoring**
Thomas Staubitz, Ralf Teusner, Jan Renz, Christoph Meinel
In Proceedings of the Fourth MOOC European Stakeholders Summit (EMOOCs)(P.A.U), 41–53 (2016), Graz, Austria.
(Outstanding paper and video award)
ISBN: 978-3-7412-4593-0

PAPER XXIX: **CodeOcean - A Versatile Platform for Practical Programming Exercises in Online Environments**
Thomas Staubitz, Hauke Klement, Ralf Teusner, Jan Renz, Christoph Meinel
In Proceedings of IEEE Global Engineering Education Conference (EDUCON)(IEEE), 314–323 (2016), Abu Dhabi, UAE.
DOI: 10.1109/EDUCON.2016.7474573

2016—Second/Third Author

PAPER XXX: **Enabling Schema Agnostic Learning Analytics in a Service-Oriented MOOC Platform**
Jan Renz, Gerardo Navarro-Suarez, Rowshan Sathi, Thomas Staubitz, Christoph Meinel
In Proceedings of the Third ACM Conference on Learning at Scale (L@S)(ACM), 137–140 (2016), Edinburgh, UK.
DOI: 10.1145/2876034.2893389

PAPER XXXI: **Using A/B Testing in MOOC Environments**
Jan Renz, Daniel Hoffmann, Thomas Staubitz, Christoph Meinel
Proceedings of the 6th International Conference on Learning Analytics and Knowledge (LAK)(SOLAR), 304–313 (2016), Edinburgh, UK.
DOI: 10.1145/2883851.2883876

PAPER XXXII: **Pre-Course Key Segment Analysis of Online Lecture Videos**
Xiaoyin Che, Thomas Staubitz, Haojin Yang, Christoph Meinel
In Proceedings of 16th IEEE International Conference on Advancing Learning Technologies (ICALT)(IEEE), 416–420 (2016), Kraków, Poland.
DOI: 10.1109/ICALT.2016.102

PAPER XXXIII: **Can MOOCs Support Secondary Education in Computer Science?**
Catrina Tamara Grella, Thomas Staubitz, Ralf Teusner, Christoph Meinel
In Proceedings of the 19th International Conference on Interactive Collaborative Learning (ICL)(IAOE), 478–493 (2016), Belfast, UK.
DOI: 10.1007/978-3-319-50337-0_45

PAPER XXXIV: **Improving Collaborative Learning With Video Lectures**
Matthias Bauer, Martin Malchow, Thomas Staubitz, Christoph Meinel
In INTED2016 Proceedings (INTED)(IATED), 5511–5517 (2016), Valencia, Spain.
DOI: 10.21125/inted.2016.0322

2015—First Author

PAPER XXXV: **Towards Practical Programming Exercises and Automated Assessment in Massive Open Online Courses**
Thomas Staubitz, Hauke Klement, Jan Renz, Ralf Teusner, Christoph Meinel
In Proceedings of the IEEE International Conference on Teaching, Assessment, and Learning for Engineering (TALE)(IEEE), 23–30 (2015), Zhuhei, China.
DOI: 10.1109/TALE.2015.7386010

PAPER XXXVI: **Collaborative Learning in a MOOC Environment**
Thomas Staubitz, Tobias Pfeiffer, Jan Renz, Christian Willems, Christoph Meinel
In Proceedings of the 8th International Conference of Education, Research and Innovation (ICERI)(IATED), 8237–8246 (2015), Sevilla, Spain.
ISBN: 978-84-608-2657-6

2015—Second/Third Author

PAPER XXXVII: **Scaling Youth Development Training in IT Using an xMOOC Platform**
Martin v. Löwis, Thomas Staubitz, Ralf Teusner, Jan Renz, Susanne Tannert, Christoph Meinel
In Proceedings of the IEEE Frontiers in Education (FiE2015)(IEEE), 1–9 (2015), El Paso, TX, USA.
DOI: 10.1109/FIE.2015.7344145

PAPER XXXVIII: **Enhancing Content between Iterations of a MOOC - Effects on Key Metrics**
Ralf Teusner, Keven Richly, Thomas Staubitz, Jan Renz
In Proceedings of the Third European MOOC Stakeholders Summit (EMOOCs)(P.A.U), (2015), Mons, Belgium.

PAPER XXXIX: **Optimizing the Video Experience in MOOCs**
Jan Renz, Matthias Bauer, Martin Malchow, Thomas Staubitz, Christoph Meinel
In Proceedings of the 7th International Conference on Education and New Learning Technologies (EduLearn)(IATED), 5150–5158 (2015), Barcelona, Spain.
ISBN: 978-84-606-8243-1

2014—First Author

PAPER XL: **Towards Social Gamification - Implementing a Social Graph in an xMOOC Platform**
Thomas Staubitz, Sebastian Woinar, Jan Renz, Christoph Meinel
In Proceedings of the 7th International Conference of Education, Research and Innovation (ICERI)(IATED), 2045–2054 (2014), Sevilla, Spain.
ISBN: 978-84-617-2484-0

PAPER XLI: **Supporting Social Interaction and Collaboration on an xMOOC Platform**
Thomas Staubitz, Jan Renz, Christian Willems, Christoph Meinel
In Proceedings of the 6th International Conference on Education and New Learning Technologies (EduLearn)(IATED), 6667–6677 (2014), Barcelona, Spain.
ISBN: 978-84-617-0557-3

PAPER XLII: **Lightweight Ad Hoc Assessment of Practical Programming Skills at Scale**
Thomas Staubitz, Jan Renz, Christian Willems,
Johannes Jasper, Christoph Meinel
In Proceedings of IEEE Global Engineering Education Conference (EDUCON)(IEEE), 475–483 (2014), Istanbul, Turkey.
DOI: 10.1109/EDUCON.2014.6826135

2014—Second/Third Author

PAPER XLIII: **Reflections on Enrollment Numbers and Success Rates at the openHPI MOOC Platform**
Christian Willems, Jan Renz, Thomas Staubitz, Christoph Meinel
In Proceedings of the Second MOOC European Stakeholders Summit (EMOOCs)(P.A.U), (2014), Lausanne, Switzerland.
ISBN: 978-84-8294-689-4

PAPER XLIV: **Handling Re-grading of Automatically Graded Assignments in MOOCs**
Jan Renz, Thomas Staubitz, Christian Willems, Hauke Klement, Christoph Meinel
In Proceedings of IEEE Global Engineering Education Conference (EDUCON)(IEEE), 408–415 (2014), Istanbul, Turkey.
DOI: 10.1109/EDUCON.2014.6826127

PAPER XLV: **Improving the On-boarding User Experience in MOOCs**
Jan Renz, Thomas Staubitz, Jaqueline Pollack, Christoph Meinel
In Proceedings of the 6th International Conference on Education and New Learning Technologies (EduLearn)(IATED), 3931–3941 (2014).
ISBN: 978-84-617-0557-3

PAPER XLVI: **MOOC to Go**
Jan Renz, Thomas Staubitz, Christoph Meinel
In Proceedings of the 10th International Conference on Mobile Learning (ML)(IADIS), 125–132 (2014), Madrid, Spain.
ISBN: 978-989-8704-02-3

REUSE (IEEE)
Substantial parts of the following papers, originally published with IEEE, have been reused in this thesis:

Team-Based Assignments in MOOCs - User Feedback
Thomas Staubitz, Hanadi Traifeh, Christoph Meinel
In Proceedings of the 2018 IEEE Learning With MOOCs Conference (LW-MOOCS) (IEEE), 39-42 (2018), Madrid, Spain.
DOI: 10.1109/LWMOOCS.2018.8534607

The Gamification of a MOOC Platform
Thomas Staubitz, Christian Willems,
Christiane Hagedorn, Christoph Meinel
In Proceedings of IEEE Global Engineering Education Conference (EDUCON)(IEEE), 883–892 (2017), Athens, Greece.
DOI: 10.1109/EDUCON.2017.7942952

REUSE (ACM)

Substantial parts of the following papers, originally published with ACM, have been reused in this thesis:

Graded Team Assignments in MOOCs - Effects of Team Composition and Further Factors on Team Dropout Rates and Performance
Thomas Staubitz, Christoph Meinel
In Proceedings of the Sixth Annual ACM Conference on Learning at Scale (L@S) (ACM), Article No. 5 (2019), Chicago, IL, USA.
DOI: 10.1145/3330430.3333619

Team Based Assignments in MOOCs: Results and Observations
Thomas Staubitz, Christoph Meinel
In Proceedings of the Fifth Annual ACM Conference on Learning at Scale (L@S) (ACM), 47:1–47:4 (2018), London, UK.
DOI: 10.1145/3231644.3231705

Collaboration and Teamwork on a MOOC Platform - A Toolset
Thomas Staubitz, Christoph Meinel
In Proceedings of ACM Learning at Scale Conference (L@S)(ACM), 165–168 (2017), Cambridge, MA, USA.
DOI: 10.1145/3051457.3053975

Improving the Peer Assessment Experience on MOOC Platforms
Thomas Staubitz, Dominic Petrick, Matthias Bauer,
Jan Renz and Christoph Meinel
In Proceedings of ACM Learning at Scale Conference (L@S)(ACM), 389–398 (2016), Edinburgh, UK.
DOI: 10.1145/2876034.2876043

Author's Contribution

The main contribution of this thesis is an in-depth examination of teamwork and graded team assignments in MOOCs. The thesis introduces the elements that constitute the basis for team assignments on the HPI platforms:

- TeamBuilder—A tool to match teams based on configurable criteria.
- Collab Spaces—A toolbox that allows these teams to collaborate virtually.
- Peer Assessment—A tool that allows the instructors to outsource the grading of complex tasks to the participants.

The tools originate from master theses that have been co-supervised by the author. In close collaboration with the major stakeholders, the author transformed these prototypes into tools for productive use. During this process, the author experimented with the features and tools in field tests on the openHPI, openSAP, and mooc.house platforms. For this purpose, he co-authored and co-facilitated a series of courses in which the features have been employed.

- On the openHPI platform: *javaeinstieg2015*, *javawork2015*, *javawork2016*, *javaeinstieg2017*, *javawork2017*, *javaeinstieg-mint-ec-2018*, *java-capstone-1*, and *international-teams2019*[1].
- On the mooc.house platform: *bizmooc2018*, *javaeinstieg-schule2019*, and *corship2020*[2].
- On the openSAP platform: *java1*.

Furthermore, the author has been involved in the teaching teams or support teams for several other courses on the openHPI platform, in which some of the features have been employed, such as *webtech2015*, *webtech2017*, and *ws-privacy2016*.

The author designed and conducted surveys and interviews with the participants of these courses and observed the way the participants worked with the features to analyze their needs and requirements and to feed them back into the upcoming development cycles. In addition, team assignments have

1. In preparation at the time of writing.
2. In preparation at the time of writing.

played a relevant role in several other courses on the openSAP platform. In total, fifteen courses offering graded team assignments have been conducted on the HPI MOOC platforms from 2016 to 2019. For these fifteen courses, the author created two datasets, merging the data from all of the platform's relevant sources. One for the team members, the second for those participants that have not participated in the team assignments. Furthermore, the author has aggregated the first data set on the team level. All datasets are pseudonymized. In conjunction with the surveys and the interviews, these datasets define the main pillars of the thesis.

Contents

Abbreviations

cMOOC	Massive Open Online Course following the connectivist model
CoP	Confirmation of Participation. openHPI certificate, issued when a participant has visited at least 50% of the course material.
CSV	Comma separated values. Plain-text based file format for structured spreadsheet data
HCD	Human centered design
HPI	Hasso Plattner Institute
IDE	Integrated Development Environment
KPI	Key performance indicator
LGC	Large Group Collaboration
LMS	Learning Management System
LTI	Learning Tools Interoperability interfaces. Enables to connect external learning tools to LMS or MMS that support the interface. https://www.imsglobal.org/activity/learning-tools-interoperability
LXD	Learning experience design
MMS	MOOC Management System
MOOC	Massive Open Online Course
NGO	Non-Government organization
OOP	Object Oriented Programming
PA	Peer assessment
QC	Qualified Certificate
RoA	Record of Achievement.openHPI certificate, issued when a participant has received at least 50% of the available points in a course.
SAP	German software company
SGC	Small Group Collaboration
SME	Small and Medium Enterprises
UI	User interface
UNED	Universidad Nacional de Educación a Distancia, Spanish distance learning university
UX	User experience
WHO	World Health Organization
xMOOC	Massive Open Online Course following the Stanford/Harvard/MIT model
XP	eXperience Points. Part of openHPI's gamification features

List of Figures

List of Tables

1. Introduction and Preliminaries

This chapter will start with outlining the big picture. The context, background, topic, and the actors involved in the research will be introduced. To define the context, a brief overview on the history and development of MOOCs, MOOC platforms, MOOC providers, etc. will be given. Next, the openHPI platform family will be introduced to provide the direct background for this work. Then, the options to assess the performance of the participants on the openHPI MOOC platform, with a particular focus on collaborative learning, project-based learning, peer assessment and teamwork will be outlined, to define the theme of the thesis. Finally, the involved actors, particularly, instructors and participants, their needs, roles, and the type of their relationships to the platforms will be examined.

1.1 Massive Open Online Courses (MOOCs)

Lifelong learning plays an increasingly important role in most societies. Technology is changing faster than ever and what has been important to study today may be obsolete tomorrow. The role of informal programs has witnessed an augmenting relevance. Particularly, Massive Open Online Courses (MOOCs) have become very popular among learners and instructors. The mere concept of MOOCs dates from the 1960s, when Buckminster Fuller [1] delineated his vision of the classroom of the future in his book *On Education* (see Figure 1.1 for an overview of the whole story). He envisioned that the world's top adept in each topic could record their lectures on video and thus would be freed of further (aggravating) lecturing and teaching chores, enabling them to focus on their research. Students would gather in masses under geodetic domes and interact with the instructors by "two-way television." With a few adjustments in the details, this futuristic vision turned into reality four decades later.

In 2008, Stephen Downes, George Siemens, and Dave Cormier, a group of Canadian education enthusiasts started the first Massive Open Online Courses or MOOCs to test and implement their cognitive conjecture, which they coined *Connectivism*. Generally, these so-called cMOOCs (or connectivistMOOCs) use existing web tools, such as blogs, discussion forums, wikis, or video hosts to concatenate existing or newly created knowledge artifacts. Their intention is

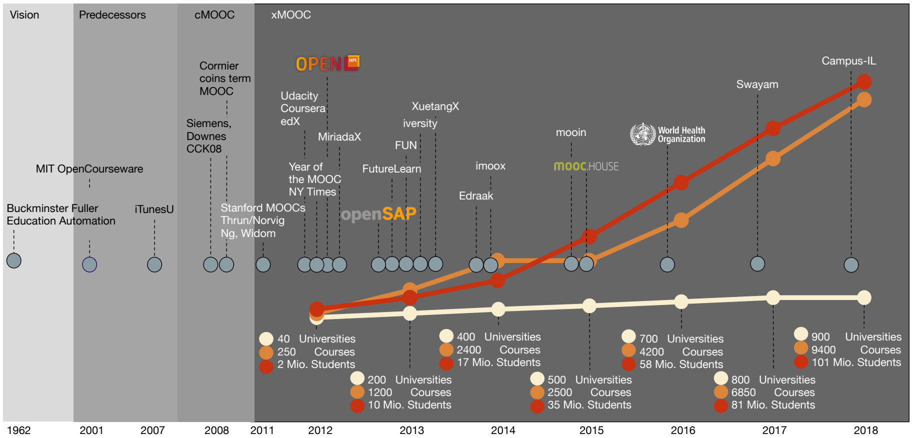

Figure 1.1: The history of MOOCs. Vision, predecessors, platforms. Note the different scales for universities (hundreds), courses (thousands), and participants (millions). The numbers have been retrieved from Class Central reports. The graph is following their wording (students) while we prefer the term participants.

less delivering learning material to the participants, they rather are facilitating the creation and collection of content. Downes [2], characterizes these MOOCs as based on the four principles of:

- Autonomy of the learner: there is no curriculum, the learner decides about the choice of contents,
- Openness: all contents are open accessible and free to be used and reused,
- Diversity: not one platform but multiple tools are employed, learners have individual perspectives and decide which path to follow,
- Interactivity: communication and cooperative learning are encouraged, knowledge emerges as the result of the interaction between learners.

Soon afterwards, a variety of American start-ups, mostly rooted in the context of major American universities[1] started to redefine the concept of these Massive Open Online Courses. Instead of following the connectivist doctrine they went back to a more traditional cognitivist approach with a focus on frontal lecturing by means of video. In 2011, three years after Siemens, Downes, and Cormier *invented* the cMOOC, Sebastian Thrun and Peter Norvig hosted their first MOOC on the topic of Artificial Intelligence (AI) [3] This was the beginning of the so called Harvard style extended MOOCs—or in short xMOOCs.

1. particularly Stanford, Harvard, and the MIT

Whereas cMOOCs follow a distributed approach, relying on existing web tools, xMOOCs, generally, are following a more centralized approach by employing specialized platforms to publish their content. These systems will be called MOOC Management Systems (MMS) from here on. In contrast to traditional Learning Management Systems (LMS), MMSs are optimized to deal with very large (or massive) numbers of participants in a single course. This is manifested in terms of performance as well as in different user interface design paradigms.

Compared to cMOOCs, the didactical model of most xMOOCs is more traditional. Videos combined with additional reading material and automated self-tests often provide a frontal teaching experience, though the details might differ from platform to platform, as well as within courses on the same platform. In the worst case, this might lead to a behaviorist teaching machine that prohibits a participant to proceed to the next learning object before she has reached a certain amount of points in a quiz. In the best case, the instructors deliver interesting courses that encourage social interaction among the participants, facilitate teamwork and offer plenty practical exercises. In which way a course is delivered, is often rather a question of the instructors' aims and creativity, than of particular possibilities of a platform.

xMOOCs, generally, attract a larger audience than cMOOCs as they are more easily consumable, and sometimes even tend towards edutainment [4]. In 2012, the New York Times declared the "Year of the MOOC" [5]. Owed to their almost immediate success, MOOCs have been hyped by the press as the new *wondertool* for online learning and tele-teaching. Following this hype, the label MOOC has been pinned on almost everything resembling a course offered online.

While the hype about MOOCs already peaked in 2013/14 and the phenomenon has often been declared dead afterwards[1], the steadily growing number of MOOCs, MOOC platforms and MOOC participants draws a somehow different picture [6]. By now, as the dust has settled, a more conservative pattern is emerging: MOOCs are a paramountly effective and valuable tool in the context of lifelong learning [7]. For universities, MOOCs serve as a vehicle

1. Gartner's hype cycle delineates the phenomenon quite well. During the recent years, MOOCs have been—and still are—placed close to everywhere on this curve. Gartner's own hype cycle for education, quite interestingly, still categorized MOOC platforms under "On the rise" in 2016. In 2017 and 2018, however, they have completely disappeared from their radar. Comment by the author: From my personal judgement, I would have located them at the peak around 2013/14 and now on their way from the slope of enlightenment to the plateau of productivity. https://www.gartner.com/en/documents/3364119/hype-cycle-for-education-2016,
https://www.gartner.com/en/documents/3769145/hype-cycle-for-education-2017,
https://www.gartner.com/en/documents/3882872/hype-cycle-for-education-2018.
All links, last access July 2019

to market their programs and thus to attract students. In addition, they are a tool to automate unpopular tasks, such as introductory courses for freshmen [8]. Most recently, there is also a trend back towards a more formal approach of certification with MOOCs. Examples are Georgia Tech's Computer Science Online Master[1] or the MIT's Supply Chain Management Micro-Masters[2]

Today, MOOCs are a global phenomenon. Next to universities—enterprises, SMEs, and NGOs are more and more discovering the format as a means to inform their employees, partners, and customers about new technologies or train them in the use of their products [9].

In the US, MOOC platforms are mainly operated by private companies, such as Coursera or Udacity, or non-profit organizations, such as EdX.

In Europe, the topic is handled differently from country to country. For example in France, a national platform is provided by the government: France Université Numérique (FUN)[3], while OpenClassrooms[4] is their privately operated competitor. In Spain, the major universities have teamed up with a private technology provider[5] and successfully serve an audience not only in Spain but also in the huge Latin American markets. Another big player in the Spanish MOOC market is UNED[6] with its offer UnedAbierta[7]. In Britain, the OpenUniversity joined forces with the BBC and the British Council to operate FutureLearn[8]. In Germany, so far, there is no central MOOC platform, a discussion about the feasibility of such a platform has only been started recently by the *Hochschulforum Digitalisierung*[9] [10]. Iversity, the largest commercial German MOOC platform has failed to become a success of its own and has been bought and sold by several media corporations, most recently[11] by Springer Nature, to deliver content that is digitally enhancing their print materials. The strongest protagonist in Germany is the openHPI platform family, which is operated by the Hasso Plattner Institute in Potsdam. These platforms will be introduced in more detail in the following section. Another player in the Ger-

1. http://www.omscs.gatech.edu/
2. https://scm.mit.edu/micromasters
3. https://www.fun-mooc.fr/, last access March 23, 2019
4. https://openclassrooms.com/en/, last access March 23, 2019
5. https://miriadax.net/cursos, last access March 23, 2019
6. Universidad Nacional de Educación a Distancia, last access March 23, 2019
7. http://catalogo.unedabierta.uned.es/, last access March 23, 2019
8. https://www.futurelearn.com/, last access March 23, 2019
9. https://hochschulforumdigitalisierung.de/, last access March 23, 2019
10. Update July, 2019: Meanwhile, the final report of the feasibility study has been published. (https://hochschulforumdigitalisierung.de/sites/default/files/dateien/HFD_AP_Nr33 _Ergebnisbericht_Machbarkeitsstudie_Hochschulplattform.pdf)
11. March 2019

man market is OnCampus[1], formerly known as MOOIN and operated by the TH Lübeck. In Austria, TU Graz and University of Graz have joined the party early on with their self-developed platform imoox[2]. In Italy, particularly the University of Naples Federico II is active in providing MOOCs on the European EMMA platform[3], which is also hosted by this university. Recently, the Italian government commissioned a national Italian MOOC platform[4] in cooperation with a network of Italian Universities.

In most other European countries MOOCs play a rather marginal role; albeit there are some universities, which are actively producing MOOCs and delivering them on the American Coursera or EdX platforms, such as the TU Munich in Germany, the TU Delft in the Netherlands, the Université Catholique Leuven in Belgium, the Universidad Carlos III de Madrid (UC3M) in Spain, or the EPFL in Lausanne in Switzerland. The European Union has adopted the topic and has been funding several initiatives for European MOOC platforms, such as EcoLearn[5] or EMMA[6], and a substantial amount of research on a variety of topics in the context of MOOCs[7].

In China, XuetangX operated by the Tsinghua university is offering courses by the top-ranking Chinese universities and some prestigious international universities, such as UC Berkeley or Rice University. Another big player in the Chinese MOOC market is icourse163MOOC[8], which is a joint effort of a state-backed educational site and a Chinese Internet company. Furthermore, ChineseMOOCs[9] is offered by Beijing University and Alibaba [10].

In the Middle East, Edraak[10]—a platform based in Jordan—and Campus Israel[11]—the national Israeli MOOC platform—are the major forces in the sector. Edraak partners with several Arab universities and institutions as well as a couple of international MOOC providers, such as MITx and Harvardx. Campus Israel is offering courses in Hebrew, Arab, and English language and also offers courses to an international audience under the brand of IsraelX on the EdX platform.

1. https://www.oncampus.de/mooin, last access March 23, 2019
2. Austrian slang for "I like it." https://imoox.at/mooc/, last access March 23, 2019
3. https://platform.europeanmoocs.eu/, last access March 23, 2019
4. https://learn.eduopen.org/, last access March 23, 2019
5. The platform is no more available, March 23, 2019
6. https://platform.europeanmoocs.eu/, last access March 23, 2019
7. To name just a few: https://bizmooc.eu/, http://tramooc.eu, https://www.openuped.eu, http://www.mooc-maker.org, https://moonliteproject.eu, all last access March 23, 2019
8. https://www.icourse163.org/
9. http://www.chinesemooc.org/
10. https://www.edraak.org/en/
11. https://campus.gov.il/

Further platforms are operated in India[1], Indonesia[2], or Russia[3]. Since EdX has open-sourced its MMS software and thus made it available for many institutions that would neither have had the resources to develop such a platform at their own expense nor to pay the fees asked from platforms, which are offered as Software-as-a-Service solutions, the list is increasingly growing. The presented list of platforms, therefore, is far from complete.

While almost all platforms first offered their content for free, nowadays, there is a strong tendency—particularly among the commercial platforms—to put certification, assessments, and even course contents behind a paywall. Additionally, some platforms offer paid tutoring models. Participants can buy personal support or even more "professional" reviews of their work than in peer assessment models.

While the xMOOC model was enormously successful in terms of massiveness, many educators castigated this re-lapse to the cognitivist (or in the worst cases behaviorist) model. In the early days, this often was a technical necessity as the evolving platforms featured little more possibilities than playing videos and some form of multiple-choice tests. Ameliorating the scalability of more modern approaches to learning and teaching, and—thus, making them available for MOOCs—soon became a major interest of research. Hands-on exercises, alternative forms of assessment, collaboration, and teamwork are some of the topics that are on the agenda. In this general context, the thesis focusses on gradable team assignments in MOOCs.

1.2 openHPI

With its openHPI platform[4], the Hasso Plattner Institute (HPI) in Potsdam, Germany, has been the first European university institute to offer interactive MOOCs. The first course on the platform was conducted in September 2012. The course topics are mostly located in the extended field of computer science and IT technology, ranging from programming courses, courses on Internet technologies and Internet security to courses about IT law or innovation techniques. The courses are offered in either German or English language.

In 2012, a proper MOOC management system (MMS) was not available on the market. Therefore, openHPI started with a heavily modified open-source

1. https://swayam.gov.in/
2. https://www.indonesiax.co.id/
3. https://www.lektorium.tv/, https://universarium.org/, https://openedu.ru/
4. http://open.hpi.de

Learning Management System (LMS)[1]. Soon it turned out that the LMS did not fit the needs of this new type of massive digital learning. We[2] decided to implement a custom, in-house solution from scratch, based on the experience that we have gained during the operation of the first few courses[3]. The key elements of the MOOC innovation for online learning were identified as:

- The synchronization of learners by providing the learning materials week by week.
- The supply of various feedback tools for self- and external evaluation of the learning success.
- The integration of social platform features, such as a course forum, to provide learners with the experience of being part of a social (albeit virtual) learning community.
- The support for the massiveness of the courses: particularly, automation and user interfaces that take this massiveness into account.

Basically, openHPI follows the Harvard/MIT style of MOOCs (xMOOCs). On top of that, however, hands-on experience and the support of collaborative learning and teamwork have been some of the focus areas of the platform development from the beginning (see [11–13]). This combination allowed for courses that support different learning styles [11] and courses that are following a social constructivist[4] or project-based[5] approach.

Next to openHPI—featuring more than 600,000 enrollments and about

1. This initial prototype was based on the Canvas LMS: https://www.canvaslms.com/
2. The text uses the "we"-form when it refers to implementation details or decisions on the platform, as all of them have been the result of discussions between many stakeholders. Furthermore, the "we"-form is used when the text refers to course design decisions. The passive form is used when the text refers to the author's research. Sometimes, the borders are blurry.
3. Several months later, when the popular edX platform open-sourced its MMS software, the development of the in-house system was already too advanced to stop the process and to switch to this now available open-source tool.
4. In short, the social constructivist learning theory is based on the assumption that people learn best while they are interacting with other human beings, creating knowledge, skills, and understanding by discussing ambiguous material or jointly solving problems. Each individual contributes their own expertise or problem solving approaches.
5. In project-based learning, students are working on a broader, often multi-disciplinary challenge. Ideally, they are working in teams, but smaller projects can also be handled individually. Instead of being lectured, the students learn by hands-on solving an (ideally real-life) challenge. Earlier acquired knowledge can be applied in context. The idea of project-based learning can be traced back to the work of John Dewey in the 1920s [14] and Jerome Bruner in the 1960s [15]. The epistemological basis for this type of learning is delineated in the psychological school of constructivism as pioneered by Jean Piaget in the 1950s and 1960s. Seymour Papert, later on, developed the constructionist learning theory based on this theoretical basis [16].

200,000 registered users[1]—the openHPI team is operating and maintaining additional instances of the platform for several customers and use cases.

openSAP was launched in early 2013. It is the world's first dedicated enterprise MOOC platform and by now features about 3,500,000 enrollments and more than 800,000 users.

Another instance of the platform—mooc.house—was announced and published in 2015. As a white label platform, it enables small and medium businesses, universities, and other organizations to provide MOOCs, it also serves as a platform to offer content that doesn't directly fit in to the openHPI profile. Currently, it features about 60,000 enrollments. Customers offering courses include the Charité Berlin (a hospital and medical research institution), Signavio (a software company), and acatech (the German academy for technical sciences).

Since 2016, OpenWHO, the World Health Organization's instance of the polatform, is providing courses on pandemic prevention and outbreak handling to first responders and other stakeholders. It recently has crossed the 100,000 enrollments milestone[2].

Next to developing, operating, and maintaining the platforms, the openHPI team is actively researching a variety of aspects of learning at scale, including automated assessment, peer assessment, teamwork and collaborative learning, online proctoring, learning analytics, mobile learning, game-based learning, and gamification.

1.3 Performance Assessment on openHPI

It is necessary to assess the participants' performance, as the platform offers certificates for successful course completion. Three levels of certification are available:

- Confirmation of Participation (CoP)—The CoP is issued to all participants who have "seen" at least a certain percentage of items. The default value here is 50%. Therefore, the most basic form of assessment is the tracking of visited items.
- Record of Achievement (RoA)—To obtain a Record of Achievement, the participants have to solve quizzes, work on practical exercises, or small

1. All listed enrollment and registration numbers in this context have been collected in March 2019
2. All listed numbers have been collected in March 2019

projects. Per default, they have to achieve at least 50%[1] of the available points.

- Qualified Certificate (QC)—The requirements are identical to the RoA. Additionally, the participants have to opt-in for a paid online-proctoring of the graded assignments[2].

The most important prerequisite for assessment tools is that they allow scalability. Owed to the massive amount of participants, manual grading by the instructors is not an option. Instead, automation and outsourcing are possible alternatives. The platform offers a quiz system that allows for automatically graded multiple-choice and multiple-answer questions and ungraded free-text questions. The platform supports the Learning Tools Interoperability (LTI) interface, which allows to simply integrate external learning tools, and to make use of their grading options. The main usage of the LTI interface on openHPI is to connect the platform to CodeOcean, a coding environment that allows to create and make use of auto-graded coding exercises in the browser (see e.g., [17–26] for our work in this area).

Multiple-choice quizzes only allow to assess a subset of the levels of Bloom and Anderson/Krathwohl's learning taxonomy (see Figure 1.2 for a visualization of Bloom's taxonomy as revised by Anderson and Krathwohl[3].) The four lower

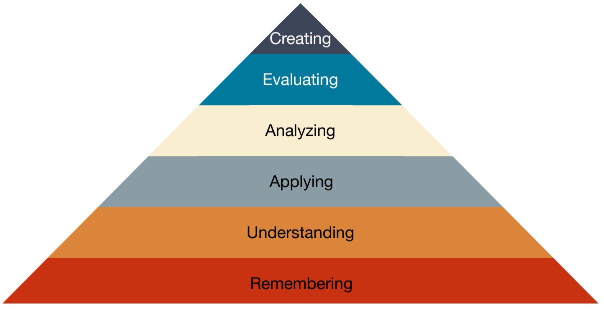

Figure 1.2: Bloom/Anderson/Krathwohl's taxonomy

levels can be assessed with multiple-choice tests as good as with constructed-response formats [27]. The complexity of creating tests to properly assess the involved competencies, however, increases level by level. The two top levels—evaluating and creating—can no more be assessed by multiple-choice

1. These values are all configurable from course to course
2. The QC is currently only offered in a small subset of courses on the openHPI platform.
3. Throughout the rest of the thesis, it will be called Bloom's taxonomy for simplicity.

tests. The other tools that are employed for automated assessment are all limited to a particular purpose. Furthermore, the complexity of creating exercises is also increasing with the complexity of the skills to be tested.

To allow the assessment of more complex projects in any domain, which are no more suitable for automated assessment, openHPI features a peer assessment tool. Peer assessment solves the scalability issue by outsourcing the assessment from the instructors to the participants instead of automating this task. Additionally, it inherently pushes all exercises to the "Evaluating" level of Bloom's taxonomy, as the participants have to assess (or evaluate) the work of their peers. Albeit it possibly can be used for exercises and tasks on all levels of Bloom's taxonomy, it is recommended to focus on the top three levels, as it comes with a certain overhead. The peer assessment feature is an integral part of the platform's tool-set to enable gradable teamwork. This tool-set is the central topic of this thesis and will be covered in more detail in the following section.

1.4 Collaborative Learning on openHPI

In 2016, Riebe, Girardi, and Whitsed [28], conducted a systematic literature review on teamwork pedagogy in higher education. According to their sources, it is no more just "desirable" being able to work in teams, it is "essential." They refer to reports from the major English speaking countries[1], as well as from Eastern Europe and China, which express the view that the ability to work in teams and the "related interpersonal skills are equally or more important than graduates' technical skills" [28] . Hughes and Jones [29], state that, nowadays, teamwork is employed in nigh on all organizations. Furthermore, they report about a poll by the Association of American Colleges and Universities (AACU) in 2009, which revealed that 71% of the employers want colleges to place greater emphasis on teamwork skills [29].

Most modern approaches to learning and teaching, such as active learning or project-based learning, work best if several students are collaborating on the same task [30]. Each student brings in a different background, different skills and problem-solving approaches. Depending on the task, these teams are ideally multi-disciplinary.

To clarify this: particularly in the context of lifelong learning, where participants often just need to learn about certain facts very quickly, these approaches are not always the ones that are suited the best for the needs of the participants and the requirements of the topic. Particularly, fact-based topics can

1. US, UK, Canada, and Australia

(and should) still be taught in a straight-forward way, e.g., by funneling the knowledge via video into the brains of the participants.

But even straight-forward subjects, have some more complex aspects that would require a lot of air-time to be explained in detail. It is often easier to explain these issues by engaging some participants in a discussion. This discussion will reveal misunderstandings and misconceptions that have been established in the participants' minds. Particularly in MOOCs, obviously, not all participants can actively contribute to such discussions. Still, passively following these discussions and watching how they evolve over time provides an analogous effect. Ellis et al. [31], define four categories of discussion based learning:

- Challenging ideas (in order to arrive at a better comprehension),
- Developing ideas (improving ideas through the input of others),
- Acquiring ideas (collaboratively amassing ideas), and
- Checking ideas (check with "authority"[1] if the developed idea is right).

For a modern digital learning environment, it is crucial to support the instructors and facilitators in offering a wide range of collaborative options. Starting with a course-wide discussion forum and ending with the possibility to form small groups or teams within the larger course community. Restricting a platform and its courses to a plain video-and-multiple-choice experience, prohibits to leverage all the advantages that come with the massiveness of MOOCs, but also in smaller contexts disables participants and instructors to go the extra-mile. Enriching the courses with hands-on tasks and exercises and enabling collaborative learning among the course participants has, therefore, been an important element of our platform since we started our first MOOC in 2012.

However, while many participants demanded for more hands-on exercises [32], collaborative learning is a top priority only for a small subset of participants [11; 12].

1.4.1 The Gap between Educational Theory and Reality

Already in the 1960s, Jerome Bruner [15], voiced the concern that:

> Short-run arousal of interest is not the same as the long-term establishment of interest in the broader sense. Films, audio-visual aids, and other such devices may have the short-run effect of catching

1. The "authority" could be either a teacher or a more experienced participant. Particularly in MOOCs where the audience often has a way more diverse background than in regular university courses.

> attention. In the long-run, they may produce a passive person for some sort of curtain to go up to arouse him. [...] The issue is particularly relevant in an entertainment-oriented, mass-communication culture where passivity and "spectatorship" are dangers.

In MOOCs, however, at least the technology does not restrict the participant to a passive role. Almost all platforms offer at least a discussion forum that allows participants to shift their role from mere consumers to active contributors. Many platforms allow for more possibilities of active course participation. In 2013, Khalil and Ebner [33], examined 30 popular courses on 6 different MOOC platforms. According to this study, 80% of the examined MOOCs provided discussion groups, 100% of the courses made use of social networks, in 20% of the courses even some local meet-up groups existed.

Unfortunately, providing the technology is by far not sufficient. For example, the participation in the course discussion forum on openHPI is somewhere between 0.02% active and 6% passive users at the lower end, and 12% active and 52% passive users at the upper end of the scale. This aligns with observations on other MOOC platforms. For example, Alario-Hoyos et al. [34], have compared the usage of a built-in platform forum versus other social media tools, such as Facebook or Twitter in a MOOC on MiriadaX in 2013. The built-in forum performed best among the examined tools. Still, 45% of the participants in a survey reported that they did not use the forum at all (1 - on a six-point Likert scale), only 2.5% answered that they use the forum very actively (6 - on the scale). Most of the rest answered in the two and three point areas of the scale [34].

Kotturi et al. [35], state that the assumption: "students will naturally populate the peer learning systems in their classes" is wrong. They reason that learning platforms do not sit "in a social setting, but in an educational setting, which has its own logic of incentives." They propose to make peer-learning systems "a required or extra-credit granting part of the course."

Grünewald et al. [11], mapped the elements of Kolb's theory of experiential learning to the techniques they applied in the second MOOC on the openHPI platform. According to them, Kolb's theory is missing the social component of learning as it is, e.g., promoted by Wenger (see [36]). MOOCs—or more general: interactive digital media—allow to append this component. A survey at the end of the examined course, however, showed that the course discussion forum and the learning groups—exactly those features that are enabling social interaction among the participants—have only been used by a smaller subset of the participants while the majority either has not used them at all or did not find them as helpful as the other course materials [11].

In 2014, Staubitz et al. [12], conducted a survey among the registered users of the openHPI platform to learn more about their preferences in terms of collaborative learning. Would they prefer to learn in groups or would they rather work alone, how active are they on other social networks, how connected are they to other learners on the platform, to which extend are they actually using the interaction channels on openHPI, are there particular interaction features that they miss? The survey revealed, that about a third of the users already knew other users either in real life or virtually. However, the platform's learning group feature has not experienced a substantial usage. Even when participants created learning groups, the activity in these groups tended to zero in most of the cases. About two-thirds of the survey participants expressed that they prefer to learn individually and expressed negative or at best neutral sentiments about learning in groups. An analogous amount of participants, however, expressed that they benefit from additional explanations by fellow participants or enjoy to provide additional explanations to others. When asked about which social features they missed most, about 50% of the participants stated that they are not interested in social collaboration features at all and another almost 20% stated that they consider the course forum to be more than sufficient for their social interaction needs within a course [12]. Obviously, there is a gap between the insights provided by cognitive and pedagogical theories—which prefer experiential and social learning over assimilative learning—and the way how most of our participants prefer to learn. To formulate it in a somewhat provocative manner: many participants prefer to consume the courses as a form of *edutainment*.

1.4.2 Participants—Motivation—Completion Rates

At this point, it is important to contemplate the participants, their goals, and motivation more closely. From the data in the user profiles, we know that most of our participants have a university degree—bachelor or higher—are working in their job since five years or more, and are in the age group between 30 and 50. A full-time job, a family, and life in general often just does not leave enough room to engage in more time intensive tasks, such as practical exercises, project-work, or team tasks. A survey conducted at the end of 2018 revealed that the motivation of most of the participants has to be considered as rather intrinsic. For example, only few of them have a concrete purpose for the received certificates [37]. Albeit the average user on the openHPI platforms has participated in about three courses, the ties between the user and the platform have to be considered as rather loose, incomparable to the ties between university students and their alma mater. So far many attempts have been made to classify the learner types in MOOCs.

Poellhuber, Roy, and Bouchoucha [38], provide an overview of these approaches and append a further element to this list. Most of these classifications are based on the users engagement and activity within the course. Starting with no-shows, dropouts, and free-riders, to active participants. They add a social component to the list by differentiating between active-socials and active-independents. In the course that they have examined, the active-socials form the smallest group in the course (5.6%). This group, however, also sports the highest survival rate (93% in the course) [38].

Staubitz et al. [20], found evidence that courses with a high and sound forum participation[1] have better completion rates, while there was no direct correlation between the forum activity of a participant and her individual course performance. Encouraging and stimulating the users to actively participate in the course discussions or to contribute to the course in another way is, therefore, considered to be important. Staubitz and Meinel [39], have proposed several approaches how to do that.

All these classifications have in common that the dropouts form the largest group. A phenomenon that has often been used to discredit MOOCs in the past. Several research efforts have dealt with this phenomenon, e.g., to predict dropouts or to determine participants that are vulnerable to drop out and encourage them to go on. More recently, this is often seen more realistically: it is just as easy to drop out from a MOOC as it is to enroll. Proper course design and most importantly a proper communication strategy targeting the course participants, including automated reminders after some days of inactivity (see e.g., [40]), are basic requirements. On top of that, there is not much to do about it. Most of the so-called dropouts are actually no-shows—users who never even visited a single learning item in the course they have enrolled for[2]. The completion rate should not be seen as the highest value in the design of a course anyway. For example, Jordan [43], has shown that MOOCs that include a peer assessment have particularly low completion rates. The reason for this, however, is not that the peer assessment itself is considered to be problematic by the participants. The main reason is that the tasks to be peer-assessed require far more effort and time to be solved than simple—or even complicated—quizzes. If the completion rate is seen as the only key performance indicator (KPI), courses with innovative pedagogical approaches will be no more possible. Next to the question "How many participants have learned something?"[3], at least, the

1. "Sound" means that reading and writing activity in the forum was alike, which indicates that the participants have communicated with each other. Activity refers to the amount of posts that have been read or written, not to the number of participants that is actively or passively using the forum.

2. This phenomenon is quite common. For example, Alario-Hoyos et al., [41] report about close to 30% no-shows in a course conducted in 2014 on MiriadaX. It is also not a phenomenon that only concerns MOOCs. Truong-Sinh, Krauss, and Merceron [42] applied data mining methods of MOOCs to other courses. They report about a 30% no-show rate in a regular face-to-face university course backed by their LMS.

3. If the completion rate of a course really is an indicator for how many participants have learned

questions “How much have the successful participants learned?” and “At which level of Bloom’s taxonomy has this learning occurred?” have to be asked.

Grünewald et al. [11], have shown that MOOCs should support different learning styles to satisfy the different needs of the different participants. Staubitz et al. [44], have shown by examining the platform’s interaction data, that different learner types exist among the participants of the platforms. E.g., John, Staubitz, and Meinel [37], have shown that achieving a certificate is not the only goal of the participants. These findings show, that the completion rate of a course is only one of many numbers.

The dropout rate, however, becomes an issue again when a course contains team projects or graded team assignments. Obviously, it can get very frustrating for the remaining team members when the others are disappearing one-by-one (or even worse, they never appear). Options and approaches how to minimize no-shows and dropouts in teams and how to deal with the resulting issues will be discussed in Chapter 4.

1.4.3 Collaboration Tool-Set on openHPI

Types of collaboration on openHPI range from low-profile, large-group collaboration (LGC), such as discussions about certain aspects of a quiz or a video in the general course forum to high-profile small-group collaboration (SGC), such as courses that set a focus on project-based learning in distributed teams—including graded team assignments. The main tool for large-group collaboration is the course forum. Additionally, the platform provides a set of diverse tools to enable small-group collaboration. This set contains a tool to match the team members, a tool to enable collaboration and communication within the teams, and a tool to assess the results of the team assignment. All involved elements will be briefly introduced now.

Course forum

The forum is the only in-platform tool for interaction between participants on the course level. Next to course or platform-wide announcements, it is also the preferred tool for communication between instructors and participants. The discussion forum in the Collab Spaces is technologically the same, but the access there is restricted to the members of the Collab Space and the instructors. The basic elements of the forum are threads and posts, where threads correspond to questions and posts can be either answers or comments. Every participant

something and how the completion rate of a course should be calculated, are different questions that are beyond the scope of this thesis.

can up- or downvote threads and posts. The author of an upvoted post receives eXperience Points (XP) for quality posting. Authors of a question also can mark an answer as the one that solved the issue. The author of the accepted answer will also be rewarded with a substantial amount of XP. Participants that have passed a certain threshold of XP, will be rewarded with a Judo-belt, which is added to their avatar in the forum.

Tool-set for graded team assignments

In the following paragraphs, the tool-set, which allows the instructors to employ and assess teamwork exercises in large scale settings will be shortly introduced. The main components of this tool-set are:

- The *Collab Spaces*—a platform feature that provides teams and groups with an area of their own within the course. It offers a set of communication and collaboration tools.
- The *TeamBuilder*—a standalone web application to form teams out of a given pool of participants, based on a variable set of parameters. It allows to limit the number of participants that will be admitted to attend the team assignments. The *TeamBuilder* can be connected to any Learning Management System (LMS) using the Learning Tools Interoperability (LTI) interface. To leverage the tool's full power, however, a tighter integration via APIs is required.
- The *Team Peer Assessment*—an extension of the platform's peer assessment tool. It allows teams to jointly submit digital artifacts. The submissions of each team are reviewed and assessed individually by the members of the other teams. It additionally, allows the team members to rate the work of their teammates in terms of contribution, organization, and social skills.

Collab Spaces

Originally, the Collab Spaces[1] mainly provided a private discussion forum and the possibility to share files for the group members. Further communication and collaboration channels have been added later on. Each participant can create such Collab Spaces and can decide whether the room will be public—open for everybody to join—or private—only invited learners can join.

1. Initially, the Collab Spaces were called *learning rooms*. Later, for reasons that are beyond the scope of this thesis, they were renamed to *Collab Spaces*. In the context of the platform, as old habits are hard to get rid of, both terms are still used synonymously in some places. *Collab Space* is the preferred term in this thesis. *Learning room* might be used in certain contexts where it particularly refers to an older version of the platform.

Basically, the Collab Spaces provide synchronous and asynchronous communication tools as well as some co-creation tools to jointly work on digital artifacts. For synchronous communication Google Hangouts have been integrated, for asynchronous collaboration a separate area of the discussion forum is provided, which is only accessible by the members of a Collab Space. The supported co-creation tools are a Tele-Board[1], basically, a brainstorming and planning tool that allows team members to jointly create virtual sticky notes on a virtual whiteboard [45] and an Etherpad[2], which allows participants to synchronously work on the same text documents. Additionally, the Etherpad provides a chat function.

The *Collab Spaces* differentiate between *Groups* and *Teams*.
Groups—are loosely coupled and have a self-set goal or a common interest:

- Several employees of the same company who want to discuss how to apply what they have learned in the context of their company[3].
- A group of participants discussing about a special interest topic that surpasses the communicated learning goals of the course.
- A school class participating in a course with their teacher.
- A group of e.g., Germans in a course offered in English, who prefer to discuss in their native tongue[4].

Groups are self-organized. The participant who creates the Collab Space has administration privileges and decides if it is public or private. Participants can join/apply for any number of groups within a course on their own initiative. They also can leave a group whenever they want. It turned out, however, that providing the participants with a set of collaboration tools and expecting that they will start to collaborate on their own initiative has only been successful on very rare occasions. This correlates with the findings of Kreijns, Kirschner, and Jochems [46], who stated that "interaction does not just happen, but must be intentionally designed into the instruction." Our experiments and findings in this context will be discussed in detail in Section 4.1.1.

1. The Tele-Board is a research prototype of a digital whiteboard and has been developed at the HPI. It is currently being replaced by its commercial variant, the neXboard (https://www.nexenio.com/en/nexboard/), mainly for better support and maintenance.
2. http://etherpad.org/
3. As this is still an open platform, they obviously should be careful not to discuss company secrets, etc.
4. We, generally, close discussions in the main discussion forum that are not in the course language. In courses that are offered in English language, but still have a large audience of German natives this happens at times. The participants are then offered to either switch to English or to create a Collab Space for German natives.

Teams—are tightly coupled and have a common task, on which their members are jointly working. The task is an essential[1] element of the course and part of the grading scheme. Teams are formed by the instructors with the help of the *TeamBuilder*. Only instructors can add members to—or remove them from—teams. Per course, each participant can only be a member of one team.

Collab Spaces for teams act slightly different than Collab Spaces for groups. For example, team members always receive an email whenever a new thread or post is created in the team's discussion forum. Thus, the forum becomes a more powerful communication tool. As with more power comes more responsibility, this behavior has been restricted to the *Team Collab Spaces*. If this was the normal behavior of the course forum—or even a group Collab Space forum—the most probable result would be that the users switch off receiving email notifications. Even if only a small percentage of users are writing in the forum, the cavalcade of emails would be overwhelming. Additionally, team members can participate in *Team Peer Assessments* (while group members cannot).

TeamBuilder

The *TeamBuilder*, a standalone tool that can be connected to any LMS supporting the LTI interface (see Figure 1.3), provides a limited set of parameters to be used for the team matching process:

- The participant's preferred language
- The participant's location or timezone
- The participant's area of expertise
- The participant's time commitment
- The participant's age and gender
- Course specific parameters

Course specific parameters might be a choice of tasks, a list of the participants' schools, or whatever the instructors of the course come up with. Instructors can activate the parameters in which they are interested and deactivate the others. Furthermore, they can decide in which order the matching-criteria must be applied and if a parameter is to be employed homogeneously or heterogeneously. The chosen parameter-set has to be published before the data can be collected from the participants to avoid incomplete datasets. Once the list of required parameters is published, it is no more possible to append additional parameters.

1. The essentiality of the task for the completion of the course, can range from a small bonus task providing a few extra points to the central assignment of the course.

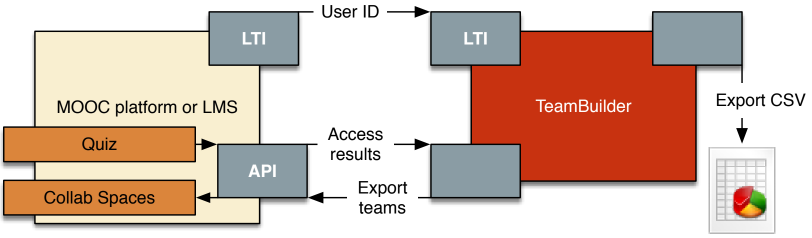

Figure 1.3: The TeamBuilder can be connected to any LMS that provides an LTI interface. To leverage its full power, however, a tighter integration is required.

It is possible, however, to deselect matching criteria later on if it turns out that the selected parameters will not allow to form expedient teams.

It was deliberately decided to work with a relatively simple and limited dataset. The participants are only loosely coupled to the platform, and the time for matching the teams is very limited. Therefore, the participants are not asked for more personal criteria, such as personality traits, etc. All data that are asked from the participants, however, are mandatory. Those who refuse to provide the relevant piece of information, will not be considered for the team assignments.

Next to the selection of matching criteria, the tool allows to define a range of values for the desired team size.

Finally, the tool allows to limit the total number of participants for the team assignments within a course. The limiter can either act on a first-come/first-serve basis, or on the participants' performance in the assignments whose deadline has passed before the team building process has started. First come/first serve simply allows the instructors to select the first N participants that have applied. If the instructor decides to select the eligible participants based on their results in the previous assignments, the tool allows two different settings:

1. Set a maximum amount N of participants to be admitted for teamwork—The N participants that performed best so far will be admitted.

2. Set a lower limit for the results in the exams that has to be achieved—In addition to the total maximum amount of participants, a lower limit for the results can be set. Only participants that have achieved at least $X\%$ of the points in the relevant exams—but no more participants than the total maximum—will be admitted. Limiting the number of participants based on the results they achieved in the previous assignments requires a tighter connection between the *TeamBuilder* and the LMS than just the LTI interface (see Figure 1.3).

The findings of Zheng, Vogelsang, and Pinkwart [47], support our idea of requiring the participants to show some commitment before they are allowed to participate in the team assignments. They state that participants that have at least partially answered surveys before joining the teams, were less likely to drop out than those who did not take the surveys at all [47].

While the *TeamBuilder* attempts to build teams fully automated, it still allows manual corrections. The final decisions are made by the instructors. Participants can be moved from one team to another, new teams can be created and existing ones can be removed, etc. Finally, there are two options to export the created teams. The *TeamBuilder* can either create a CSV file for further usage, or it can directly create Collab Spaces on the platform and add the team members to their respective Collab Space.

Peer assessment

The openHPI platform's peer assessment tool allows both, summative and formative assessment. For the summative part, the instructors define rubrics, which are used by the participants to grant points to their peers. In the formative part, the participants are asked to write a review that provides their peers with constructive criticism. Furthermore, submissions as well as reviews can be reported to the instructors if a participant suspects that there is something wrong with the submission, or that the submission contains plagiarisms. Basically, the peer assessment consists of three mandatory and two optional steps:

1. Step 1: Submission—Submit your digital artifact for review (mandatory).
2. Step 2: Training—Review exemplary submissions of peers that were already graded by the instructors and compare the results (optional).
3. Step 3: Review and grading—Grade the work of your peers and write helpful reviews (mandatory).
4. Step 4: Self-evaluation—Evaluate your own work in comparison to the work of your peers (optional).
5. Step 5: Grade the graders—View your results and the reviews that you received and reward the peers that wrote helpful reviews with a good rating (mandatory).

"Optional" means that the instructors can either omit them completely, or they can include them and define participation in the step as either mandatory or optional. Step 5 is only partially mandatory. It has to be included, but the participants do not necessarily have to view their results or rate the received reviews to receive their points. However, if they want to see how many points they have received from a particular reviewer (still double blind), they have to take this step and rate the received review. If they have missed any of the other mandatory steps, they do not receive any points. For more details about the single user peer assessment feature see [48].

Team peer assessment

Extending this system to allow the assessment of teamwork, first required the decision "who grades whom." Should teams grade other teams, should individual team members grade other teams, or should the members of a team grade their teammates? Figure 1.4 shows a schematic view of the chosen option. Basically, it was decided that individuals rate the work of other teams.

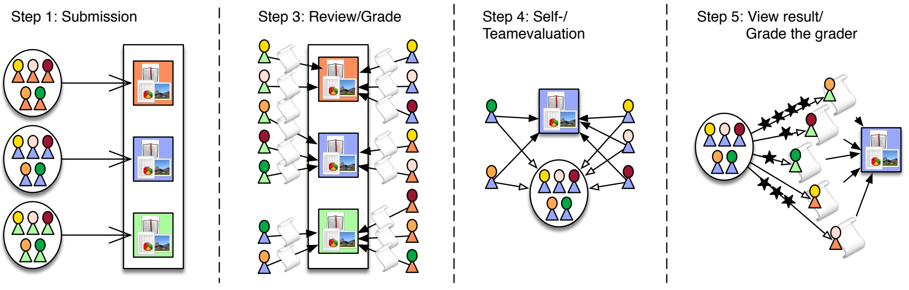

Figure 1.4: Team peer assessment steps

Optionally, team members can rate their teammates within the team. This approach provides us with several advantages:

1. Submissions receive more reviews, while the workload for the team members is kept low.
2. Managing the review process is easier, as the team members do not have to be coordinated and no versioning is required.
3. Participants have fewer opportunities for free-riding. Even if they manage not to contribute to the actual task, they still have to write reviews if they intend to receive points.

The following list introduces the most significant—user facing—changes.

Step 1—Every team member can upload the team's solution. As long as the solution is just drafted—and not finally submitted—every team member can upload a new version of the solution and overwrite the old version. Every team member can finally submit the team's solution. The submission of a team's solution is always accredited to all team members.

Step 2 and 3—The system ensures that participants will review only submissions of the other teams and not their own.

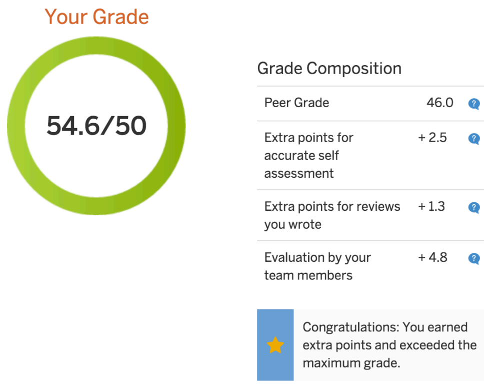

Figure 1.5: Screenshot of a participant's results view. The "Peer Grade" component contains the points that the team received together. The points in the other components have been achieved individually by each team member. The "Peer Grade" component defines the possible maximum of achievable points. Individual components provide bonus points. The example shows that, therefore, the maximum can be exceeded.

Step 4—Each team member can evaluate the team's submission in comparison to the work of the other teams. Additionally, each team member can rate her teammates in terms of contribution, organization, and social skills.

Step 5—The overall grade consists of a team component and several individual components. Figure 1.5 shows how these components are presented to the participant. The team component defines the maximum of achievable points, the individual components are added as bonus points. Therefore, the maximum of achievable points can be exceeded. As in the single user peer

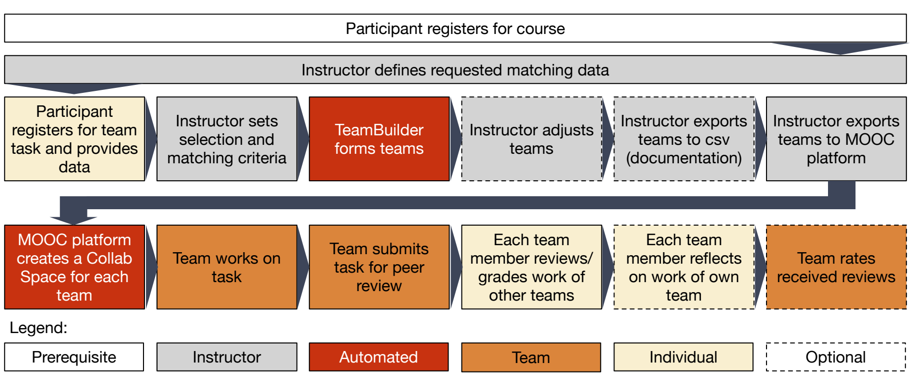

Figure 1.6: Complete workflow of team tasks on openHPI

assessment, participants can report submissions and reviews, whereas reporting submissions is an individual effort, while reporting reviews is a team effort. The team can also reward reviewers that wrote helpful reviews with additional points. Figure 1.6 shows the complete workflow that participants and instructors have to go through to set up and participate in a peer-assessed team task.

1.5 Structure

The rest of the thesis is structured as follows:

Chapter 2 provides an overview on the research questions, the applied methodology, and the examined datasets. Chapter 3 presents related work in the areas peer assessment, collaborative learning in large and small groups, project-based learning, and teamwork—team matching, team formation, and grading of teamwork. Chapter 4 discusses differences between loosely coupled groups and tightly coupled teams. Chapter 5 is the centerpiece of the thesis. It explores graded team- and project-based assignments in a selection of MOOCs on various topics. These courses/assignments have been accompanied by experiments, surveys, interviews, and an analysis of the course data. Chapter 6 and Chapter 7 conclude the thesis and present ideas for future experiments and improvements.

The following terms will be used, for the rest of this thesis:

- *MOOC*—If not stated otherwise, the term MOOC will be used in the meaning of xMOOC.
- *Platform*—Generally any MOOC platform. In the context of this thesis, it mostly refers to the HPI MOOC platforms: openHPI, openSAP, OpenWHO, and mooc.house.
- *Users*—Registered users on one of the HPI platforms.
- *Participants*—Users who have enrolled in a course.
- *Courses*—Courses on any of the HPI platforms will be referred to by their course code in italics, e.g., *javawork2017*. A mapping of these codes to the course title and a detailed description of the course including the course URL and details about the team assignment is provided in Appendix I - Courses.

2. Research Questions, Methodology, and Datasets

Hitherto, the stage for the thesis has been set. In summary, the general topic is fostering collaborative learning in MOOCs. The particular focus is on forms of small group collaboration within MOOCs and even more particular on gradable team assignments. The tool-set that has been implemented to allow such gradable collaborative assignments has been introduced in Section 1.4. In the context of MOOCs, collaborative work and particularly small group collaboration (SGC) as team assignments face particular challenges:

- Because of the high dropout rate, it is anything but sure that those who have registered for a team task will still be there when it actually starts. Timing of the registration for the team task, proper communication and predicting probable dropout candidates are key issues.

- Owed to the short duration of the courses (2-6 weeks), it can be challenging to fit even one of these assignments in a course. As many participants are full-time employees, the instructors attempt to ascertain that the participants can work on more time-consuming tasks on weekends. As team assignments are graded by peer assessment, at least two weekends are required. One to work on the task, a second to review the work of the others, maybe even a third to rate the reviews.

- As in any other learning context, teamwork does not fit for all courses, tasks, and participants. In our courses on openHPI, we, therefore, carefully select tasks that make sense to be assigned to teams and try to offer an alternative to the team tasks where possible.

In the following sections, an overview on the research questions that are to be answered, the methodology that has been employed, and the datasets that have been examined, will be given.

2.1 Research Questions

The thesis attempts to answer the following general questions:

1. Why are users participating in the team assignments?
2. Can teamwork reduce the attrition rate in MOOCs?
3. How can we reduce the attrition rate in the teams?
4. Can we predict who will drop out of the team assignment?
5. What is the best way to match the team members?
6. What are the needs of the teams in terms of tool support and mentoring?
7. How should the assignments be graded?

More specific subquestions will be added in context.

2.2 Methodology

To answer the research questions, a long-term study following an iterative mixed methods approach, has been designed and conducted. Prototypes of platform features have been developed and employed in the courses. Where possible, they have been tested in pilots—hidden courses with a limited number of selected participants. Feedback has been collected in surveys. One-hour semi-structured interviews have been conducted in two of the courses. Some of the interviewees have also been observed while interacting with selected platform features. The collected feedback as well as the platform's learning

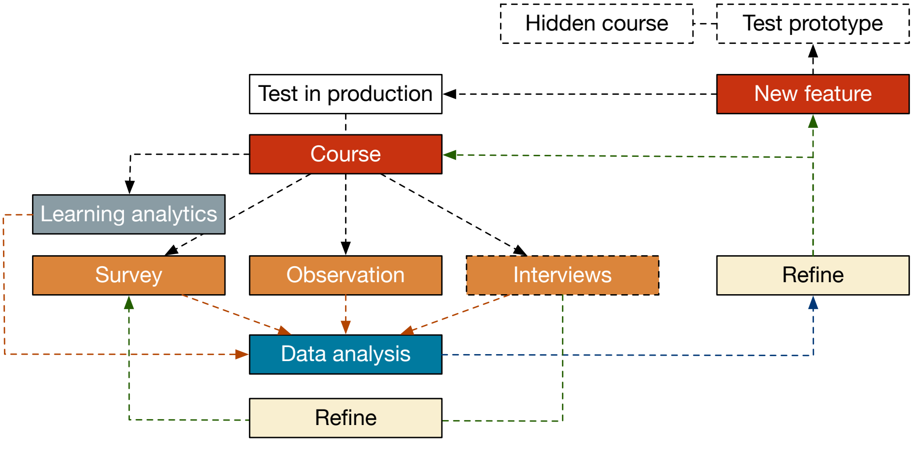

Figure 2.1: Iterative mixed methods approach.

analytics data have been analyzed, and the features have been refined according to the results (see Figure 2.1). This process went through several iterations from 2016 to 2019. Preliminary work on some of the features had already started in 2013. In a keynote at L@S2019[1] in Chicago, Candace Thille[2], presented a virtuous circle of “Better learning activity interactions” →“Better learning data and models” →“Better insights” →“Better learning experience” (see Figure 2.2). She stated that “the connection between research and teaching is broken” and that, therefore, this virtuous circle currently does not work. The research approach that has been employed for this thesis, attempts to fix the broken link between research and teaching (and engineering). Platform-design, course-design, course-delivery, communication with the participants, and research are combined under one roof.

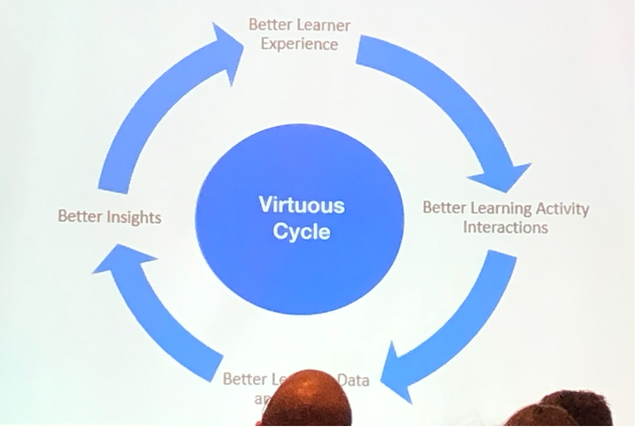

Figure 2.2: Virtuous circle between research and teaching. Keynote by Candace Thille at L@S2019 in Chicago.

Particular issues in the context of this thesis are *scalability* and the participants’ *ability to communicate*. Resulting from the participants’ ability to communicate, classical experiment setups—e.g., A/B testing—are often hard to put in practice. The participants, generally, ascertain rather quickly that not all of them are treated the same way. Therefore, often different iterations of a course had to be compared, sometimes even different courses. Furthermore, several experiments had to be conducted simultaneously, as not every course includes a team peer assessment and it would have taken ages to do otherwise.

1. L@S is the ACM conference that sets a focus on investigating large-scale, technology-mediated learning environments, such as e.g., MOOCs.
2. Former director of the Stanford Open Learning Initiative and co-director of the Stanford Lytics Lab. Currently, director of learning science and engineering at Amazon.

This comes with a price. The data are not always comparable and there are always additional aspects next to the examined ones that might influence the results.

Owed to the scalability requirement, even the prototypical implementation of a feature requires a massive effort. Small errors that could easily be fixed in a laboratory setting with a few selected participants might lead to a severe catastrophe in a time-boxed course with thousands of participants. So before a feature can be tested for its scalability, it has to have reached a certain state of maturity. Fail often and early is not really an option in this context. Unfortunately, it can easily become an unwanted reality as many errors only show when a feature is driven to its limits by a massive amount of participants. In our research as well as in the operation of the platform, we are heavily dependent on the users' satisfaction with the learning experience.

A failing feature that renders the participants' work obsolete, is not an option. If it happens, proper mechanisms to work around the technical issues have to be found. Often enough, these workarounds come with a high amount of manual work[1]. Whenever something goes wrong, the reputation of the Hasso Plattner Institute or a partner—particularly SAP, which played a very active role in the development of these features—is at stake. So even a prototype always requires a substantially higher amount of effort for its implementation than features that are tested in a laboratory environment.

In 2016, the first course containing a graded team assignment has been conducted on the openSAP platform. Instructors on the openHPI and openSAP platforms, so far, have employed peer assessments in about fifty courses[2]. The author has been directly involved with producing and conducting about twelve of these courses and had a consulting role in several others. Team-based tasks have been offered in about fifteen courses, in a variety of settings and contexts, so far. Most of these tasks have been team peer assessed. The author has been actively involved in the teaching team of about five of these courses.

1. An example: In one of the first peer assessments, the users had to submit a piece of HTML. The task did not state clear enough, that the HTML should be attached as a separate file. Many participants copy/pasted their HTML into the feature's text input field. Resulting from a small bug in the feature—basically, one line of code was not indented correctly—the peers could not download the other attached files. The author and one of his colleagues had to re-grade several dozens of submissions manually, which cost each of us about one day of work.

2. Additionally to that, Signavio—a course provider on the mooc.house platform—employs peer assessments with great success in almost all their courses. These courses have not part of the evaluation, because of their comparably small number of participants.

2.3 Datasets

This section provides an overview on the types of data that have been examined in the context of this thesis. Basically, the data can be categorized in three types. First, self-disclosures of the participants, e.g., the users' profile data, surveys, and interviews. Second, general platform data, such as course information and course properties, e.g., enrollment numbers, completion rates, etc. Finally, the dataset, which is produced by the participants' interaction with the platform, while they are engaging in learning.

2.3.1 User profile data

On openSAP and openHPI the participants are asked to provide some socio-demographic background data, such as age, gender, previous education, job status, etc. The profile data are available for about a third of the participants, as it is voluntary to provide these data. On mooc.house, user profile data are not collected in this detail.

To register for the teamwork, the participants have to provide the data that are needed for the team matching. The type and amount of data that have to be provided can differ from course to course. These data are complete for this subgroup as it is mandatory for the registration. Finally, geographic information is collected via the participants' IP-address, once they become active within a course. Geographic data, currently, are missing for no-shows.

Generally, the participants come with a variety of backgrounds—ranging from school-kids to retirees—and previous experience—ranging from beginners to experts.

Except for a few courses that have particularly targeted pupils in school, the strongest group of participants are in their thirties to fifties, working in a regular job since about ten years, and often have at least a bachelor's degree. About 80% of the participants are male. Most participants on openHPI reside in Germany. Courses that are offered in German language, have close to 100% German audience, plus a few people in Austria and Switzerland. Courses that are offered in English language have an international audience with strongholds in Germany, India, and the US. The situation on openSAP is analogous—with the distinction that most courses on openSAP are offered in English language, while a little more than half of the courses on openHPI are offered in German language.

2.3.2 Learning analytics data

Visited items, points in quizzes and assignments, general performance in the course, and participation in the forum discussions play an important role in the study. Particularly, to examine the team assignments, two datasets have been created in which all relevant data sources, such as course reports, peer assessment reports, and TeamBuilder reports have been merged. The first set contains all the data on the participant level. A detailed description of this set is available in Appendix V - Team Member Data. The second set aggregates the participant data on the team level. A detailed description of this set is available in Appendix IV - Aggregated Team Data. A third dataset has been created that contains the available data for those course participants who have not participated in the course's team assignment for comparison. A detailed description of this set is available in Appendix VI - Non-Team Participant Data.

2.3.3 Course data

A detailed description of the relevant courses, the number of participants, and particularly the type of exercises and assessments that they provided, is listed in Appendix I - Courses. The courses that are listed there have played a role in several particular research efforts that are discussed in this thesis. The given grade of detail, depends on the relevance of the course for the central theme of the thesis. A *tag* has been created for each of the particular research efforts and has been applied to all courses that played a role in it. These tags are also explained in more detail in Appendix I - Courses.

2.3.4 User surveys

User surveys about different aspects of peer assessment, the Collab Spaces, and team assignments have been conducted in many courses that contained one of these elements. When comparable surveys have been conducted in different courses, they are aggregated or compared. Where possible, they are also compared to the results of other researchers. The list of relevant survey questions and the context in which they have been asked is available at Appendix III - Surveys.

2.3.5 Interviews

Altogether thirty-two semi-structured one-hour interviews have been conducted with users who participated in the team assignments. Fourteen interviewees participated in the team assignment of *javaeinstieg2017*, eighteen interviewees participated in the team assignment of *java1*. Some of the interviewees have been asked to interact with the relevant platform features while being observed by the interviewers to detect usability issues. The complete, pseudonymized list of interviewees is available in Appendix II - Interview Summaries. An Excel file containing the summaries of all interviews is available there as well. The

Figure 2.3: Interviews word cloud *javaeinstieg2017* (left) and *java1* (right)

.

interview summaries have been processed using MaxQDA[1]. Figure 2.3 shows word clouds that have been extracted from the interview data. The clouds show that in both interview sets, the topics have been analogous. It will be shown later on, however, that the general perception of the teamwork has shifted in a positive direction between the two courses.

1. MaxQDA is a tool to explore, tag, analyze, and visualize qualitative and mixed methods data https://www.maxqda.com/what-is-maxqda

3. Related Work

This chapter presents related work in the areas of assessment, collaborative learning, and teamwork.

3.1 Assessment@Scale

Resulting from the massive amount of participants in MOOCs, manual assessment of tasks and exercises by the instructors is unfeasible. Basically, there are two possible approaches to deal with this challenge:

1. Automating the assessment
2. Outsourcing the assessment

Automated assessment has a long history in e-learning. For the assessment of project work, however, it comes with severe limitations and will, therefore, only be discussed briefly for completeness. Outsourcing assessment, basically, comes in two flavors:

1. Paid outsourcing—Models such as Amazon Mechanical Turk[1] or Udacity's mentoring model[2]
2. Outsourcing the assessment to the course participants themselves—Peer assessment

Paid outsourcing is not an option in the thesis' context and, therefore, will not be covered here. Peer assessment is one of the thesis' focus areas and will, therefore, be covered in depth.

3.1.1 Automated Assessment

Multiple-choice tests and their derivatives are the simplest, and therefore most common, form of automated assessment. This form of automated assessment—

1. https://www.mturk.com/
2. https://www.udacity.com/start-mentoring

and to some extent automated teaching[1]—dates back to the teaching machines of the late 19th and 20th century. The most famous protagonist, probably is B.F. Skinner[2], a psychologist at Harvard, who is often credited to be the inventor of the teaching machine. However, e.g., Sidney Pressey in the 1920's and Halcyon Skinner in the 1860's have preceded B.F. Skinner's work [54].

Derivatives of multiple-choice quizzes (MCQ) are multiple-answers, fill-in-the-gaps, or matching terms. If employed intelligently, these quizzes can be a very valuable form of assessment and can be used way beyond simple examination of remembered knowledge. Literature, e.g., Zheng et al. [55], often states that these quizzes are fit to assess at least the lower four levels of Bloom's taxonomy. The fifth and sixth level—Evaluating and Creating—are only very hard if not at all assessable with MCQs and the likes.

Next to MCQs, several domain-specific assessment tools exist. In the domain of mathematics, e.g., STACK[3] by the University of Edinburgh is such a tool. CodeOcean[4] (see [20]) is frequently used, on the HPI platforms, for the automated assessment of programming exercises. Further tools in the domain of programming are e.g., Praktomat[5] or Web-CAT[6]. [19] Even in formalized domains, which are generally supportive to automated assessment, such as mathematics or programming, automated assessment comes with certain limitations, however. Automatic assessment of programming exercises often relies on unit tests. Creating such exercises is not quite easy and requires much time[7].

---

1. The multiple-choice quizzes on openHPI can be configured in several variations. They can be self-tests or graded exams. Particularly, as self-tests they can be used to expand or enrich what has been said in a previous video. Explanations can be added, which are only shown when the participant has submitted her answers.
2. Skinner is best known for his work on the behaviorist education model [49] [50]. This model was built on the theories of J.B. Watson, who himself was influenced by Pavlov's work on conditioned reflexes [51]. Skinner proposed that children learn to speak in their mother tongue based on a given set of words and sentences, positively reinforced by receiving little rewards when they do it right [52]. The behaviorist learning theory assumes that all behavior are mere reflexes on consequences that result from previous behavior. The teacher is assumed to be a benevolent dictator rewarding or chastising the learner for right or wrong behavior, whereas rewarding correct behavior is rated to be superior to punishment [52]. Noam Chomsky already back in 1959 doubted that—if Skinner was right—humans would be able to learn to use their language in its infinite variety of combinations [53], one of the abilities that define the distinction between human and animal.
3. https://www.ed.ac.uk/maths/stack/
4. https://github.com/openHPI/codeocean
5. https://github.com/KITPraktomatTeam/Praktomat
6. https://github.com/web-cat
7. As an alternative, static code checking could be used to replace or complement dynamic checks, such as unit tests. In contrast to dynamic checks, static code checks do not test if the submitted code delivers the expected result when it is executed. It rather checks the source code if all formal requirements are fulfilled. This starts with testing if the code follows a certain style guide, but, particularly in strictly typed languages, such as e.g., Java, goes well

Particularly, in "open" programming projects[1], it quickly can result in a huge workload for the instructors [21]. The complexity of the required tests is increasing with the complexity of the assignment.

The openHPI team has developed further tools to test if e.g., emails are signed and encrypted correctly, or if RDF triples[2] follow the instructors' specifications. All these tools have in common that they are very domain-specific and cannot be used for general purposes.

3.1.2 Peer Assessment

As shown in the previous section, automated assessment either does not cover all levels of Bloom's taxonomy—particularly, not the ones associated with higher-order thinking skills—or it is domain specific and separate tools for each domain have to be developed. Machine-grading of hardly quantifiable criteria, e.g., elegance, style, and creativity, (still) is difficult or virtually impossible. Additionally, individualized feedback is an integral part of education, but (still) cannot be delivered by using automated assessments [56].

Peer assessment (PA) is employed in today's MOOCs as an attempt to address these issues. This method allows participants to receive personalized feedback and to engage in challenges that go beyond the capabilities of automatic machine-grading by allowing the participants to grade and comment each other's work [56]. Benefits of peer assessment include improvement of higher-order thinking skills, consolidation of topical knowledge, and individualized feedback for each participant [57; 58]. Educational assessment does not only aim to measure the students' achievement it also aims to deepen their comprehension [59]. The learning effects of PA can thus not only be seen as welcome byproducts of the process, they might even be set in the focus. Basically, each task, notwithstanding its nature, is lifted to the "evaluating" level of Bloom's taxonomy.

PA as a form of educational assessment is very flexible and can be used to serve summative and formative assessment alike [60]. It is a quite common application of formative PA that students are reviewing each other's work and are giving written feedback [61; 62]. Summative PA of fellow students' work, however, is a more complicated matter and requires careful guidance by a teacher, since grades should be fair, consistent, and comparable for all students [62–64].

---

beyond that. The advantage of such static code checks is that they are often well supported by existing libraries and most of all, that they do not have to be specified at the exercise level, but can be defined and configured once per supported programming language.

1. "Open" in contrast to exercises, where many prerequisites, such as e.g., identifiers, etc. are predefined by the instructors
2. https://en.wikipedia.org/wiki/Semantic_triple

A common way to enable participants to peer-grade other participants' work is to use teacher-designed grading rubrics. Literature distinguishes between *objective* and *subjective* assignments. In objective assignments, the correctness of an answer is clearly defined. In contrary, subjective assignments allow the students to solve the given task in multiple ways without a clearly defined correct outcome [65]. Creative writing is an example of an eminently subjective task. Gauging the quality of such a piece of work through PA is difficult. But, particularly in a MOOC environment, such tasks are the ones that most-likely will be peer assessed, as the massive nature of these courses forbids manual grading by the instructors, and an automatic grading of creative work is only hard to imagine. Rubrics are used to counteract the subjectiveness during the grading by providing students with teacher-designed categories that communicate the quality expectations that a piece of work should meet [65]. They seek to guide students through the grading and often elaborate on the criteria, e.g., by giving examples how many points to award for which expertise and completeness displayed in the work of a peer [66]. Knight and Steinbach divide peer assessments into three guidance categories:

1. open-ended—little guidance, e.g., in form of rubrics, is given,
2. guided—general hints and questions to contemplate are provided,
3. directed—a detailed, checklist-like guidance is provided.

They argue that directed PA is superior to the other categories, since it also enables less knowledgeable students to assess their peers' work [58].

Self-assessment is often recommended as a complement to peer assessment, because of its analogous didactic and cognitive benefits [67; 68]. Having seen and assessed the work of their peers, students evaluate their own work on the same criteria that they used to evaluate the work of their peers [69].

Studies examining student's performance improvements through PA participation report varying results. There is a consensus that performance improvements largely depend on the specific application and learning environment where PA is employed [70; 71]. If applied responsibly, students have been found to improve in general course performance if they participated in PAs for this particular course [70; 71]. Feedback generally is perceived useful by students. Some studies suggest that some students take comments from their peers more seriously than teacher comments [64; 67; 70]. Accurate grades have been reported for both peer and self-assessment, whereat the accuracy relative to an accepted standard, such as teacher-assigned grades, is ameliorating with the number of reviews to be averaged [70; 71].

Responsibly applied PA should allow an explicit grading training to increase the students' aptitude and confidence to assess the work of their peers [71].

Based on the work of Kulkarni et al. [72], Coursera was the first MOOC platform to use calibrated peer review (CPR), which transfers the idea of training sessions to online environments [73].

Peer assessment in face-to-face, on-campus courses, mostly has been employed in the context of smaller groups guided by a teacher or teaching assistant and often as a supplement to teacher assessment [56]. In the context of MOOCs, increasingly heterogeneous groups of students participate in PA. Owed to the different backgrounds and knowledge of students, student aptitude and grading accuracy is doubted. PA itself, as a valid assessment form, is sometimes being challenged by course participants [56; 58; 74][1]. Therefore, it depends on the instructors to narrow down quality expectations, for example by providing detailed rubrics to ensure the success of PAs and to keep assessments comparable, consistent, and fair [64; 71]. Students' subjectiveness based on their culture, education, and knowledge of the given topic will influence the way in which a student grades to a certain degree [72]. Bias can partially be counteracted with multiple reviews per peer (averaging a grade), well-defined expectations, trainings, and anonymity, [58; 68]. Double-blind peer reviewing allows participants to provide more critical feedback and freely express their opinions without having to consider interpersonal factors, which in turn results in more honest and, ideally, more useful reviews [75; 76]. Then again, Gamage et al. [77], report about an experiment in which they showed that non-blind PA leads to significantly longer and more useful feedback.

Several authors identified rogue reviews as a challenge for peer assessment. Rogue reviews are insufficient reviews caused by laziness, collusion, dishonesty, retaliation, competition, or malevolence [58; 70]. These were always a challenge for PA, but were diagnosed to be a bigger challenge online, owed to an increased anonymity and a decreased feeling of community affiliation [61; 75].

With the rise of MOOCs, peer assessment became ever more important. Wu et al. [78], presented an approach to increase the accuracy and scalability of peer assessment based on game theory. They claim that their mechanisms are outperforming traditional peer assessment mechanisms. They tested and verified one of their mechanisms through a crowd-sourced experiment. The participants had to "grade" submissions containing a randomly generated set of colored objects by counting the objects with a certain color [78]. The dilemma with such an evaluation, however, is that it has little to do with reality, where the peers have to assess way more complex assignments, and the decision

1. A detailed analysis of the forum discussions in the openHPI and openSAP courses featuring peer assessments, revealed that the participants have a generally positive attitude to PA. The study was intended to be part of this thesis, but had to be removed because of the length limit. It will be published separately.

how to grade is often by far less obvious than in this test. Hicks et al. [79], examined the difference in the quality of formative reviews if the participants have to assign a numerical value (rating) to the submission first—and observed a significant positive effect on the content of reviews[1]. Sajjadi, Alamgir, and von Luxburg [80], experimented with different statistical and machine learning methods to aggregate several peer grades throughout a semester to come up with accurate final grades for the submissions. Interestingly, the simple mean outperformed all of them [80].

The consequences of not reviewing the submissions of the peers differ between the peer assessment systems. Some regard the evaluation step as mandatory. No reviews written→zero points. Others penalize only the final result by deducting a certain percentage of points for the submission [81; 82]. Most peer assessment systems recommend to request three to five reviews from each participant. The challenge is to receive a sufficient amount of reviews to calculate proper averages without overloading the participants with reviewing chores. Kulkarni et al. [72], have shown that the median of the received grades performs better than the average of grades. Raman and Joachims [83] suggest to employ ordinal instead of cardinal grading. Participants order the submissions they have been assigned to for grading as: *a better than b better than c* [83].

3.2 Collaborative Learning@Scale

The participants of a MOOC, particularly those, who—actively or passively—participate in the discussion forum, form a *community of practice*. Wenger [84] has defined a community of practice as based on two pillars:

- Sharing a concern, challenge, or topic,
- Deepening the knowledge on that topic by ongoing interaction.

According to Wenger, communities of practice basically are everywhere and downright emerge on their own [85]. Grünewald et al. [11], delineated how this concept maps to MOOCs. Albeit the observations in the courses on the HPI platforms confirm Wenger's view, it also became obvious that the participation in these communities can be improved by some rather simple interventions of the instructors.

1. We conducted an A/B test on openSAP to determine the difference in the length of the written reviews when the succession of summative and formative assessment is changed (First the buttons to rate and then the text input for the review or the other way round). The test revealed that the results were identical

The benefits and challenges of collaborative learning, project-based learning, and team-based learning are not a new topic, so there is no shortage of literature to explore. Springer et al. [86], show positive effects on achievement, persistence, and attitude. Kreijns et al. [46], list deep learning, long term retention, improved social and communication skills, and the formation of social relationships. Laal and Ghodsi [87], list the improvement of soft skills, building social support systems, reducing potential anxiety, active involvement in the learning process, better classroom results, and increased critical thinking skills. Hiltz et al. [88], state that the results achieved in online collaborative learning can be as good or even better as in a traditional classroom setting. The results of those online learners who studied alone (serving as a control group) turned out to be the poorest of all [88]. Vygotsky [89], delineates learning as a collaborative activity in which the environment influences the individual. According to his "Zone of proximal development" [89], learners perform better

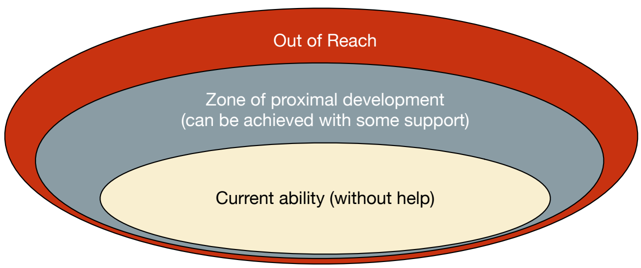

Figure 3.1: Vygotsky—Zone of proximal development

if they receive guidance by more experienced people, which might be either instructors or peers (see Figure 3.1). In this context, Mitra [90], showed that these "guides" do not necessarily need to be more knowledgeable of the task's topic to have a substantial effect on the learning process of the "guided." He also showed that, at least for children, this only works if they work on the given task in groups [90].

Social constructivism, a learning theory building on the work of Piaget and Vygotsky, defines learning as the process of appropriation of knowledge, not only internalizing it but transforming it in an individual way. Learning occurs through interaction and collaboration, thus, it is an inherently social activity [91]. Social constructivist researchers postulate that

> (...) by drawing upon a larger collective memory and the multiple ways in which knowledge could be structured among individuals working together, groups could attain more success than individuals working alone [91].

Social constructivism is not to be confused with Bandura's social learning theory. Bandura [92], basically picks up the behaviorist position that humans are learning by stimulus-response patterns. He states, however, that individuals are not learning by receiving positive or negative reinforcement themselves, but by observing and cognitively processing the consequences of other individuals' actions [92].

3.2.1 Virtual Tutors and Peer Tutors

Virtual tutors to stimulate the interaction among the participants, are an active field of research since more than 30 years (see e.g.,[93–96]). The HPI platforms currently do not support virtual tutors, and this topic is not a part of the thesis. A valuable alternative to virtual tutors in MOOCs are peer tutors. Rosé et al. [97], present a tool called Quick Helper, which matches participants that are struggling with a question to other participants who have been identified as possibly being able to help. Teusner, Hille, and Staubitz [25], follow an analogous approach with requests for comments in their auto-grader for programming exercises. While the system prompts struggling participants to ask for help, it asks other participants who have already successfully solved an exercise to support those that are still struggling with hints.

3.2.2 Learning in Small Groups

While collaborative learning generally is quite well researched, collaborative work in MOOCS, currently, is still under-researched. The available literature either proposes concepts how it could be approached, reports about the instructors' experience in a single course, or proposes technical solutions for implementation details, which often—at best—have been tested in simulations or laboratory settings.

In 2014, McKinsey [98], reported about the experience with pair programming sessions in Berkeley's *Engineering Software as a Service* MOOC on EdX. Cambre et al. [99], reported about several MOOCs in which they have employed TalkAbout, a tool that matches participants at a certain time slot to join a video chat and discuss a given issue. The students enjoyed discussing in diverse groups, many of them spent twice as much time in these discussions as required. Lim et al. [100], developed a tool to form small ad hoc discussion groups in MOOCs. The tool "organizes students into groups, allowing them to first individually answer questions and then see each others' answers and discuss those answers, while a timer counts down. When the time is up, the students choose their final answer." In 2015, Zheng et al. [101], presented a design for MOOCs as interactive collaborative learning spaces. Their contribution, are four scenarios how they imagine collaboration in MOOCs could happen. Even a prototype to test their ideas is still in their list of future work. Zheng, Vogelsang, and Pinkwart [47], conducted an experiment in a MOOC on Crystallography, where they tested a tool to automatically create teams. The platform did not come with any team support, so, they had to use external tools such as facebook or Google plus. They observed that there was close to no interaction or discussion within these groups. In the beginning of 2016, several California community colleges announced to bring teamwork to their online classes. Bazaar[1], 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 [102]. In 2017, Cheng et al. [103], presented ProjectLens, a tool that allows to match teams and provide them with a communication channel. Their paper yields no evidence that the tool also supports the assessment of the team assignment. So far, they have conducted one case study with eighty-eight students who enrolled in a project-based course on UI design, offered by the University of Minnesota [104].

3.2.3 Team Assignments in MOOCs

NovoEd is one of the few major MOOC platforms supporting 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 [105]. NovoEd is considered to be state-of-the-art in terms of teamwork in MOOCs. In 2014, Manning [106], compared NovoEd, Coursera, and OpenEdX and stated that (of these platforms) only NovoEd supports gradable team based assignments. Sankaranarayanan et al. [107], state that there is “a shortage of affordances for collaborative learning experiences in [MOOC] platforms” and that NovoEd is the one notable exception. NovoEd’s website states: “NovoEd provides the best online learning platform for social, collaborative [...] learning [...] at scale.”[2] In 2014, a course on Human Centered Design (HCD) was offered on that platform, but, according to Cheng et al. [104], little is known about the success of the teams in this course. Wen, Yang, and Rosé [108], report about their experience in two courses on NovoEd in 2014. The courses had a size of round and about 2000 participants. Of those, about 200 per course joined a team. About half of the teams, successfully submitted the final team project. Recently, the author, in a self-experiment, participated in NovoEd’s course “Learning Experience Design (LXD).” The platform follows a laissez-faire approach for team building[3]. Participants are asked to either start or join a team. The partici-

1. http://www.cs.cmu.edu/~cprose/Bazaar.html
2. meta name="description" on view-source:https://www.novoed.com/
3. Different approaches to team building will be discussed in one of the following s

pant who starts a team, automatically, is the team leader. She can then invite classmates to join the team, based on their profile information [108]. Except for ongoing encouragement by a course facilitator and options to find other participants with corresponding profiles, there is no platform support to match teams. The platform offers a team forum and private messaging to support the communication within the teams. Next to formative assessment in the form of comments by instructors and peers, there was a binary automated assessment for both the teamwork and the individual tasks—submitted/not submitted.

In 2017, Ju, Glassman, and Fox [109], presented Teamscope, a tool that analyzes and visualizes the team members GitHub[1] and Slack[2] data. Teamscope enables instructors to peek into the teams and get an intuition if everything is fine or something is going wrong within a team. It has been employed in an on-campus software engineering class at Berkeley. In 2018, Ju et al. [110], addressed the "scalability issue of software engineering projects" and laid out a blueprint for software engineering team projects in MOOCs.

3.2.4 Team Building

The terms *team building* and *team matching* are used from here to delineate the process of finding and matching the right members for a team. The term *team forming* is used to delineate the process of transforming a group of individuals into a well performing team. Basically, there are two different approaches to build teams in an educational context:

- Teams are built by the learners
- Teams are built by the instructors

Kizilcec [111], calls these approaches *laissez-faire* and *interventionist* and this wording is adapted for this thesis. Instructors might want to form local teams to alleviate the *team forming* process, or they might want to form culturally diverse (and therefore probably distributed or virtual) teams to achieve a certain learning goal. Kizilcec [111], examined differences between geographically distributed groups that were forced to rely on computer mediated communication and local groups that were able to meet face-to-face. While his sources state that the face-to-face-groups had a better learning experience concerning the group discussions, they did not find significant differences in learning itself when the students were tested before and after the teamwork [111].

1. GitHub is a popular platform hosting an online repository and version control tool. https://github.com
2. Slack is a popular team chat. https://slack.com

Instructors might form teams randomly or based on certain homogeneous or heterogeneous criteria. Naturally, there might also be teams with a heterogeneous mix in one criterion and a homogeneous mix in another. Whether the instructors pick a random approach or a criterion-based approach, depends on the number of learners that have to be "teamed" and the tools that are available to the instructors to support them in matching the team members. If no such tool is available, the laissez-faire approach is often the only choice. However, there are some arguments against it:

- Albeit teams that consist of friends seem to perform better on tasks with high quantity output, this is not the case for tasks with high quality output. Teams of strangers are stronger in "constructive disagreement" [112].
- Teams that consist of some members who already know each other and others who do not, 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 [113].
- 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, generally, do not know each other at all, the laissez-faire approach is hard to manage for the participants.

Kizilcec [111], confirms the need for other ways to build teams than random selection or self-organization. Shimazoe and Aldrich also discourage the use of self-selection in the team building process [114]. Layton et al. [115], compare pros and cons of self-selected, randomly assigned, and instructor-assigned teams in depth. They conclude that self-selected teams come with the most disadvantages. Random assignment has no particular strengths, while it at least avoids "some of the negative effects of self-selection" [115].

Wen [116], proposes an approach to tackle the issue of choosing team members out of a large crowd of strangers. Students start with the submission of an individually produced artifact in a course wide forum. The submitted artifacts are discussed by the course participants. 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 this way produce a better learning outcome than those that are randomly formed [116].

Agrawal, Golshan and Terzi [117], identified "partitioning of students into non-overlapping teams that also maximize the benefit of the participating students" as to be NP-complete[1] and propose an approximate algorithm for solving it. Zheng and Pinkwart [118], propose a matching algorithm based on discrete swarm optimization (DPSO). It starts with randomly-generated teams and allows to dynamically re-compose teams, based on their performance in the tasks and the students' satisfaction with the composition of their previous teams [47]. In an experiment, conducted on the German MOOC platform Iversity, they built teams based on motivational surveys and the learners' preferred way of communication tools. From task to task, the teams have been reassembled, attempting to predict the team with the best possible performance for each student [47]. Roepke et al. [119], have developed a Moodle plugin[2] that employs a long, sophisticated list of parameters, including learning styles, personality traits, or previous knowledge, to match the optimal team members. *MoodlePeers* is based on the *GroupAL* algorithm [120], which takes a list of vectors—each representing a participant—and forms teams of a given size. Er et al. [121], attempt to predict a student's possible commitment in upcoming collaborative activities based on her previous performance. They have built a predictive model, based on features, such as number of discussion posts, quiz attempts, quiz scores, time spent on quizzes, page views, etc. They match the teams based on these features and aim to predict how many team members will actively participate in the group discussions. So far, they have only applied their approach to a course translating business terms between Spanish and English. The paper states that the predictions were correct, but does not discuss if the approach has been conducted during the runtime of the course and actual teams have been built, or if just the course data have been used to verify the approach in a simulation. If real life teams have been formed, the paper does not discuss the success of the teams [121].

The *CatMe Team-Maker*[3], which was developed at Purdue university, is probably one of the most mature team-matching systems for on-campus courses on the market. When the students sign up for teamwork, they have to complete a detailed survey, including questions about their schedule, gender, race/ethnicity, GPA, prerequisite courses, skills, leadership preferences, etc. The Team-Maker then builds the teams based on this data. Instructors have many options to define how they prefer the matching to be done.

Belbin [122], identified nine archetypical team roles. According to his theory, teams perform best if preferably all these roles are present. He developed a test to determine the team role of employees or students. Teams can then be

1. NP-complete problems are defined as not-solvable within a reasonable amount of time by current computers.
2. https://github.com/moodlepeers/
3. https://info.catme.org/research/

built according to the test results [122].

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 time frame 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, usually, is not an option. Therefore, we can neither rely on lengthy processes as the one delineated 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. The Team-Maker lacks certain features owed to its focus on on-campus teams. For example, it matches members based on their schedules, while we need a less fine-grained matching mechanism, based on the timezones, because of our international audience.

3.2.5 Team Forming

Tuckman and Jensen [123; 124], defined the process of transforming a group of individuals into a performing team as a sequence of the five stages: *forming*, *storming*, *norming*, *performing*, and *adjourning*. Building on Tuckman's model, Rickards and Moger [125], investigated what happens if the storming stage of a team just does not end and what enables teams to excel in the performing stage. They identified a "weak behavioral barrier" between the storming and norming stages, which needs to be crossed in order not to fail as a team, and a "strong performance barrier" between just plain performing and outperforming. They studied MBA students who received creativity training for two years. In the second year, the number of teams that passed the barriers, significantly increased when the teams received improved support [125]. Gersick [126], examined several teams and came to conclusions differing from Tuckman and Jensen's. Instead of consecutive stages of similar behavior, she observed congruities in timing in terms of phases of activity and inertia between the groups [126]. Built on these observations she built a model that she termed "punctuated equilibrium." While the other authors examined local teams, Johnson et al. [127] examined the team forming process of distributed teams in an online master program. They noticed that the distributed teams formed analogous to local teams. The performance of the teams was strongly dependent on the ability of the teams to establish procedures and solve conflicts [127].

3.2.6 Team Grading

Finally, when the teams have been built, when they have formed and performed, we need a way to assess or grade the results of the participants and to provide them with feedback. In this context, it is of particular interest if the grade will be a collective grade, an individual grade, or a grade composed of both elements. Furthermore, it is of interest if the grade is received for a single task or for a sequence of tasks. Another question is if the result should be graded, or rather the team-process itself [28]. Finally, particularly in the context of large scale learner bodies, such as e.g., MOOCs, it is of interest how the grades will be determined. Individual grading by the instructors is not feasible here. Depending on the task, automated grading or peer grading are suitable candidates.

Carnegie Mellon's *Eberly Center for Teaching Excellence* suggests to compose grades based on several components, including the team's final product but also the team processes and the functioning of the team. They also recommend to translate the team's overall performance into individual grades [128].

Kennedy [129], conducted a study among ninety students to examine if peer assessment within a group is a valid instrument to determine individual grades for the team members. Kennedy differentiates between the team's outcome and the team process to arrive there. While the outcome was graded by the teacher, the students graded the contributions of each member during the process. The grades given by the students varied largely. While some students were reluctant to give lower grades to teammates who contributed less, others were less troubled to do so. He observed that the requirement to grade each other created a certain amount of frictions within the teams [129]. This can become particularly problematic if the team members also share some aspects of life outside their team. Participants might give better grades to bow to "peer-group pressure or as a consequence of friendship" [130]. Willcoxson states that teamwork often lays a burden on higher-achieving students, as they perform less well in teams, while lower-achieving students often perform better[1].

Nepal [131], states that a team grade often does not reflect individual effort and, therefore, proposes to add individual weighting factors for each team members. Hughes and Jones [29], point out that teamwork is not the same as team success. They state that the outcome of a team is not necessarily an evidence that the members worked particularly well together:

> A team might be successful because one member made uniquely

1. Willcoxson does not indicate which are the reasons for this phenomenon. It could be that teamwork is leveling out the results, tearing the high-achievers down to a similar amount as the low-achievers are lifted up. On the other end of the spectrum, it might be that the low-achievers just fit themselves better in and perform better in teams.

> important contributions that ensured a quality product despite marginal efforts by most team members; or a team might be successful because it was operating in a particularly munificent environment virtually guaranteeing a successful outcome[...]

Willcoxson, as well as Hughes and Jones, therefore, recommend to rather assess the teamwork process than the outcome.

Willcoxson [130], introduced a paper based assessment framework that consists of four phases. Before the team assignment starts, all participants individually have to reflect what was good or bad in previous teams or groups and what they would want to avoid or repeat. In the next step, the team sets up the rules for the project work, then a planning and attachment sheet are developed. In the attachment sheet, the participants note the discrepancies between what was planned and what actually happened—and why.

Next to paper tools, there are several digital tools to support the grading of teamwork in university courses. Examples are *Expertiza*, a peer assessment tool developed by Gehringer at North Carolina State University. Expertiza offers the option to peer-grade the work of teams. Teams jointly submit their solutions, while grading the work of the other teams and their teammates, is done individually [132]. *Catme*, developed at Purdue university offers a whole suite of tools to assist instructors offering teamwork in their courses. Next to the previously mentioned matching tool "Team-Maker", it also offers a solution to assess the work of the teams and their processes [133].

4. Collaborative Learning in Small Groups

While in many courses, a general course wide discussion forum is sufficient to cover the collaboration needs of instructors and course participants, others require more dedicated levels of collaboration, such as options to create smaller units of participants within the learning community of the course. The HPI platforms support two different types of such smaller units: *groups* and *teams*.

4.1 Groups vs. Teams

We use the term *groups* for loosely coupled participants, while we use the term *teams* for more tightly coupled participants. *Groups* are formed by the participants themselves. They are loose conglomerates of participants that join forces based on a common interest or preference. *Teams* are typically created by the instructors, using a tool that allows a semi-automated matching of participants based on instructor-defined criteria.

4.1.1 Groups—Loosely Coupled

We have developed the possibility to create subgroups within the overall learning community very early on. The platform's Collab Space feature supports the participants to create such subgroups and provides them with a private discussion forum, a set of online collaboration tools, a possibility to share files, and a video chat.

Special interest groups

Originally, the participants were enabled to create any number of Collab Spaces and invite friends or peers with similar interests or needs. The created Collab Spaces can be either open (for everybody to join) or invite-only. Use cases that we envisioned, included:

- enabling participants to create a space where they can talk in their native tongue if it differs from the course language,

- enabling shy participants to create a more cozy environment where they can discuss in a smaller group,
- enabling participants who already know each other from a different context, e.g., the same school or company to create a more private environment,
- enabling special interest groups to dig deeper into a certain course topic or to extend the course by collaboratively exploring topics beyond the official course scope.

However, the participants do not make use of it to the extent that we have expected. Actually, it only worked well in very few examples:

- a German speaking group in the course *intsec2018* which was offered in English language,
- a group of Debian[1] users in *linux2018*, while the course focused on the Ubuntu distribution[2],
- a group of teachers who shared and discussed hands-on experiences in a course about the Calliope micro-controller[3].

But these examples are rare. We asked the participants about the Collab Spaces in several post-course surveys.

Less than 5% of the participants have used the Collab Spaces. The usefulness of the Collab Spaces was perceived rather indifferently. Usually, the Collab Spaces have been created, sometimes other participants joined, and then nothing was happening there.

Based on these results and the general lack of interest or proper use of the feature, it would have been a consequent step to remove the feature and focus on more successful elements of the platform. The following section discusses why we took a different approach.

4.1.2 Teams—Tightly coupled

The main pillar of our definition for tightly coupled small group collaboration is a common graded project that has to be solved collaboratively or cooperatively by a small group of participants. We use the terms *collaboration* and *cooperation* according to the following definitions by Roschelle and Teasley [134]:

1. A popular Linux distribution
2. Another popular Linux distribution
3. The Calliope mini has been developed to ease the step into the digital world for kids in grade 3-6. The course targeted teachers and introduced the possibilities and options of the tool.

> Collaboration is a coordinated, synchronous activity that is the result of a continued attempt to construct and maintain a shared conception of a problem.

and Power [135]:

> Cooperation is accomplished by the division of labor among participants as an activity where each person is responsible for solving a portion of the problem.

The decision whether the to collaborate or to cooperate to get the task solved, is left to the teams. To provide a flexible and scalable grading mechanism for these projects, the platform's peer grading tool has been extended to allow *team peer assessments*. In a *team peer assessment* one of the team members hands in the team's solution for all other members. Each team member then, individually, has to review and evaluate the work of 3-5 other teams. Writing reviews is mandatory. As a side-effect it filters inactive team members, preventing them from benefitting of the work of active members.

Graded team assignments in MOOCs come with several challenges. Some of these challenges are of a more general nature and concern team assignments in all contexts, others are specific in the context of MOOCs. A general aspect is e.g., the grading of the submitted work. Will each team member receive the same grade for the submitted work? Can (and should) individual components be added to the overall grade? Another general aspect is the composition of

Table 4.1: Number of threads in group, team, and course wide forums.
C-CS: # of courses with Collab Spaces (CS)
C-T: # of courses with team CS (percentage of courses with CS)
T-GCS: # of threads in group CS (percentage of total)
T-TCS: # of threads in team CS (percentage of total)
T-C: # of threads in course wide forums (percentage of total)
Total: # of threads in discussion forums

	C-CS	C-T	T-GCS	T-TCS	T-C	Total
openSAP	145	12 (8.3%)	924 (1.9%)	6050 (12.7%)	40806 (85.4%)	47780 (100%)
openHPI	64	4 (6.3%)	1232 (6%)	1096 (5.3%)	18293 (88.7%)	20621 (100%)

the teams, which matching criteria are important to leverage successful teams?

Which strategies are successful to create successful teams with low dropout rates? These questions will be addressed in Chapter 5. One of the indicators that the approach to shift the focus from *groups* to *teams* was successful, can easily be observed in the forum activity. Table 4.1 compares the forum activity in teams, groups, and the general course forum. The table lists only those courses that have any type of forum activity—groups or teams—in the Collab Spaces[1]. On both platforms, there is only a very low activity in the group Collab Space forums (1.9% (openSAP)/6% (openHPI)) of the total forum activity). Considering the fact that only a small subset of courses on both platforms actually feature team tasks, the activity in the team Collab Space forums is significantly higher. The average amount of group threads per course are 6.4 (openSAP) and 19.3 (openHPI). The average amount of team threads per course are 504.2 (openSAP) and 274.0 (openHPI).

1. There are many reasons why a course does not have any Collab Space forum activity. It might be a very old course that has been delivered before the move to the new platform, it might be an upcoming course that is not published yet, it might be a course where the Collab Spaces have been disabled, etc.

5. Gradable Team Assignments

Gradable project-based assignments allow instructors to deliver courses that are based on the model of active learning. Enabling teams to work collaboratively on these assignments takes the concept to the next level as it includes a tighter social learning component than the more loose forms of social interaction, such as collective forum discussions or special interest groups. The tool-set that has been introduced in Section 1.4 provides us with the technology to include these scalable and gradable project-based assignments for both individuals and teams; and thus to develop a completely different type of MOOCs that leave the traditional frontal teaching style behind. The big question now is, "Does it work?" Only very few aspects of this question have a technical nature. So, more accurately formulated the question rather is "Do the participants accept this type of courses and assessments?" In the following, the prototypical implementations and experiments that have been conducted on the HPI platforms will be discussed in detail.

Throughout this chapter the terms *participants* for course participants, *team workers* for participants in the team assignment, *free-riders* for inactive participants in active teams, and *solos* for participants who decided to solve the given task alone and not in a team (*javawork2017* and *java-capstone-1* allowed the participants to decide, whether they want to work on the task in a team or alone) will be used. Participants who never have visited a single item in the course are called *no-shows*. No-shows and late enrollments have been removed from the completion rate equation as they distort the results. The research conducted in this context, consists of several analyses of the course and learning analytics data, surveys and interviews. Based on the results of these efforts, the platform features have been revised and updated. Figure 5.1 provides an overview on the development of the platform features, courses, surveys, data analysis, and interviews, as well as the publications in this context.

In 2016, the course "Design Thinking for Software Developers (SAP internal)"(*dt1-pilot3* was the first course that offered a team-based task. As the peer assessment feature did not yet support the option to grade the work of teams, the assignments in this course still have been graded manually by the team's mentors. This was only possible as the course was internal and only had a handful participants. A few weeks later, "Enabling Entrepreneurs to Shape a Better World" (*sbw1*) followed as the first course on the HPI platforms

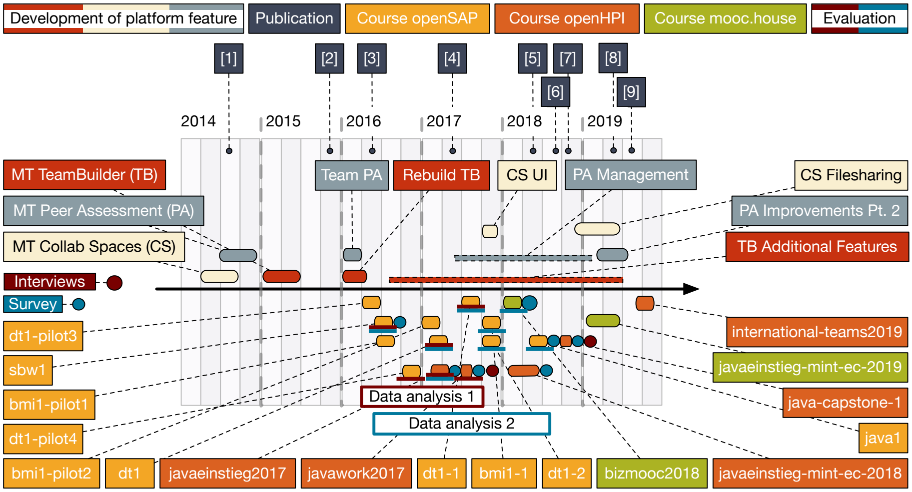

Figure 5.1: Overview courses and steps of study.
Publications:
1: Supporting social interaction and collaboration on an xMOOC platform [12]
2: Collaborative learning in a MOOC environment [13]
3: Improving the peer assessment experience on MOOC platforms [48]
4: Collaboration and teamwork on a MOOC platform: A toolset [136]
5: Team based assignments in MOOCs: Results and observations [137]
6: Team-Based Assignments in MOOCs - User Feedback [138]
7: Collaborative Learning in MOOCs - Approaches and Experiments [39]
8: MOOCs in Secondary Education - Experiments and Observations from German Classrooms [139]
9: Graded Team Assignments in MOOCs - Effects of Team Composition and Further Factors on Team Dropout Rates and Performance [140]

containing a full-fledged scalable and gradable team assignment. *sbw1* was accompanied by a survey and, additionally, user feedback was collected in the form of "I like... I wish..."[1] posts in the course discussion forum. Based on the results of the survey, the participants' feedback, and the analysis of the collected interaction data, the initial prototype was refined. For example, an introductory week has been added to the 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 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

1. 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.

was limited by the availability of mentors.

In 2017, the first course that included an unmoderated, short (two-weeks), graded team task (*javaeinstieg2017*) was delivered. At the end of the course, a survey has been conducted among the team-task participants. Furthermore, fourteen of the team-task participants have been interviewed for about one hour each. The Collab Spaces have then been refined, based on the results of the survey and the interviews. For example, a video was added to explain the possibilities of the Collab Spaces, several menu items were renamed and features that have not been used and rather confused the participants were removed. These improvements were tested with a very small set of participants in an offline workshop, some further adjustments were made and the next large-scale experiments were run in the courses *bizmooc2018* and *java1*). The surveys for these courses have been overhauled, and a new set of 18 one-hour interviews in the *java1* course was conducted.

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 assignment is a task that would also be solved by teams in a real work environment.

The focus of the tasks was less on the team-processes than on the outcome of the teams. The outcomes had to be submitted to the platform and have been peer-assessed by members of the other teams. Additionally, the team members were able to rate the work of their teammates and, thereby, to add an individual factor to the overall grade. Detailed descriptions for the tasks to be solved and the settings of the teams and the courses are available in Appendix I - Courses. The list of interviewees and a link to the transcript of the interviews is available in Appendix II - Interview Summaries.

5.1 Data and Methodology

The study accompanied the courses on the HPI platforms that featured graded team assignments. It consists of several, sometimes iterating, steps. The collected course data was analyzed. Prototypical implementations and experiments have been accompanied by surveys and interviews. Significant parts of this study have been previously published in [137–140] and are reused with permission. In the following, each individual part of the study will be delineated in some detail. Throughout the rest of the section, however, the results of these parts are combined and ordered by aspects and questions rather than chronologically. Where necessary, references to the course or part of the study will be made in context.

5.1.1 Platform Data Analysis

First iteration

The first step of the study was conducted in late 2017/early 2018 (published in [137]). The data have been collected from four courses on openSAP and two courses on openHPI. On openSAP the courses "Enabling Entrepreneurs to Shape a Better World" (*sbw1*) and several iterations of the course "Design Thinking for Software Developers" (*dt1-pilot4, dt1, dt1-1*) were examined. On openHPI the courses "An Introduction to Object-Oriented Programming in Java" (*javaeinstieg2017*) and "An Introduction to a Programming IDE" (*java-work2017*) were examined. A detailed description of the involved courses is available in Appendix I - Courses, where they are tagged as *sgc-team1*. This step of the study attempts to answer the questions:

- Can teamwork reduce the attrition rate of MOOCs?
- Which role does extrinsic motivation play in terms of completing the assignment?
- Is there a higher probability that local teams will complete the assignment?
- Can we predict by the previous weekly scores who will complete the assignment and who will not?
- When is the best time to match the teams?

Second iteration

In late 2018/early 2019, the second iteration of the course data analysis has been conducted (published in [140]). At that point, fifteen courses containing team assignments had been delivered on the platform. A dataset containing the anonymized course data, team registration data, peer assessment data, and user profile data for all of these courses has been created in order to obtain a holistic view. The dataset in total contains 846 teams. During the analysis, it became evident that some courses deviate substantially from the standard courses—e.g., the pilots and the workshops. The collected data from these courses, often rather distorts the picture and leads to conclusions that are more likely the result of anything but the examined feature. The dataset, therefore, has been reduced. In the end, only eight of the fifteen courses were directly comparable. All eight examined courses have been completely open to the public and their size ranged from about 2,000 to 20,000 enrolled participants. The remaining seven courses have been removed from the analysis as they introduced factors that were not directly related to the team assignment itself. Four of these deselected courses have been pilots, open to a selection of explicitly invited participants only. One of them has been offered to schools to be integrated in their regular computer science classes or as an extracurricular activity. Two of them have been offered in an experimental format that differed significantly from the other courses. One commonality of the deselected courses is that they have been comparably small and hosted only a few teams. Some of the removed courses, however, still have been examined in particular contexts within this study. For similar reasons, the dataset has been reduced even further in a few cases. Wherever it becomes relevant, this will be explained in more detail in context. The eight selected courses 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
   (English language, *sbw1*, 39 teams)
   Business Models for the Digital Economy
   (English language, *bmi1-1*, 49 teams)
   Intrapreneurship - Make your Business great again :-)
   (*bizmooc2018*, 28 teams),
3. Design Thinking[1] (DT):
   Developing Software using Design Thinking (Three iterations)
   (English language, *dt1*, 62 teams; *dt1-1*, 66 teams; *dt1-2*, 48 teams)

The selected courses contain 703 teams (371 in the OOP category, 156 in the BI category, and 176 in the DT category). A detailed description of the involved courses is available in Appendix I - Courses, where they are tagged as *sgc-team3*. The remaining seven courses that have not been examined for this part of the study are tagged as *sgc-team3-not*. This step of the study, attempts to answer the following questions:

- Which are the differences (if any) between the total course population and the subset of this population that is registering for a team task?
- Which constellations in the composition of teams have particularly positive or negative effects on the teams' performance or dropout rates?

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 [141] [142].

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

The relevant data sources of the platform—course reports, peer assessment reports, and TeamBuilder reports—have been merged and serve as the basis for the following three datasets:

1. A dataset aggregated on team level, for all 846 teams in the examined courses. A detailed description of this dataset is available in Appendix IV - Aggregated Team Data.
2. A complete dataset on the user level for all 6246 team members. A detailed description of this dataset is available in Appendix V - Team Member Data
3. A complete dataset on the user level for all participants who have not participated in a team task in the fifteen examined courses. A detailed description of this dataset is available in Appendix VI - Non-Team Participant Data

Differences and commonalities between the total course population and the team-task participants have been identified. The effects of certain team compositions on the teams' performance and dropout rates have been analyzed. Finally, the effects of some platform modifications on team performance and team dropout rates have been evaluated. To verify the validity of the conclusions, the group-wise distribution of observations has been examined. In most cases, further investigations have been stopped if not all groups contained a similar amount of observations. Furthermore, each of the examined variables has been double-checked 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 on closer examination the significance of the results vanished into thin air.

5.1.2 Surveys

In 2016, a post-course survey, targeting the participants who have been working in teams, was conducted in *sbw1*. In early 2018, a set of surveys—conducted among the participants of *javaeinstieg2017*, the follow-up workshop *javawork2017*, and *bizmooc2018*—have been evaluated (published in [138]). These surveys aim to procure more insights about the participants' view on the team-building process, mentoring, and the tools provided in the Collab Spaces. In *javaeinstieg2017* and *bizmooc2018* an almost identical post-teamwork survey was provided that only addressed those users who participated in the team-task. In *javawork2017* a few team-related questions have been added to the regular end-of-course survey. Particularly, it was attempted to answer the following questions:

- Are the provided communication and collaboration tools sufficient?
- Are the participants fit to find and employ alternative communication and collaboration tools?
- Are there tools that are not used at all?
- Would the participants prefer to be teamed or find teammates on their own?

In *javaeinstieg2017*, 340 users (22% of the team task participants) submitted the survey. In *javawork2017*, 224 users submitted the survey, 23 of them were team-task participants (58% of all team-task participants). In *bizmooc2018*, 42 users (26% of the team-task participants) submitted the survey.

Later in 2018, an overhauled survey has been conducted among the team-task participants of *java1*. The results of this survey have not been published so far. Additionally, some data have been collected in an extended version of the introductory Java course that is used in schools *(javaeinstieg-mint-ec-2018)*. Furthermore, a survey regarding the team task in this course has been conducted among the participants. A detailed description of the involved courses is available in the list of courses in Appendix I - Courses, tagged as *sgc-team2*. Furthermore, the original survey questions and their answer-options are available in Appendix III - Surveys.

5.1.3 Interviews

In 2017, fourteen team-task participants of the course *javaeinstieg2017* have been interviewed for about one hour each. During some of these interviews, the interviewees have been asked to demonstrate how they have solved certain tasks in their team. These efforts have been observed by the author and a colleague[1] to examine the workflows and detect usability issues on the platform. The fourteen volunteers closely cover the variety of the participants that have joined the team task. Eleven male and three female team members, three pupils 17-19 years old, otherwise mostly professionals, many with a bachelor, master, or doctorate. Four of the interviews have been conducted face-to-face, ten via video chat. Six of the participants have a computer science or engineering background, three participants have a business and management background. The remaining come from media, creative design, life sciences or did not share this information. All interviewees had German nationality, as the course was

1. Hanadi Traifeh from the HPI-Stanford Design Thinking Research Program (DTRP)

offered in German language. Half of them had an urban, the other half a rural or small town background. The fourteen interviewees represent thirteen teams as one team was represented twice. Seven of the represented teams ended up with half of the team members remaining, three had more than half, two less than a half, and two ended up dysfunctional with just one member left. Five interviewees contributed strongly to the course discussion forum, the remaining nine contributed only a little. Thirteen interviewees completed the course among the top 5% participants. The group of interviewees has a strong bias towards successful course participants, however, as will be shown in the remainder of this chapter, this is very common for the participants of the team tasks. Therefore, they can be considered to be highly representative for this subset of the course population. In 2018, a second set of one-hour interviews has been conducted with eighteen team-task participants of *java1*[1]. Thirteen male and five female team members have been interviewed. Fourteen of the interviewees have been between forty and sixty years old, mostly professionals with a computer science or engineering background, three of them had a business and management background. Six of them have a bachelor degree, two have a master degree and two have a high school diploma. The other eight have not shared this information. All interviews have been conducted via video chat except for one that failed due to technical reasons and has been sent per mail. This course was offered in English language, and therefore, had a more international audience. Six interviewees originate from different countries in South America, nine from Europe, and three from India. The eighteen interviewees represent fifteen teams as some teams are represented double. Eleven (including all teams the are represented twice) of the represented teams finished the task with more than half of the team members remaining, three teams had half or slightly less than half members remaining, one of them was dysfunctional. Thirteen interviewees contributed highly to the course discussion forum, the remaining five contributed only a little. Seventeen interviewees completed the course within the top 5% participants another one was in the top 20%. Again, the group of interviewees has a strong bias towards successful course participants. Nevertheless, they can be considered to be highly representative for this subset of the course population. Appendix II - Interview Summaries provides an overview about the pseudonymized background profile of the interviewees. A structured summary of all interviews is available as an Excel file. The link to this file can also be found in Appendix II - Interview Summaries. The summaries of

1. Again by the author and Hanadi Traifeh. Both sets of interviews also covered the follow-up workshops *javawork2017* respectively *java-capstone-1*. An attempt to do the same in *bizmooc2018* failed owed to a lack of volunteers.

these interviews have been processed with a tool called MaxQDA[1]. First, the author defined a set of categories to structure the participants' answers to the given open-ended questions:

- Tools used — Platform and third party tools that have been used by the participants, while working on the given task. Sub-codings exist for each single tool that has been explicitly mentioned in this context during the interviews.
- Tools unused — Explicit mentions of platform tools that have **not** been used by the participants, while working on the given task. Sub-codings exist for each single tool that has been explicitly mentioned in this context during the interviews.
- Preferences — How to match the teams, how to collaborate, when to create the teams, how to activate inactive team members, contact after course end, future plans for team tasks.
- Suggestions — Additional tools, different approaches, intro video, etc.
- Initiative — Who took the initiative, the interviewee or other team members.
- Good learning experience — Several subtopics, will be explained in more detail in Section 5.12.
- Suboptimal learning experience — Several subtopics, will be explained in more detail in Section 5.12.

Then, all answers have been tagged with these codings and the data have been analyzed (see more in Section 5.12).

5.2 User Satisfaction

User satisfaction and learning experience are key in a continuous-learning environment. Nobody forces the learners to complete a course. There are absolutely no consequences (except for not achieving a certificate of questionable value[2]) for the learners if they drop out. So, our approach in the design and development of the platform has to be learner centric and our aim is to make the platform as comfortable for the participants as possible. This does not mean,

1. https://www.maxqda.com/what-is-maxqda
2. Albeit there seems to be a tendency that the certificates are used more often in job applications, recently; they do not have any official accreditation. The Hasso Plattner Institute recommends to value the Qualified Certificate with two ECTS credits. Qualified certificates, however, are offered only in a few selected courses. None of them plays a role in the context of this thesis.

however, that any compromises concerning the quality or the difficulty-level of courses are made.

5.2.1 General Acceptance of Teamwork

At the end of *sbw1*, the first course featuring a team-based assignment on the HPI platforms, a survey was conducted to determine if the learners liked to work in teams and to get to know their opinion about the learning outcomes. 824 participants had applied for the teamwork, 267 have been admitted. Out of those, 57 participants answered the survey. Generally, the participants have been satisfied with the team assignment. The majority considered the task relevant, manageable, and suitable for virtual teamwork and enjoyed working in a team. Figure 5.2 shows the results of this survey in detail. The survey also

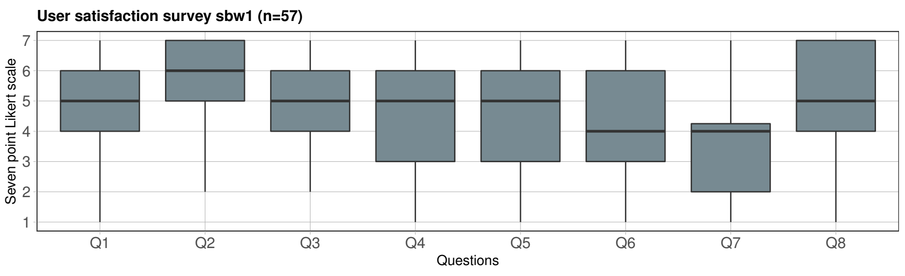

Figure 5.2: User satisfaction - post-teamwork survey *sbw1* (n=57)
All answers are on a 7-point likert scale (1-I do not agree, 7-I fully agree).
Q1. Generally, how satisfied have you been with the Team Challenge?
Q2. The Team Challenge tasks were relevant.
Q3. The Team Challenge tasks were manageable.
Q4. The Team Challenge tasks were suitable for virtual teamwork.
Q5. I enjoyed working in my team.
Q6. The support our team received from our mentor was valuable for our project.
Q7. The time the mentor spent with our team was sufficient.
Q8. I was able to acquire important know-how through the Team Challenge tasks.

contained a "What did you like, what should be changed" section providing some more qualitative feedback. On the positive side, the participants mainly mentioned that they liked to work in international teams with people from different educational and cultural backgrounds. Other positive statements mentioned the possibility to work on a relevant task in a team:

> It was an amazing way to practice the material and gain a lot of insights about myself and the challenges on the way to become an entrepreneur.

The survey also revealed, however, that the mentoring-related questions received less positive feedback. Two regularly mentioned issues that were requested to be changed were more time to solve the tasks and the relation to the mentors.

> Mentors should have good role to play. Initially we have great problem with mentor.
>
> Also, I thing Mentors should be chosen carefully[1].

Another issue that has been mentioned quite regularly, are tensions and frustrations resulting from different time commitment of the team members and dropouts in the teams. We have addressed the first issue by adding the time commitment as a matching criterion to the *TeamBuilder*. Our approach to predict the dropouts among teamwork applicants based on the results of their previous quizzes and assignments will be discussed in Section 5.6, our approach to team matching will be discussed in Section 5.7.

Although small in number of participants, *javaeinstieg-mint-ec-2018* provides an interesting glimpse into a completely different target group. While the participants in most of the other courses are lifelong learners, 30–50 years old, mostly with at least a Bachelor's degree and working in their job since five or more years, the majority of the participants in this course have been pupils 16–20 years old. We asked them if the course in total, and the team task in particular, enriched their school class. Almost 60% of the participants perceived the course in general as enriching, another 20% found it "so/so", 20% did not like it. Still, as much as 30% also perceived the team task positively, 40% perceived it as neutral, and 30% perceived it negatively. Given that the workload in the team tasks is often comparably high and teamwork itself is not always everybody's darling, this value is surprisingly high.

5.2.2 Usefulness of the Provided Tool-Set

In the surveys that have been conducted in *javaeinstieg2017*, *javawork2017*, and *bizmooc2018*, the participants have been particularly asked about their satisfaction with the provided tool-set. The Collab Spaces are offering a discussion forum for asynchronous communication, Google Hangouts for video chats, a chat tool within the Etherpad, and before *bizmooc2018*: TogetherJS[2]. The surveys attempted to show how the communication and collaboration tools

1. Spelling and grammar as in original answers.
2. TogetherJS allows participants to synchronize the currently active web-application (One clicks, all surf). It supports also to display the mouse pointer of the other participants, and it also contains a text chat. (https://togetherjs.com)

in the Collab Spaces have been used and which further tools are requested. Next to figuring out if the tool support of the platform is sufficient, the surveys investigated if the participants can manage to find more suitable tools on their own if necessary.

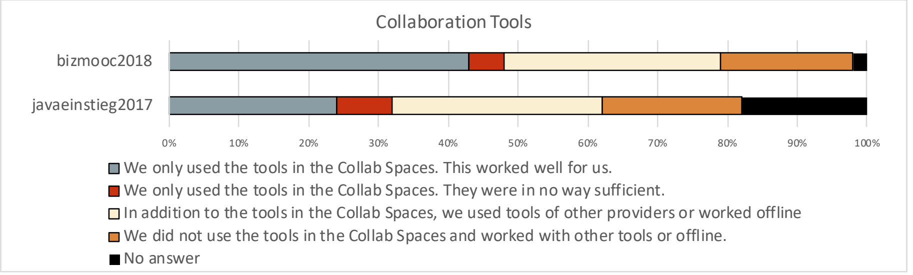

Figure 5.3: Post-teamwork surveys *javaeinstieg2017* (n=340), *bizmooc2018* (n=42) Have the provided collaboration tools (Etherpad, Tele-Board) been sufficient to solve the task?

Collaboration tools

The majority of the participants managed to complete their task with the tools offered by the platform or by combining them with additional online or offline tools. About 20% used only third party tools and did not use the tools provided by the platform at all. 5-15% of the participants did not manage to use additional collaboration tools even if they felt that the provided tools are not sufficient to solve the task. Fig. 5.3 shows the survey results in the courses *javaeinstieg2017* and *bizmooc2018*.

The most asked for tool in *javaeinstieg2017* was an online UML diagram tool[1]. Other tools that were missed are a more sophisticated word processor and other office tools. The Etherpad's possibilities to format text have been perceived to be too limited even for producing a simple glossary. Fig. 5.4 shows the result of a survey that has been conducted among the team task participants of *java1*. In this survey, it has no more been differentiated between communication and collaboration tools. Instead of asking if the provided tools are sufficient, it was asked which of the listed tools would be appreciated most if it was added to the platform. By far most participants asked for a tool that shows which other team members are currently online (see also next section), followed by a team calendar or meeting planner. As in the previous surveys, a word processor with better formatting options than the Etherpad was also quite

1. This is not so astounding as the task in this course was to create a UML diagram.

popular. Those participants who opted for "Other" most often asked for better file sharing support (more allowed file formats, updating files, file descriptions, etc.).

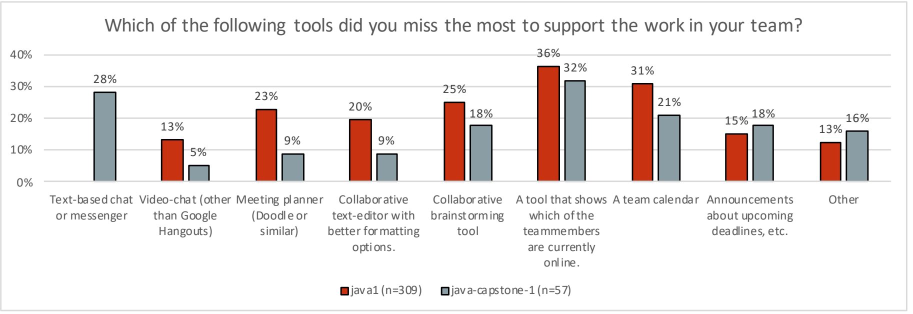

Figure 5.4: Post-teamwork survey *java1* (n=309) and *java-capstone-1* (n=57). Which of the following tools did you miss the most? (Multiple answers have been possible.)

Communication tools

Fig. 5.5 shows that most participants either found the provided communication channels to be sufficient or were able to enhance them with third party communication tools of their choice. Only the participants in *bizmooc2018*, seemed to have problems finding suitable alternatives when they rejected the tools that we have provided. The most requested additional communication channel was a regular text chat with a proper notification function. The interviews with some of the team task participants in *javaeinstieg2017* confirmed this as well.

The text chat options that are offered by the Etherpad and TogetherJS, have been perceived as too hidden to be useful for a general purpose. Particularly, TogetherJS has rather confused the participants than to support the communication in the team. Several participants used alternative or additional communication channels, such as *Skype*[1] and *WhatsApp*[2], and in the *java*2017* courses also *Discord*[3] and *Teamspeak*[4].

1. https://www.skype.com/en/
2. https://www.whatsapp.com/
3. https://discordapp.com/
4. https://www.teamspeak.de/

Common issues with Google Hangouts were:

1. Participants do not have a Google account or perceive Google as “evil”[1].
2. Participants perceive a video chat as too intrusive towards their privacy[2].

A particular issue with the video chats has been revealed during the *javaeinstieg2017*-interviews. Many participants have tried to start a video chat without scheduling a meeting first. For some reason they imagined that some one of their team will be online as well, so that they can talk. Interestingly, some of these participants are well used to work with synchronous meeting tools from their daily jobs, where they would never get the idea of just starting a video chat without scheduling a time slot first. This elucidates the strong request for a tool that shows who is online (as many participants are used to nowadays by their daily use of synchronous communication tools, such as Slack[3] or Rocket.chat[4]), as well as for calendar and scheduling tools. Several participants, however, stated that they are strictly averse to the display of their online status, as they consider this, too, to be a strong privacy intrusion. However, the

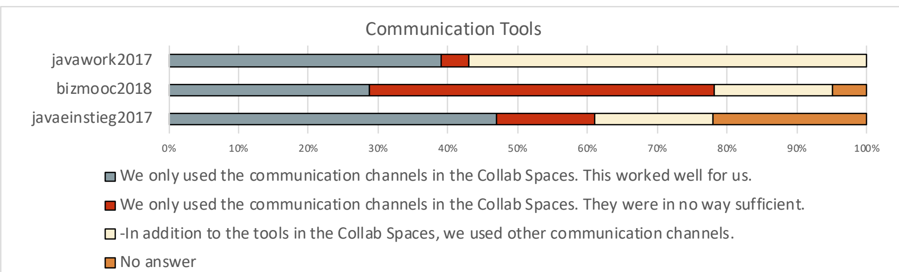

Figure 5.5: Post-teamwork surveys *javaeinstieg2017* (n=340), *bizmooc2018* (n=42), *javawork2017* (n=23) Have the provided communication tools (Teamforum, Hangouts, TogetherJS, Chat in the Etherpad) been sufficient to solve the task?

surveys that have been conducted among the team task participants of *java1* and *java-capstone-1* revealed that the by far most used communication tool was the discussion forum. In the *java1* survey, the participants have not been asked

1. They are afraid that Google might use their data to do something creepy. In most cases, this perception is completely irrational, as the same participants have been using Skype instead, without questioning that Microsoft is at least as “evil” as Google.
2. In this case, it is not Google that is intruding their privacy but the other team members. Less tech-savvy participants, e.g., often do not know that a video chat can easily be turned into a voice chat by a single click, or a camera lid.
3. https://slack.com/
4. https://rocket.chat/

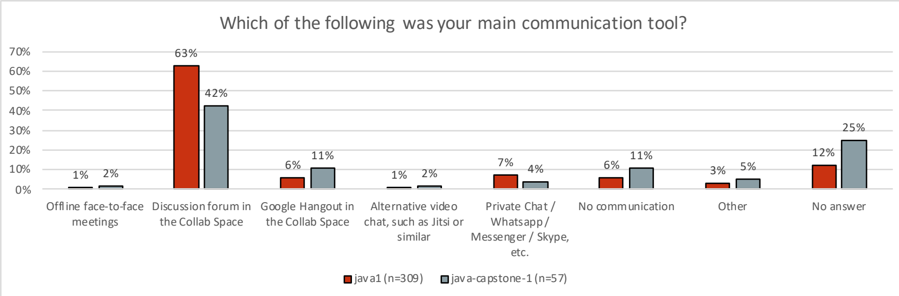

Figure 5.6: Post-teamwork surveys *java1* (n=309) and *java-capstone-1* (n=57). Which of the following was your main communication tool?

about a text-based chat or messenger tool. The survey in *java-capstone-1* seems to add a strong request for an additional text-based chat (see Figure 5.5). Most of the *java1*/*java-capstone-1* interviewees, however, expressed that a separate text chat would rather complicate communication than improving it.

5.3 Learner Background

In order to learn, which learner-type is registering for the team tasks, the team members have been examined in terms of socio-demographic and geographical background, previous team experience, and course participation. Where possible, the data have been compared to the total course population.

5.3.1 Socio-Demographic and Geographical Background

First, it was 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 are collected in the users' profile. It is voluntary to provide these data. About 35% of the team members and 25% of the course population have provided this data.

The geographical data are automatically collected based on the users' IP address, whenever they access the course material. Therefore, geographical data are available for 100%[1] of the team members and for about 60% of the total course population (this closely represents the overall show rate in the examined courses). Figure 5.7 shows the average age of the team members versus the total course population. The maroon bubbles represent the total course population,

1. Team registration is a course item. Therefore, all team members have at least accessed one course item, are tracked geographically and will no-more count as no-shows.

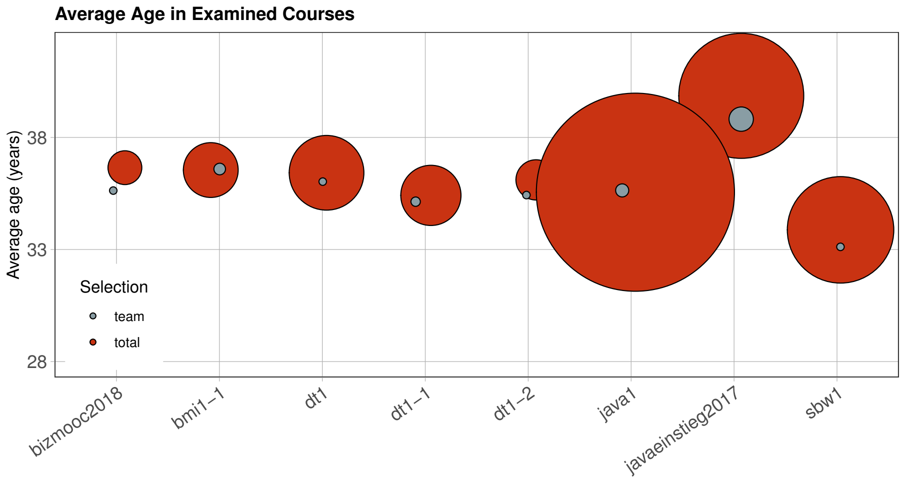

Figure 5.7: The average age (total course population) compared to the average age in teams. The size of the bubbles represents the relative size of each group. The maroon bubble for *java1* represents ~20,000 participants.

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. The graph indicates that only a minority of the course participants register for the teamwork. In addition, 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 are more complete than the age data of the other participants.

Figure 5.8 shows that the vast majority of the participants come with a bachelor's or master's degree[1]. This applies for both, 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 others are teachers, or academic researchers. 30% of the participants have the position of a technician, 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

1. To simplify things, we do not differentiate between a master's and its older German counterparts Magister and *Diplom-Ingenieur*

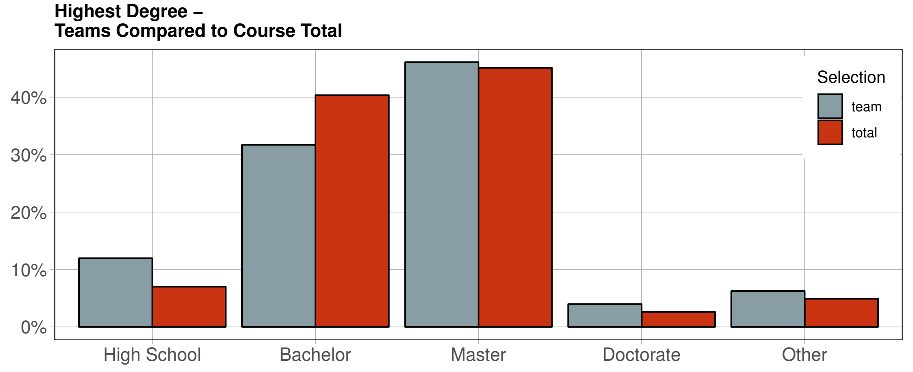

Figure 5.8: Most participants in the examined courses has a Bachelor's or Master's Degree.

for all courses. Teams and course population also do not differ significantly in terms of first time platform users. *javaeinstieg2017* and *javawork2017* were 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' geographic origin. Except for the courses offered in German language, all courses have particularly strong groups of participants from Germany, India, and the United States of America. We have not found any fundamental differences in geographic origin between team members and total course population.

To sum it up—the socio-demographic and geographical background of the team members roughly parallels the background of the total course population.

5.3.2 Previous Team Experience

According to the pre-team-task surveys in *bizmooc2018* and *java1*, most of the team members feel comfortable working in teams (see Figure 5.9). About 60% of the participants have not worked on a team task in a MOOC before. The other 40% have participated in at least one team task in a MOOC (see Figure 5.10). The very vast majority ($>$ 90%) of those who have participated in team tasks in MOOCs before, rated their previous experience as very good, good, or at least neutral. Only a very small number ($<$ 10%) rated their experience as disappointing or very disappointing.

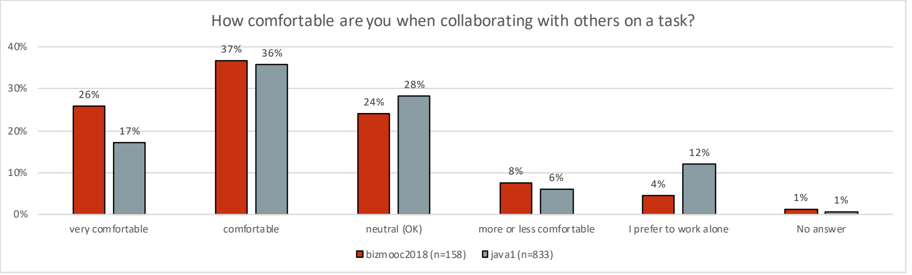

Figure 5.9: Pre-teamwork survey: How comfortable are you when working with others in a team? *java1* (n=833) and *bizmooc2018* (n=158)

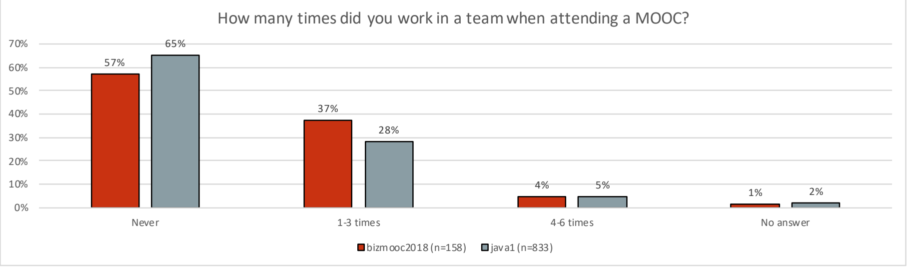

Figure 5.10: Pre-teamwork survey: How often have you worked in a team in a MOOC before? *java1* (n=833) and *bizmooc2018* (n=158)

5.3.3 Course Participation

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

Visited items

Figure 5.11 shows that team members in all examined courses, have visited significantly more items in each section than the total course population[1]. 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 significantly lower percentage of item visits. This can always be explained by "anomalies" in the course structure—e.g., Section 11 in *bizmooc2018* hosts a couple of video outtakes that have been added after the

1. The same phenomenon has been observed in the courses that have been removed from the selection

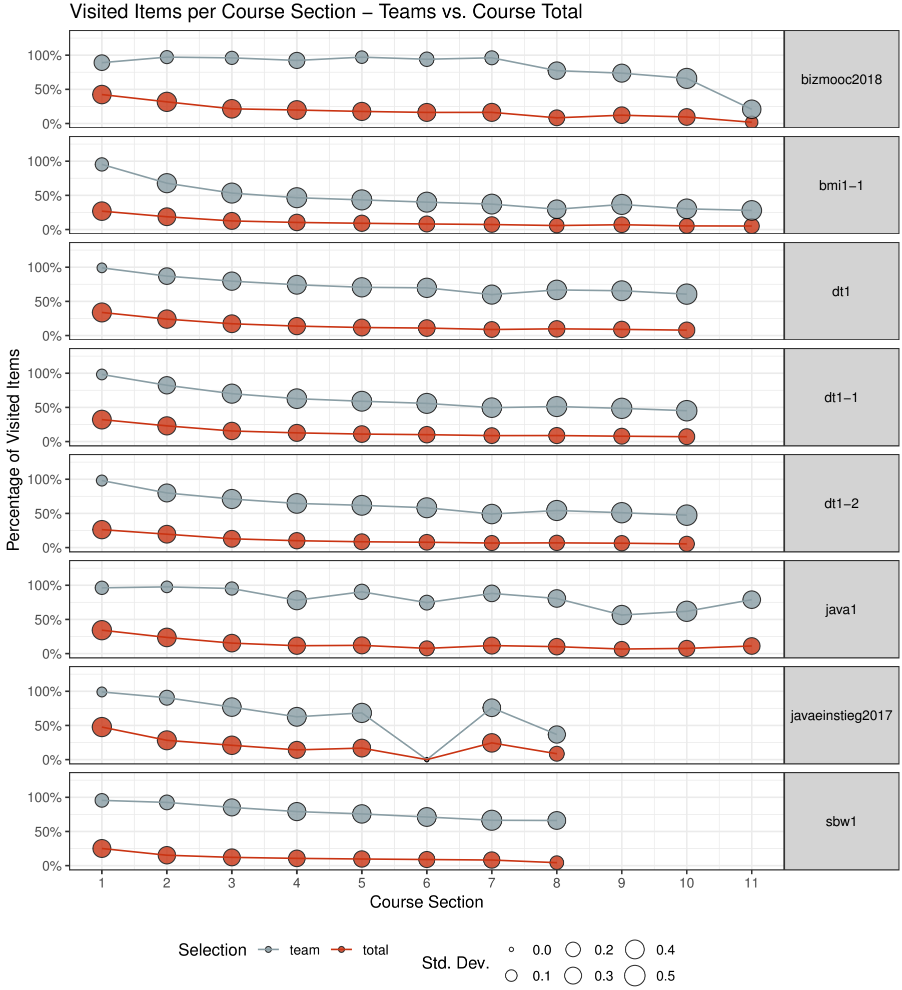

Figure 5.11: Team members have visited significantly more items throughout the course.

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 lead to higher course completion rates—measured in certificates (see Figure 5.12). In total, we can state with great confidence that it are generally the high performers who register for the team tasks.

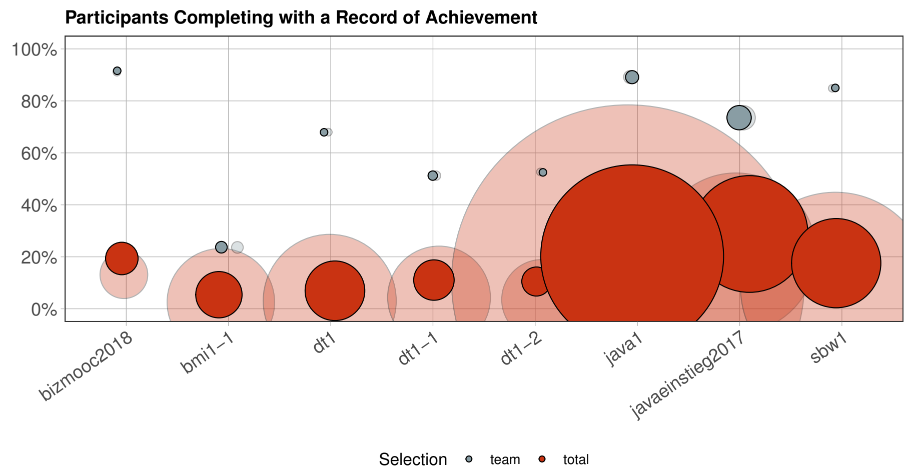

Figure 5.12: The y-axis represents the percentage of participants who earned a certificate. The color of the bubbles represents the selection of examined participants (team members - grey, course total - maroon). The size of the bubbles represents the size of the selection. The transparent bubbles represent 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%.

Forum activity

Finally, the differences in the forum contribution between team members and the total course population have been examined. Figure 5.13 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 astounding 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 TeamBuilder settings confirms that location has been a matching criterion, a quick check of the team data, however does not necessarily confirm this. Another possibility is that they have rather used the video chat or external communication channels than the Collab Space forum for communication. A concluding statement is not possible here as, currently, data about the amount and length of the video chats are not yet collected.

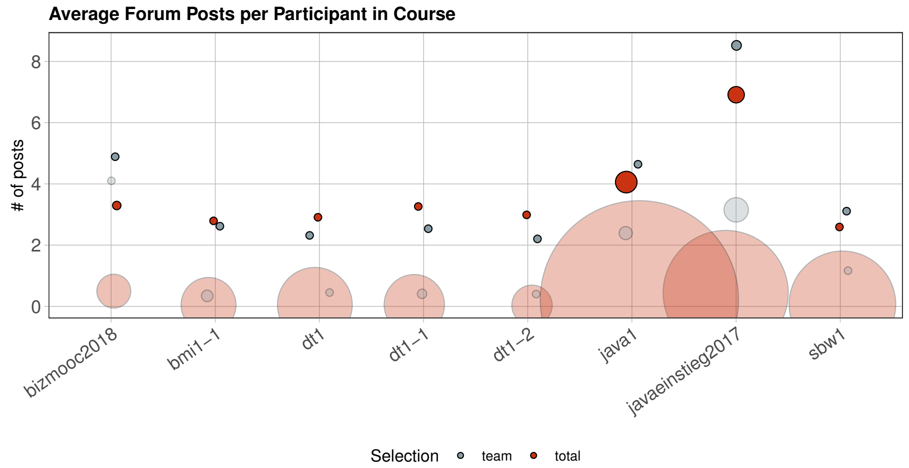

Figure 5.13: The transparent bubbles show the average amount of forum posts per participant. The opaque bubbles show the amount of average forum posts per active forum contributor. Both separately for team members and course population. The size of the bubbles represents the relative size of each group. The transparent bubble at *java1* represents ~20,000 participants.

5.4 Influence of Teamwork on Course Attrition Rate

The high dropout rates in MOOCs are often attributed to a certain feeling of loneliness or isolation in a MOOC (see e.g [143; 144]. Collaborative exercises and team tasks are often presented as an option to break up this isolation and a means to reduce the attrition rate in MOOCs (see e.g., [145; 146]). The question remains, however, is this true?

During the first iteration of the conducted data analysis (*sgc-team1*[1]) the course completion rate of the team workers (CCTM) has been compared to the total course completion rate (CCT) (see Figure 5.16). As a common pattern, the course completion rate among the team workers was significantly higher

1. A detailed description of the involved courses is available in the list of courses in Appendix I - Courses, tagged as *sgc-team1*.

than the overall course completion rate. The team workers also often complete the courses with particularly good results. In *sbw1*, 60% of the learners who successfully completed the team assignment were among the Top 5[1] of the course. 97% of the course participants among the Top 5 were team workers. In the *java*2017* courses, about 80% of the successful team workers were among the Top 5 and in the *dt** courses it was about 20%. Although only 3% of the enrolled users in *javawork2017* had opted to work on the project in teams[2], 16% of the Top 5 were team workers. In *javaeinstieg2017*, 23% of the enrolled users had applied for the team assignment but 74% of the Top 5 were team workers.

Generally, the completion rates of the courses on the HPI platforms are comparatively high. The average completion rate for the ten most recent courses without team assignment was 29% on openHPI and 35% on openSAP[3]. The completion rate of most of the courses with team assignments is slightly below this average, which is expected as these courses generally come with a higher workload than courses that fully rely on quiz based assignments. Compared to their previous iterations in 2015, which did not feature a team based assignment, the completion rates of the *java*2017* courses are similar. In *javaeinstieg2015* a very similar task to the team task in *javaeinstieg2017* was offered as a peer-assessed (but not team-based) task. The completion rate for this task is also very close to the team task completion rate in *javaeinstieg2017*.

Considering the observations in Section 5.3.3, however, it is questionable to conclude that team tasks help to decrease the attrition rate of a course. The more reasonable conclusion is, that the relation is the other way round. Only the high performers register for and succeed in the team tasks.

5.5 Team Task Completion Rates and Free-Riders

To successfully complete a team task—at least in theory—the participants have to do two things: 1. They have to contribute to the team's submission in a substantial way. 2. They have to assess the submissions of a given number of other teams. As all team members are automatically pulled to the next step, when one team member submits the work, it is hard to tell who really did contribute and who is free-riding. The team peer assessment feature is built

1. The Top 5 are the best 5% of those course participants who successfully completed the course with a certificate.
2. In this course the participants had the choice to either work on the task in teams or solo. The task was exactly the same for both groups.
3. At the time when this research has been conducted (end 2017/beginning 2018). The average number of enrollments at course middle on both platforms: ~5600.

Table 5.1: Comparison of interviewees' statements about active team contributors vs. platform data of completed tasks.

	# of teams	Details
Platform data fits interviews (all free-riders detected)	10	Based on 14 participant statements. All team members that represented the same team, confirmed each other's statements.
Data likely fits interviews (probably most free-riders detected)	2	Two of the statements have been somewhat unclear about the active members, or the interviewees did not really remember the details. Most likely, however, the data fits.
Active team members, who have not written reviews (false positives)	13	Considering only the actual team task, these teams would have had a higher completion rate.
Free-riders not detected (false negatives)	3	At least in one of the three cases, free-riding seems to be an intentional act. Only one team member was active. When the deadline approached, another team member submitted the active member's work.

on the assumption that those who have not contributed to the submission, will also not assess the work of the other teams, and therefore, these free-riders will be filtered out automatically in the next step. So far, however, this is just an educated guess. Figure 5.16 (ITAT) shows that there are "free-riders" in active teams that are filtered out by the system. We do not know, however, which of those actually have contributed to the teamwork, and just have not assessed the work of others (false positives, so to say[1]). We also do not know how many

1. If "no reviews→no points" is merely seen as a filter mechanism to detect free-riders, these people would be false positives as they actually have contributed. As writing reviews, however,

of those, who have assessed the work of others, actually have **not** contributed to the team's submission (false negatives, or actual free-riders) As the system is somewhat non-transparent here, it is hard to ascertain. The interviews with the teamwork participants in *javaeinstieg2017* and *java1*, have opened a little peephole here. For the teams of the interviewees, the platform's task completion data have been compared to the interviewees' statements about the contribution of their team members. Albeit the number of comparisons is rather small, as the data results from only thirty-two interviewees representing twenty-nine teams, it confirms our assumption that the vast majority of free-riders are filtered out by the requirement to review the work of other teams in order to receive points for their own work. Only in three out of the twenty-nine teams, known free-riders have not been "detected" by the mechanism. The Pearson correlation between the number of "known" inactive team members and the number of "detected" inactive members reveals that the result is statistically significant ($n = 29$, $R = 0.53$, $p < 0.002$). On top of that, the data indicate that there are several false positives; team members, who have actively contributed to the task, but who did not assess other teams' work (See Table 5.1).

5.6 Minimizing the Team Dropout Rates

Given that the platform supports tools to assist the participants keeping on track—such as automatic reminders—and that the course quality meets the platform's standards, and, finally, that the instructors meet at least the minimal communication standards; the dropout rate, generally, should not be considered to be a big issue anymore. MOOCs are "easy come—easy go". Actually, this is even one of their strengths—people who otherwise would demur from registering for a course, because of the obligations that come with it, can easily do so and do not have to fear any consequences if life gets in the way of their plans. In team tasks, however, dropouts can cause problems and frustration for the other team members. Figures 5.14 and 5.15 show the results of the post-teamwork surveys in *java1* and *java-capstone-1*. The participants have been asked how many members had been left in their team at the end of the task and how the others have vanished. The teams in both courses started with a size of about 6-8 members. The tasks in the courses differed a lot in terms of complexity and also in their significance within the course. The task in *java1* was a small side project to introduce object-oriented modelling and offered

is more than just a filter mechanism, but has a value of its own, it is absolutely correct, that these people have not received points. It should also be stated explicitly that the vast majority of inactive team members are not inactive for shady reasons, such as intentional free-riding. Most of them just have underestimated the effort, or overestimated their spare time.

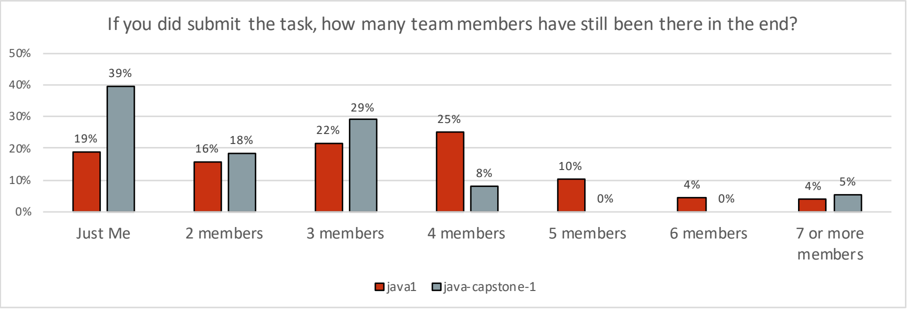

Figure 5.14: How many active members have been left at the end of the task?
Post-teamwork survey in *java1* (n=245) and *java-capstone-1* (n=38)

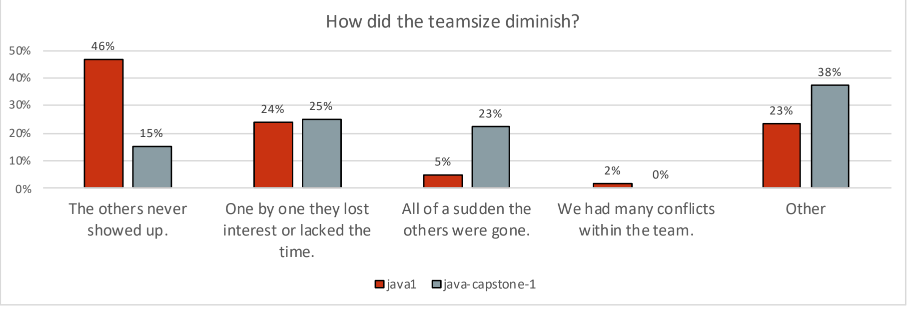

Figure 5.15: How did the team size diminish?
Post-teamwork survey in *java1* (n=254) and *java-capstone-1* (n=40)
"Other" was most often specified in more detail as 1. "the task was too advanced", 2. "they said hello, but never came back", 3. "everything's fine, no one dropped out."

only a few bonus points. In *java-capstone-1*, the participants had to deliver a complete software project in the form of a game. The success in this project also defined the success in the course. *java-capstone-1* was a follow-up course of *java1*. This explains to some extent, the differences in the way the team members dropped out in these courses (see Figure 5.15). While in *java1* the majority just did not show when the team task started, this group was much smaller in *java-capstone-1*. More participants already were familiar with that type of task from the previous course, fewer participants just wanted to peek in to see what it is. In *java-capstone-1* the task turned out to be either too complex or too time consuming for many participants at a certain point of time, which shows in a higher amount of team members that all of a sudden left the team. It is very interesting, that conflicts within the teams did **not** play a role in any of the examined courses. Generally, three categories of dropouts have to be differentiated :

1. Participants who drop out of the **course** between registration for team task and begin of team task.
2. Participants who drop out of the **course** during the team task.
3. Participants who drop out of the **team task**, but still complete the course.

The following three subsections will discuss several characteristic aspects for each of these categories. First, the effects of extrinsic motivation to complete the team task will be examined. This is particularly relevant for the third dropout category. Then, the effects of alternatively offered "solo" activities will be discussed. This is particularly relevant for the second dropout category. Finally, options to predict probable dropouts will be examined. This is particularly relevant for the first dropout category.

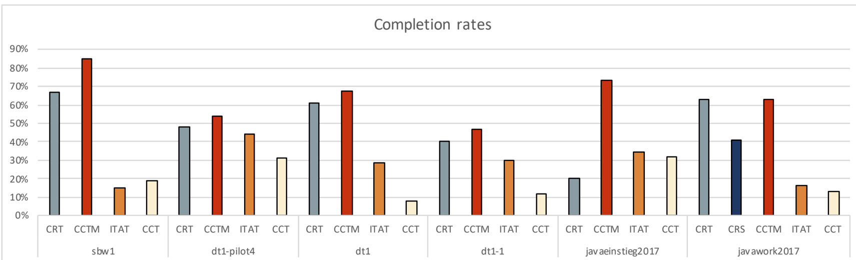

Figure 5.16: Comparison of completion rates in *sbw1*, *dt1**, *java*2017*.
CRT - Team workers who successfully completed the team assignment
CCTM - Team workers who completed the course with a certificate.
ITAT - Potential free-riders in active teams (Members who have not assessed the work of other teams (and, therefore, have not received any points)).
CCT - Course completion rate (total).
CRS - Completion rate of solos (only *javawork2017*).

The Role of extrinsic motivation

Figure 5.16 shows that the third category can be a rather large group, if the team task does not contribute a significant amount of points to the total result (compare CRT with CCTM for *javaeinstieg2017*). Extrinsic motivation is offered in the form of course-points in the examined courses. The level of extrinsic motivation grows with the points' significance to pass the course. In some courses, e.g., *javaeinstieg2017*, it was possible to pass the course with an 100% result without participating in the team task at all, as it was only a bonus exercise. In other courses, e.g., *javawork2017*, the team task was basically the defining element to pass the course. The question is now if this increased level of significance is reflected in the tasks' completion rates, and which effect it has on the courses' completion rates.

On the one hand, the amount of points seems to correlate closely to the team task's completion rate. The completion rate in the mandatory tasks is significantly higher than in the bonus tasks. The comparably low task completion rate in *javaeinstieg2017* results from a high workload and only few available points (see Table 5.2). On the other hand, almost none of the successful team workers in this course, was in need of these points to complete the course with a good result. The completion rate of 20% for the team assignment, therefore, has to be seen very positively. The task completion rate, however, is measured

Table 5.2: Value of team task within course vs. completion rate of team task
([1]) Bonus
([2]) Regular
([3]) Including milestones
([4]) In this course the task was available both as a team task and as an individual task. The number only refers to those that have started the task as a team.

Course	Value of team task	Team task completion
sbw1	33%[1]	67%
dt1-pilot4	57%[2,3]	48%
dt1	57%[2,3]	61%
dt1-1	57%[2,3]	40%
javaeinstieg 2017	6.25%[1]	20%
javawork 2017	100.00%[2]	63%[4]

relatively to the number of participants who actually started to work on the task. The question is, which effect does it have on the course completion? Table 5.3 shows the number of participants in total versus those who de facto became active in the course assignment—in a team or solo—in *java-capstone-1*. The significance of the assignment within the course has been very high. The participants had to pass this assignment to complete the course with a certificate. The course was very well appreciated by a few more advanced users. In total, however, it had a disastrous completion rate (see Table 5.3). So, increasing the significance of such a task within the course has a negative effect on the overall

Table 5.3: *java-capstone-1* completion rates. Nine of the thirteen teams have passed as teams, four of them have passed dysfunctional (only one member left). The completion rate for teams, lists the overall value including the dysfunctional teams.

# of Participants	Total	Passed	Completion
Course population	2408	51	2%
Team members	218	33	15%
Teams	32	13	40%
Solo	144	32	22%

completion rate. An analogous effect has been observed in *javawork2015*, and to some extent also in *javawork2016*, and *javawork2017* (see Table 5.4). The main reasons that have been given by the participants were:

1. a general lack of time on their side,
2. the time frame to solve the task was too short,
3. the learning curve from the introductory Java courses to the follow-up workshops increased too steeply.

“Solo” alternatives

Some participants are only interested in the task itself, but not at all to work in a team. Many participants have either underestimated the team task’s workload, or the circumstances in their lives have changed, so that there is not as much time left for the course as they originally had planned. One option would be to discourage already shaky candidates from registering for the team task by informing them in detail about the consequences for the rest of the team. Willcoxson [130], suggests to take this even a step further and to make the students sign a contract on registration that commits them “to contribute positively to team dynamics.” A slightly different approach was taken in *javawork2017* and *java-capstone-1*. The same task was offered in parallel, both as a team assignment and as a solo assignment. Many participants have expressed that they appreciate the possibility to work on the same task alone and the resulting flexibility. On the other hand, we have also received feedback from many participants that they originally had not intended to work in a team, but later

Table 5.4: Course completion vs. peer assessment completion.
Course: Course completion ($^{issuedCertificates}/_{participantsAtCourseMiddle}$)
PA: Peer Assessment completion ($^{peerAssessmentSuccessfullyFinished}/_{peerAssessmentStarted}$)
Type: Teamwork or individual work
Sig.: Significance of the task for course completion (low: mostly bonus, high: required for certificate, additional: required for extended certificate.)

Completion rates	**Course**	**PA**	**Type**	**Sig.**
javaeinstieg2015	38%	22%	solo	low
javawork2015	7%	24%	solo	high
javawork2016	18%	41%	solo	high
javaeinstieg2017	32%	21%	solo	low
javawork2017-t	13%	73%	team	high
javawork2017-s		37%	solo	high
bizmooc2018	27%	61%	team	add.
java1	24%	48%	team	low
java-capstone-1-t	2%	19%	team	high
java-capstone-1-s		24%	solo	high

on enjoyed it very much and would not have wanted to miss it. It is hard to say whether offering the option to work in teams or solo really affected the completion rate of those starting to work on the task in teams. Too many other factors are involved, and none of the other courses are really comparable in this context. Nevertheless, based on the data in Tables 5.3 and 5.4, several statements can be made:

1. Courses that strongly depend on project work[1] will have a lower completion rate.
2. It cannot yet be stated for sure—given that the task can be submitted solo or as a team—who performs better, the solos or the teams. While *javawork2017* tends strongly towards the teams, *java-capstone-1* points slightly in the opposite direction. Tallying the numbers, results in a slight

1. The term peer assessment has been avoided here on purpose, as we are convinced—based on the combined results of our research—that it is not the peer assessment, which is the culprit for the low completion rates, but the given task or project and the workload that comes with it.

advantage for the solos.

3. Raising the significance of the peer-assessed task within the course's grading model is increasing the completion rate of the task (∅55% (high) vs. ∅38% (low)). It is, however, also significantly decreasing the overall completion rate of the course. (∅30% (low) vs. ∅10% (high)). Providing such tasks as add-ons with a separate certificate (as done in *bizmooc2018*), seems to be the most promising approach.

Dropout prediction

During the first iteration of the data analysis, it was attempted to predict probable dropouts based on the participants' results in the course's exams and exercises before the team registration. The goal is to minimize the dropouts in the teams by selecting the ideal point in time to start the team registration. The examined courses are *sbw1*, *javaeinstieg2017*, *dt1-pilot4*, *dt1*, and *dt1-1*. Figure 5.17 explores the differences between team members who completed the team task and those who dropped out of the task. Participants who achieved at least 80% of the possible score in the weekly assignments were defined as high performers. The solid lines show the participants who successfully[1] finished the

1. We defined participants who received more than zero points in the team task as successful. Having received more than zero points, indicates that the participants completely went through the task, including grading/reviewing the work of their peers.

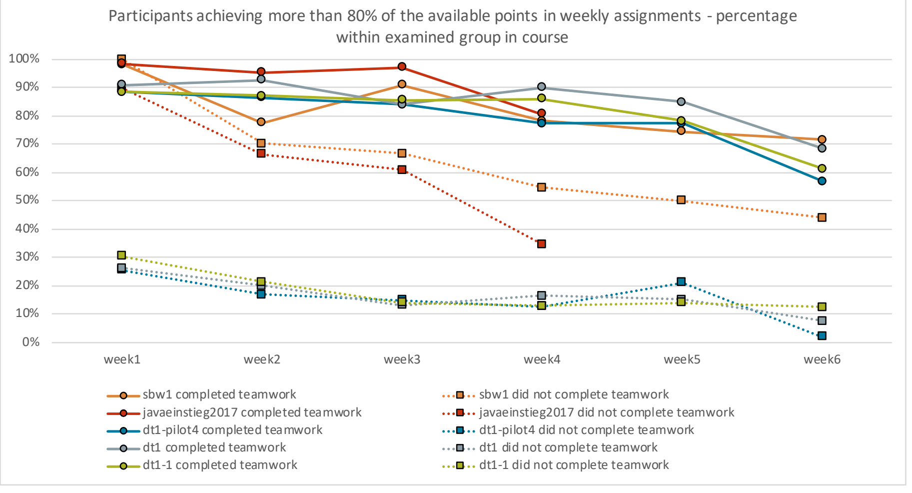

Figure 5.17: High performers in terms of weekly course scores. Clustered by success or non-success in the team task.

team assignment. The dotted lines show the participants who dropped-out of, or did not even start the team assignment. While in the *dt** courses, the distinction between successful and non-successful team workers is well-defined, it looks quite different in *sbw1* and *javaeinstieg2017*. In both courses, successful and non-successful team workers started with an analogous percentage of high performers. To predict which learners will—based on their previous course results—very probably not complete the team task, it does make a difference **when** the data for this prediction are collected. Filtering on the results of Week 1 will often not lead to satisfactory results[1]. In *javaeinstieg2017* the correlation between the results in the weekly scores and the results in the team assignment was investigated in more detail. All participants who registered for the team task and have achieved less than 50% of the points in Week 1, have dropped out of the team assignment. Table 5.5 shows the correlation between the result in the team assignment and the scores in Week 1 to Week 4. The correlation is increasing each week, while the significance stays on a constantly very high level. Looking at the data from another angle, there is a probability close to

Result in ...	**Correlation (R)**	**Significance (p)**
Week 1	0.18	6.81E-13
Week 2	0.26	6.10E-26
Week 3	0.31	1.91E-35
Week 4	0.38	1.00E-55

Table 5.5: The Pearson correlation between the result in the team assignment and the score in Week 1-4 in *javaeinstieg2017*

100%, that a learner who applied for the teamwork and achieved less than 50% of the available points in a certain week, will drop out in the following week. However, filtering out all applicants with less than 50% in Week 1 is not very effective as most applicants have a better result anyway. The effect is much better when looking at the results of Week 2 and still improves a little in Week 3. The difference between Week 3 and Week 4, however, is not really significant anymore. Except for that, waiting till long past Week 2 is not an option anyway, as then time runs short to organize the assignment. The pink rectangle in Figure 5.18 visualizes this observation. Its height is defined by the worst result of a successful team member in the respective weekly assignment (except for

1. It seems as the results in *sbw1* and *javaeinstieg2017* are rather the norm than the exception when they are compared to analogous numbers in courses without team assignments.

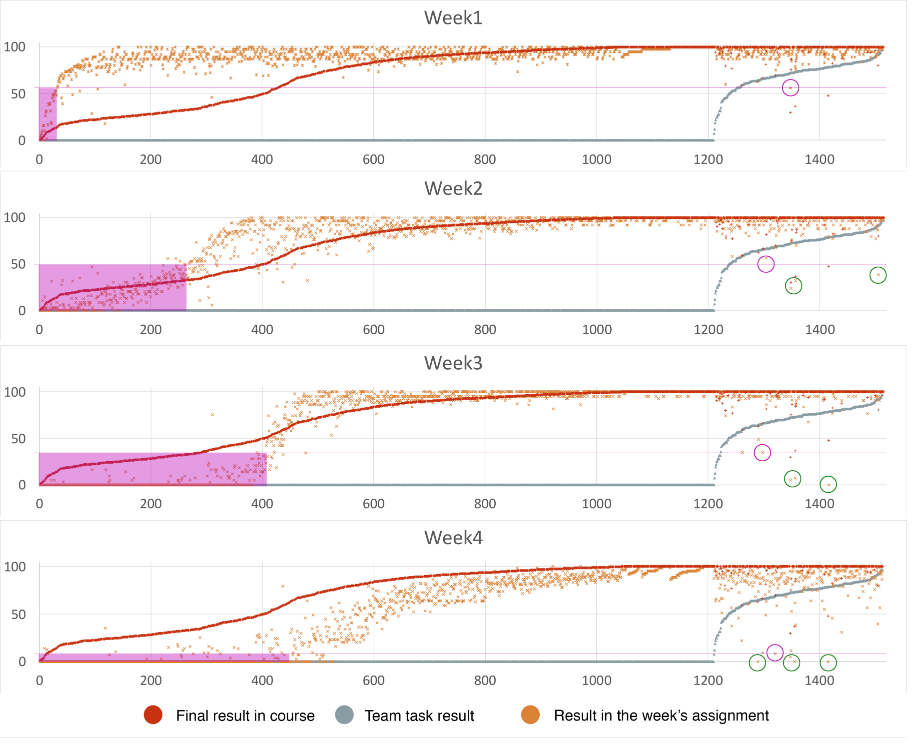

Figure 5.18: Dropout prediction in *javaeinstieg2017*.
Y-axis: percentage of normalized values. X-axis: team workers ordered by result in team assignment and result in course. Pink rectangles: approximate amount of participants that can safely be excluded from the team task. Pink circle: worst result of successful team member in the week's assignment (defines height of rectangle). Green circles: max. three outliers per week.

max. 3 outliers in Week 2 to 4). The width of the rectangle approximates the number of participants who can safely be denied to participate in the team task as they will very likely drop out. Although the situation in *javaeinstieg2017* definitely can be considered as special, owed to the huge amount of team task registrations, Figure 5.18 still illustrates, that denying access to participants who have been identified as dropout candidates based on their results in the course so far, can at best play a minor role in minimizing the team dropout rates. When the approach was applied a year later in *java1*, using the thresholds that had been determined in *javaeinstieg2017*, it allowed us to deny access to only sixty-five of the 810 team task registrations. It definitely contributes to the phenomenon that the first weekly assignment is rather easy to solve, compared to the exams of the following weeks.

Providing an additional assignment— dedicated only to filter dropout candidates—might help to improve the situation but still needs to be evaluated[1].

5.7 Team Matching

Previous studies have shown that working in teams of "strangers" does not only come with disadvantages (see Section 3.2). Furthermore, the situation is often similar in real life. Teams are assembled based on many criteria, personal preferences of team members being the least important. The nature of MOOCs is a further reason, why we also are in favor of this way to match teams. The majority of users is participating in the course on their own, not with a group of friends with whom they would like to form a team. Also, the general participation in the forums is too low to match the participants based on some sort of previous forum interaction. Our approach is, therefore, to (semi-)automatically match the participants, based on a set of criteria, which are defined by the instructors, according to the needs of the course topic and the task. Some instructors prefer local teams that are able to meet face-to-face, while others set a stronger focus on cultural diversity.

Participants' preferences

The results of the surveys that have been conducted among the team-task participants in several courses on the HPI platforms—*javaeinstieg2017*, *javawork2017*, *bizmooc2018*, and *java1*—confirm our assumptions and show that the participants as well have a strong preference towards our approach.

In the post-teamwork surveys in *javaeinstieg2017* and *bizmooc2018*, the participants were asked if they prefer to be teamed or prefer to self-select their team members. In both courses, the vast majority supports our interventionist approach (Fig. 5.19). The slightly larger amount of laissez-faire supporters in *bizmooc2018* can easily be explained by the fact that there was a comparably large group of students from the same university who had to submit the assignment as a part of their grade for an on-campus course and would have preferred to team up with others, who were under the same pressure. In *javawork2017*, we used the laissez-faire approach and, therefore, asked who would have rather worked in a team than solo if we had used an interventionist matching approach. 17% of the participants supported that[2].

1. Actually, this approach was a part of the feature's original design. An additional "Week 0" had been planned as an introduction to the team assignment, including an exam just for the purpose of determining the commitment of the applicant. At some point this assignment was abandoned in favor of employing the regular course assignments.
2. As those who answered "No" to the question, have not worked in teams but on their own, we

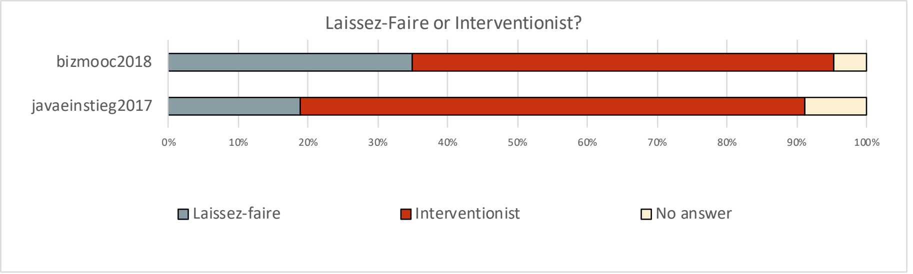

Figure 5.19: Post-teamwork surveys *javaeinstieg2017* (n=340), *bizmooc2018* (n=42). Would you rather be teamed (interventionist) or would you rather self-select your team members (laissez-faire)?

Matching criteria

One factor that turned out to be of particular interest for the participants, was the time commitment—the number of hours that a participant is willing to spend per week on the given team task.

In *bizmooc2018*, and *java1* the participants have been asked in a pre-teamwork survey about their expectations towards their teammates. In both courses, the most common expectation among the participants was a "similar (to their own)" time commitment (about 40%). "A high level of commitment" and "respecting my time"—options with a similar direction of impact, made up for another 30%. "A safe environment to communicate ideas", made it to rank three and "having fun with the teammates" was the least popular option. The results have been very similar in both courses (see Figure 5.20). The post-teamwork

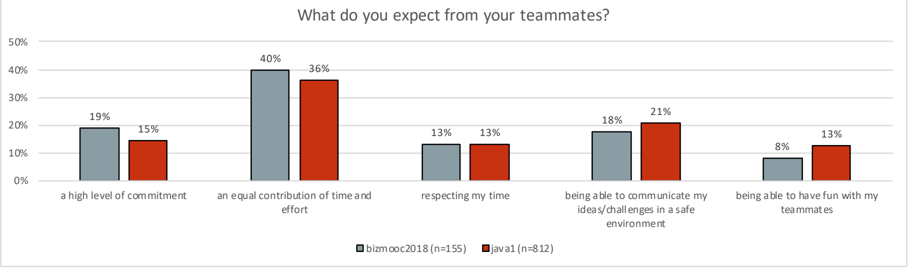

Figure 5.20: Pre-teamwork surveys *bizmooc2018* (n=155), *java1* (n=812)

survey among the team task participants in *java1* (see Figure 5.21) also shows

are safe to assume that they rather meant that they prefer to work alone than that they prefer the laissez-faire approach.

that a similar time commitment is the most important matching criterion for the participants. Furthermore, it shows that Wen's [116] approach to match team members based on their previous interaction in the forum is one of the least favorite approaches among our participants. Given that only a very small subset of the course population is actually actively posting in the forum, this approach is not feasible in our context anyway. The interviews provided similar results.

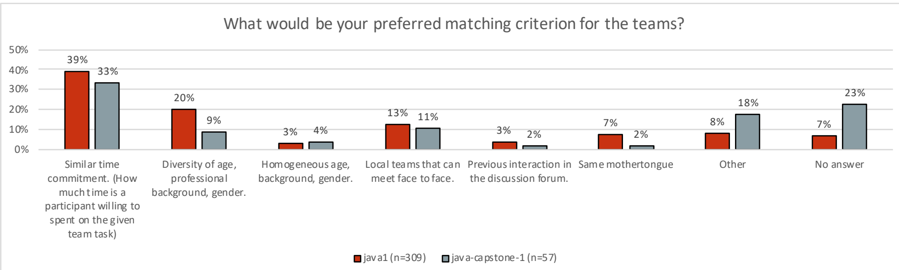

Figure 5.21: Post-teamwork survey *java1* (n=309), *java-capstone-1* (n=57)

None of the participants—no matter if their teamwork experience was good or bad—was interested in having fun with the other participants. Only marginally few of them have stayed in contact after completing the team task. Many interviewees stated that they enrolled in the course to gain new knowledge or to acquire certain competencies. They emphasized that they are not interested in another social media platform, or to make new friends.

Re-Matching

Once a team has started with the peer assessment process, re-matching teams requires so much effort that it is nigh on impossible. Our common practice is to start with the peer assessment process from day one of the project time. However, in most cases this would not be necessary if the description of the project/task was delivered separately as a plain text-item instead of being part of the peer assessment's submission step. Actually, this approach would solve several issues at one sweep. Some participants, e.g., expressed that they wouldn't have joined the course in the first place if they had known what expected them in the project. When matching the teams is decoupled from the start of the peer assessment process, the participants could also be asked to start working on a small pre-project task to prepare the actual project[1]. Issues within the teams would emerge before the start of the actual task and many no-shows

1. A lightweight version of openSAP's milestone approach.

and dropouts could be detected before "*rien ne va plus*." As long as the actual peer assessment process has not started, re-matching teams is comparably easy (at least from the technical perspective). Active participants in inactive teams can be gathered to form new teams. As this, currently, still is a mostly manual process, it does not scale yet. A prototypical project exploring the possibilities of this approach is currently developed and will be conducted in October 2019. If the results are promising, we will consider to develop software support for such a solution.

5.8 Approaches to Teamwork

In all given team tasks on the openHPI platform, we explicitly allowed both: a collaborative, discussion-oriented approach or a cooperative, divide-and-conquer approach. The participants in *javaeinstieg-mint-ec-2018*, *java1*, and

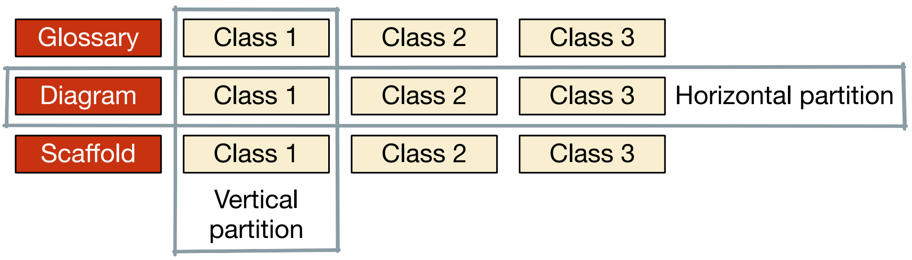

Figure 5.22: Vertical vs. horizontal partition of tasks

java-capstone-1 have been asked how they approached the task in each of the courses' post-course surveys. The options have been:

- We mostly collaborated. We scheduled meetings, discussed solutions, and did most of it together.
- We mostly cooperated. We horizontally split the task (along several subtasks—one member creates the glossary, the next one the class diagram, the next one the documentation).
- We mostly cooperated. We split the task vertically along the several (Java-)classes that finally result in the solution (see Figure 5.22).

While *javaeinstieg-mint-ec-2018* addressed mainly high-school pupils, *java1* mainly addressed an older, professional audience. The differences were manifested mainly in the course settings. The time frame of *javaeinstieg-mint-ec-2018* was stretched to better match the realities of schools. The course *java1*

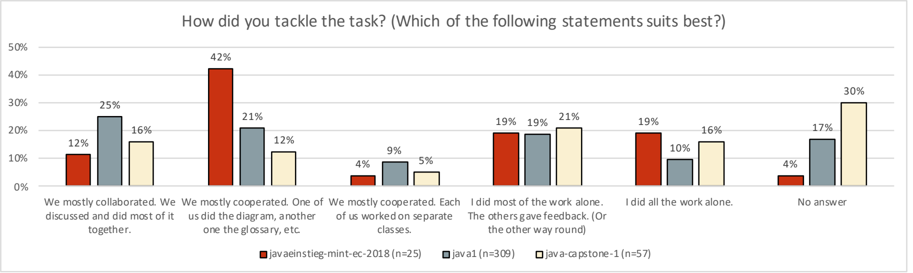

Figure 5.23: How did you tackle the task? (*java1*: n=309, *javaeinstieg-mint-ec-2018*: n=25, *java-capstone-1*: n=57)

was offered on openSAP, where the share of professionals among the participants is similar—maybe even slightly higher—as on openHPI and the audience is more international. The contents have been almost identical. *javaeinstieg-mint-ec-2018* was offered in German, *java1* was offered in English language. The school versions of the courses come with minimal instructor support. The forum activity in these courses is generally very low as most participants have local support from their teachers. While the two previously mentioned courses addressed Java beginners, *java-capstone-1* addressed an audience with more advanced Java skills. Previous knowledge has not been enforced, but it was communicated and expected, that the participants have at least completed one of our introductory Java courses.

Interestingly, the professionals in *java1* tended stronger towards a collaborative solution than the pupils in *javaeinstieg-mint-ec-2018*. Even those who had opted for a cooperative approach, more often have chosen a vertical partition of the work, which by nature is closer to a collaborative approach than the horizontal partition along the given subtasks (see Figure 5.22). The survey results from *java-capstone-1* seem to put this in perspective again. One explanation might be that many participants in *java-capstone-1* have already answered the same survey in *java1* and, therefore did not complete the survey, as shown by the comparably high amount of "no answers." Another explanation might be that the very tight schedule of *java-capstone-1* has not offered a sufficient amount of time to tackle the relatively complex task collaboratively. Finally, the number of survey participants in *javaeinstieg-mint-ec-2018* (25) and *java-capstone-1* (57) is very small, compared to *java1* (309), and, therefore, might introduce some inaccuracies.

5.9 Aggregation of Team Data

So far all analyses have been performed on the participant level. Now that a basic comprehension of the differences and commonalities between team members and the total course population has been established, the aggregated team data will be examined in more detail and the teams' dropout rates and performance in regard to various aspects of the composition of their members will be compared. The data of all team members have been aggregated, to

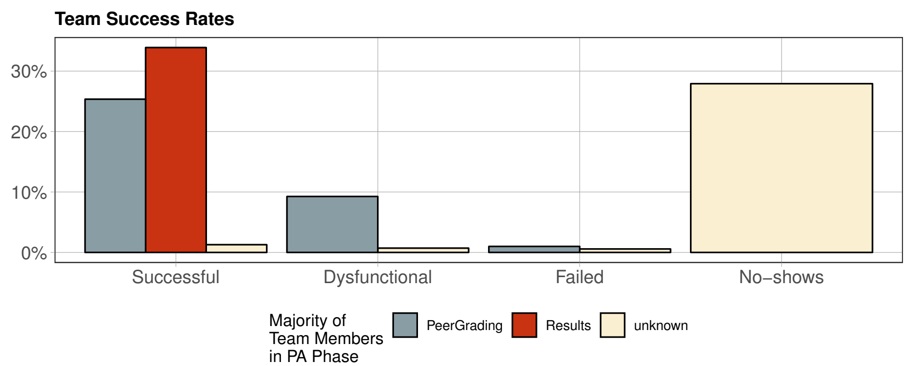

Figure 5.24: Team success rates. Successful teams: two or more members have finished the task, including the peer review. Dysfunctional teams: one team member has finished the task. Failed teams: The team has started to work on the task but none of the members have completely finished the task. No-shows: teams that have not started to work on the task. Unknown: The participants have no entry in the peer assessment data.

generate a single observation for each team in all courses. Appendix IV - Aggregated Team Data lists in detail how the data have been aggregated.

Figure 5.24 shows an overview of the teams' success in the examined courses. A team has been defined 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 have successfully finished the task. A team is considered a *no-show* if none of the members have started to work on the task.

The color of the bar indicates in which phase of the peer assessment the majority of the team members have terminated to work on the task. The grey bar indicates that the majority of the team members have not reviewed the work of other teams—and most probably also have not contributed to the team's submission—the maroon bar indicates that the majority of the team members successfully have finished the task in all phases. The beige bar indicates that the majority of the team has not started to work on the task at all[1].

5.10 Effects of Team Composition

The teams' composition has been examined with regard to all socio-demographic aspects that are available in the data. Sometimes, however, nothing worth to report about has been found. In some cases, the 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 there have been exclusively—or at least dominantly—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. The analysis given here, therefore, has been restricted to observations in the context of the teams' composition in gender, countries of origin, initial team size, and the team members' time commitment. Next to the team's composition, the effects of mentoring (see Section 5.11) and several platform modifications (see Section 5.12) have been examined. For each of these aspects, the teams' dropout rates and the grades that they have received from their peers have been compared. The dropout rate of a team has been defined 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—all team members receive the same grade for their team's submission. Therefore, the team grade has been aggregated as the **maximum** of the team members' grades. Additionally, the points they received for their individual contribution from their teammates and the individual bonus points they received for valuable reviews[2] have been compared. Both are individual components and have, therefore, been aggregated as the **mean** of the team members' grades.

1. The few occurrences of teams that have a majority of members with an *unknown* peer assessment state in the *successful*, *dysfunctional*, and *failed* areas result from *javaeinstieg2017*, where the instructors have removed inactive team members on request.
2. The instructors define the minimum amount of **required** reviews. The system doubles this number to define the maximum amount of **possible** reviews. The available points for the required reviews have been defined as 100%. Therefore, teams can reach up to 200% for the reviews they wrote

5.10.1 Local vs. Remote Teams

The questions of particular interest in this context are if there are differences between local and remote teams in terms of completion rate and performance. The data for this aspect of the study has been collected in the courses *dt1-pilot3*, *dt1-pilot4*, and *dt1* on openSAP and *javaeinstieg-mint-ec-2018* on openHPI. A full description of these courses is available in Appendix I - Courses. All these courses have in common that there have been teams whose members had the possibility to work face-to-face—as they live in the same city, work in the same company, or go to the same school—and there have been teams whose teamwork was restricted to virtual meetings as the spatial distance between the team members did not allow face-to-face meetings. Most of these courses have been pilots or have been run under very different conditions and settings as the other courses. Except for *dt1* they have been removed from most of the other analyses in this chapter, as they added too much noise there. Table 5.6 shows in which state local and remote teams have finished the task. The possible team states (*passed*, *dysfunctional*, *failed*, *no-shows*) have been defined in Section 5.9. The last row in the table aggregates the data for all local, and remote teams in

Table 5.6: Completion on team level by course and Local / Remote.
P = passed, *D = dysfunctional* (only one member completed the task and received points), *F = failed* (started to work on the task, but no team member received any points), *N = no-shows* (none of the team members started to work on the task)
C1 = dt1, *C2 = dt1-pilot3*, *C3 = dt1-pilot4*, *C4 = javaeinstieg-mint-ec-2018*

		# Teams	% Teams	P	D	F	N
C1	Local	15	25%	80%	7%	0%	13%
	Remote	46	75%	87%	2%	0%	11%
C2	Local	11	58%	100%	0%	0%	0%
	Remote	8	42%	100%	0%	0%	0%
C3	Local	11	53%	80%	10%	0%	10%
	Remote	9	47%	45%	22%	11%	22%
C4	Local	5	50%	100%	0%	0%	0%
	Remote	5	50%	40%	20%	20%	20%
Total	Local	41	38%	88%	5%	0%	7%
	Remote	68	62%	79%	6%	3%	12%

all examined courses. The dataset contains a slightly higher number of remote teams (60/40). More local teams have passed the assignment and more remote teams have not started to work on the task. There is not much difference for dysfunctional and failed teams.

In contrast to the other examined courses, *javaeinstieg-mint-ec-2018* did have a very special target group: high school pupils. To fit better in the daily routine of regular school life, the course was stretched so that it ran three months instead of four weeks—with only minimal changes in the course content. The participants—or in some cases their teachers—have decided whether they want to work in local teams within their school or in distributed teams with members in classrooms throughout Germany. Additionally, the teams were matched based on the participants' time commitment. Sixty-two participants had registered for the team task, fifteen of them have been teachers or other grown-ups, another five of them have been adult pupils of a vocational school. Twenty-six of the sixty-two have successfully finished both components of the peer assessment. Twelve of the thirty-six "dropouts" have been teachers who only wanted to see what the students were supposed to do and how the process works. Examined from a team perspective, there have been thirteen teams in total. Five local teams in three different schools, five remote teams, and three teams of teachers. The teachers' teams have not been considered in this evaluation. All five local teams succeeded, two of the five remote teams succeeded as well, one remote team was *dysfunctional*, two dropped-out completely. The number of teams in this course is not sufficient to make a bold statement, but there seems to be a tendency that remote teams are more likely to drop out than local teams. Figure 5.25 seems to imply that the performance of remote teams in terms of *grade-by-peers* is lower as well. The box plot, however is somewhat misleading in this case as the two teams that haven't started, and therefore, received zero points are distorting the results of those who actually have worked on the task. Obviously, local teams have certain advantages as they need less technical support than the remote teams and many of the tasks to be done can be accomplished with less organizational overhead. However, the best result—considering both: achieved points and team cohesion—has been achieved by a remote team. The results of those who have submitted their work also have not been worse than those of the local teams. Furthermore, there have been additional aspects, particularly in this course that might have had an influence on the success of a team. Albeit we do not know for sure in all cases, there is some evidence that the course had been included more closely in the school curriculum for at least some of the local teams, while it was less formally included for most of the remote teams (in other words: the grade of the pupils in class was dependent on their course result or not). The two *dt1-pilots* have been been mostly internal courses with an invited audience in order to test course and platform features. *dt1-*

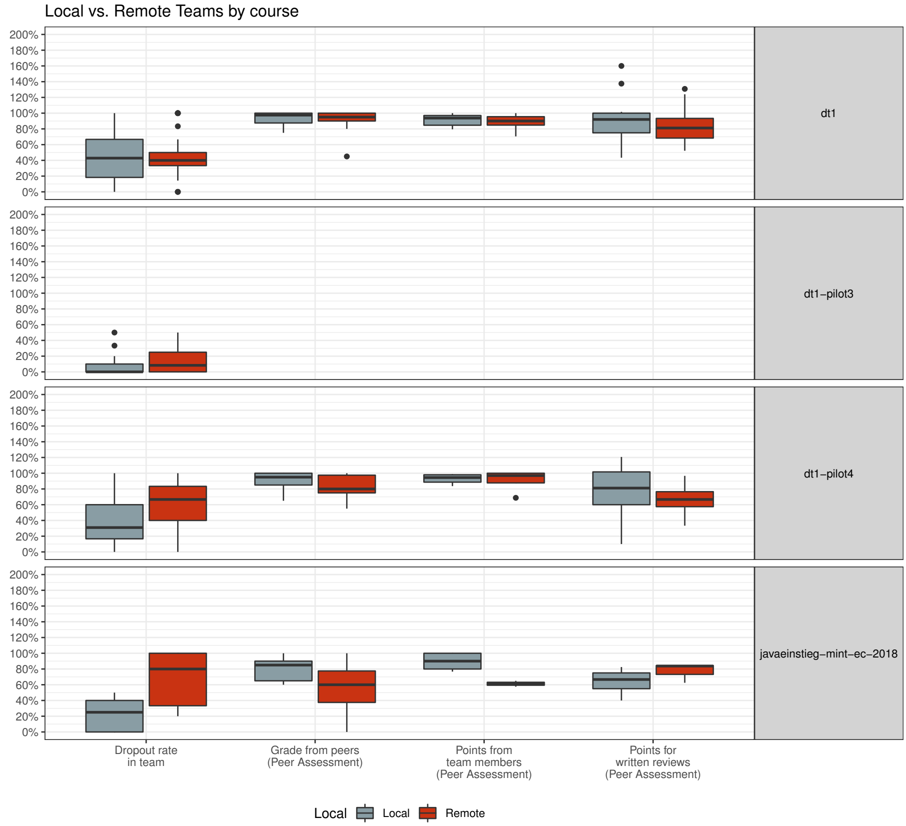

Figure 5.25: *javaeinstieg-mint-ec-2018*: 5 local, 5 remote teams; *dt1*: 15 local, 46 remote teams; *dt1-pilot3*: 11 local, 8 remote teams; *dt1-pilot4*: 11 local, 9 remote teams

pilot3 addressed selected SAP employees only, *dt1-pilot4* addressed invited students of partner universities. *dt1-pilot3* did not have a peer-assessed task yet, therefore, there's no information about the received points. In none of the *dt1** courses, evidence for a performance difference between local and remote teams has been found. Analogous as to *javaeinstieg-mint-ec-2017*, however, there is a higher dropout/no-show rate in the remote teams.

In conclusion, we can state that the performance of the remote teams is comparable to the local teams. This aligns with Kizilcec's [111] finding that, in terms of learning outcomes, there is no big difference between local and distributed teams. In total, however, remote teams tend to have a higher dropout rate than local ones (see Figure 5.26), whereas the higher dropout rate was mostly caused by teams that did not get started at all.

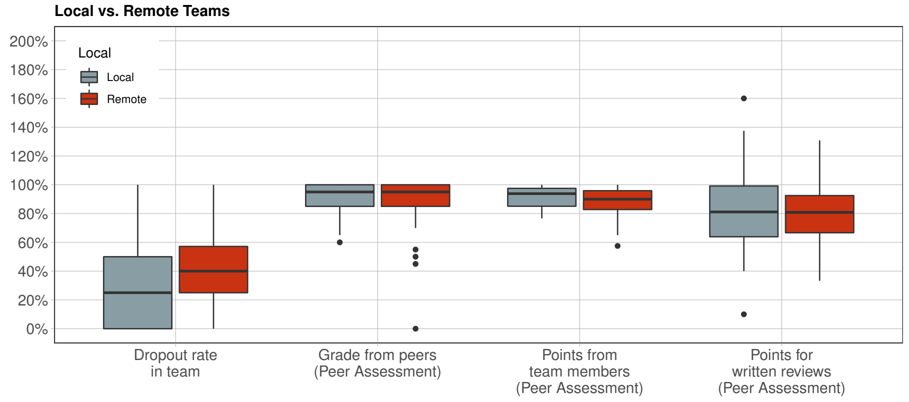

Figure 5.26: 4 courses: 42 local, 68 remote teams

5.10.2 Geographical

For the analysis of the teams' geographical composition, the data of *javaeinstieg2017* has been removed as this course was conducted in German language and, therefore, had a close to 100% German audience. Furthermore, only the countries with the largest populations in the courses—Germany, India, and the US—have been examined. To those, the teams with a heterogeneous geographic background have been added. For this thesis, a team is considered 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 course (see Figure 5.27). Figure 5.28

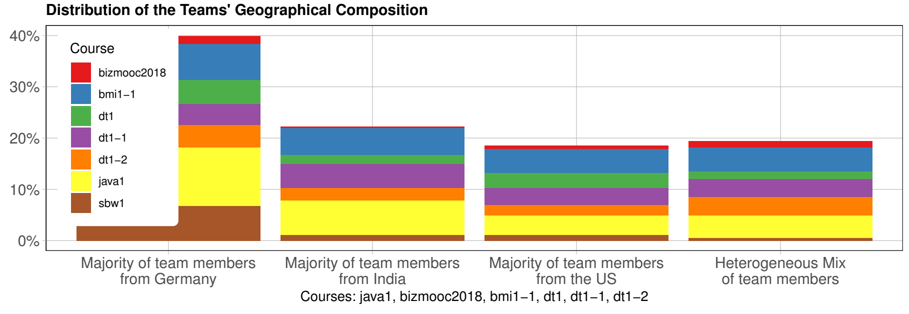

Figure 5.27: The individual geographic groups are of comparable size in each of the examined courses.

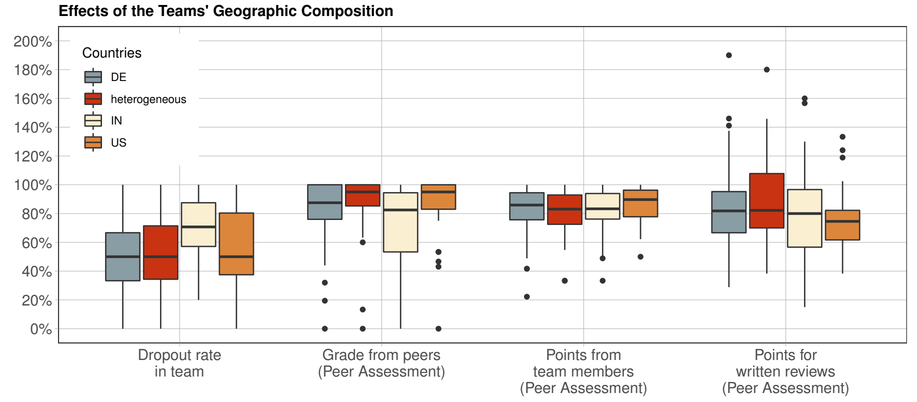

Figure 5.28: Team performance in terms of countries of origin. (451 teams)

shows that different characteristics in teams with particular geographic origins can be observed. The dropout rate is significantly higher in teams with dominantly Indian members. It is lowest in teams with dominantly German members and in teams that are heterogeneously mixed. Dominantly US teams have a dropout rate close to the previous two, but the variance is higher here. US dominated and heterogeneously mixed teams received the highest grades from their peers. Dominantly Indian teams have performed substantially worse. US dominated teams tend to rate each other better within the teams, which might be an indicator for a better team spirit. In terms of received points for written reviews German, mixed, and Indian teams outperform the US dominated teams.

5.10.3 Gender

Examining the teams' gender composition, we have observed a slight peak in the dropout rates of teams with a 80:20 male-female ratio. Given a team size of 5-6 members, this represents teams, where one woman is working with an otherwise all-male team. 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*. John, Staubitz, and Meinel [147], have examined these gender-specific phenomena in teams in more detail. At the time of writing, the results of this research have been accepted for publication at the Learning with MOOCs conference 2019.

In *javaeinstieg-mint-ec-2018* it has been observed that all teams that have not submitted anything, have been all-male teams. All teams that have managed to submit something have been mixed teams. Twenty-one of the thirty-three male participants (64%) dropped out of the team task, while only four of the

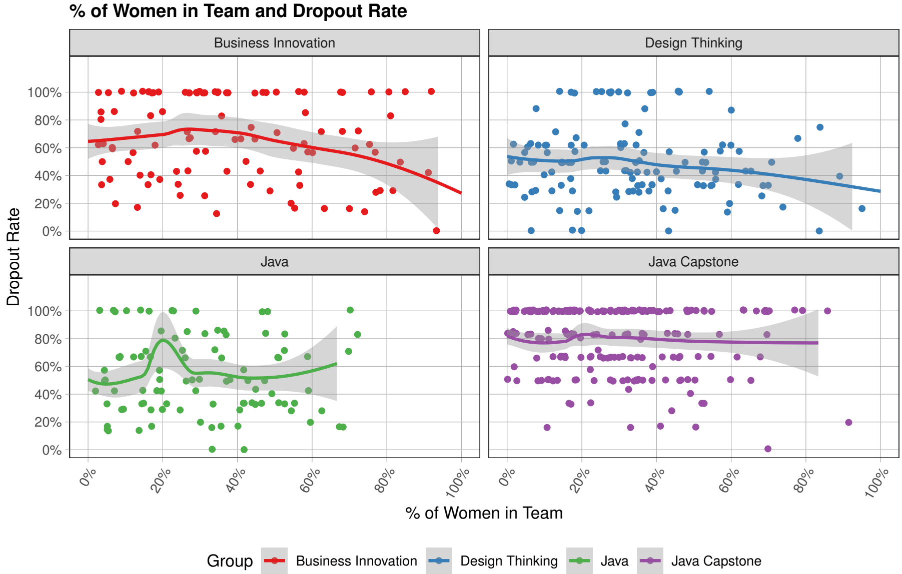

Figure 5.29: Higher dropout rates in teams with an 80:20 male-female ratio. The curve is not a linear regression, but a smoothed conditional means as provided by ggplot2's geom_smooth() method[1].

fourteen female participants (29%) dropped out. These findings are in line with previous empirical studies, which state that on average girls do better in school than boys, as they are particularly diligent and better compliant to our educational system [147].

5.10.4 Initial Team Size

The initial team sizes have been grouped 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 abandon our initial conclusions. The same applies for the investigation of teams with an even vs. odd initial size. So far, the data do not support any evidence if our policy to start with large teams, to provide enough dropout buffer, is not a self-fulfilling prophecy. This still has to be investigated in more detail in the future. Two courses to investigate smaller initial team sizes are already in preparation. One of them will be conducted in

October 2019, the other in February 2020.

5.10.5 Commitment

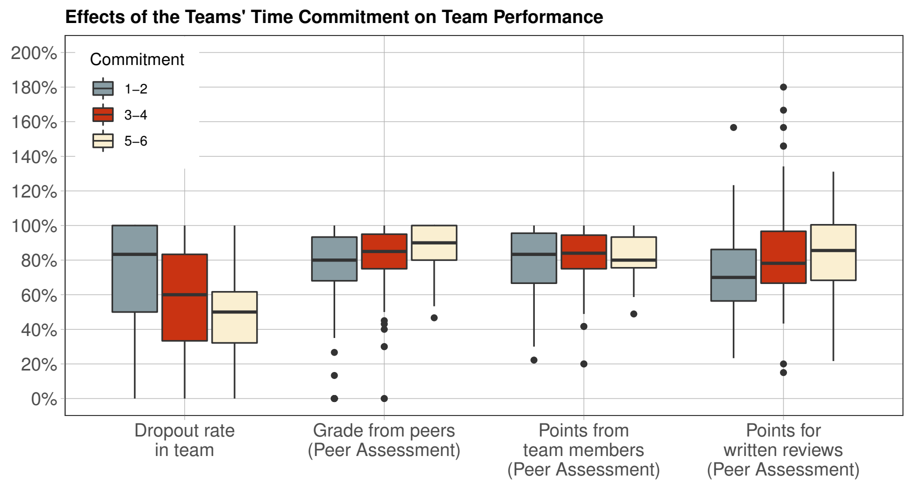

Figure 5.30: Time commitment and team performance (399 teams).

As the surveys have shown, a similar time commitment is a strongly requested and well accepted matching criterion. It was employed as the main matching criterion in *javaeinstieg2017*, *bizmooc2018*, and *java1*. 61% of the examined teams committed to spend 1-2 hours per week on the given task, ~32% committed to 3-4 hours, ~7% committed to 5-6 hours. As expected, the teams with lower time commitment have higher dropout-rates and a lower performance in terms of the grade received from their peers (see Figure 5.30).

The courses in which the time commitment was employed as a matching criterion, have then been compared with those where time commitment has not been used for matching. Surprisingly, the examination of the complete dataset revealed a higher dropout rate in those courses that matched the participants by their time commitment. The data are not suitable to provide significant results for the complete selection, however. Too many other factors are sharing the same dividers and, therefore, invalidate the result. Most of the commitment-matched teams are in the OOP category, one is in the BI category. All the non-commitment-matched teams are in the BI or DT categories. All commitment-matched teams worked on a two-week task while all non-commitment-matched teams worked on a six-week task. It is impossible to tell which of the factors are responsible for the observed difference as all the dividers are aligned. Reducing the dataset to the BI courses only, removes some of the dividers, as now all courses are at least in the same category. Still, the commitment-matched course has a two-week task while the others have a six-week task. Furthermore, this introduces a new divider as now all the non-commitment-matched courses are pre-platform-modification, while the commitment-matched course is post-modification (see also Section 5.12).

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*[1]).

It has been shown in Section 5.6 that participants, who did not score 100% of the PbT, will almost certainly drop out of the team task. Based on the analysis of the data in *javaeinstieg2017* (see Section 5.6), the TeamBuilder's filtering mechanism has been used to deny sixty-five of the 811 registered teamwork participants in *java1* access to the team assignment. Additionally, the *PbT* has been used as a matching criterion for some of the teams. Teams

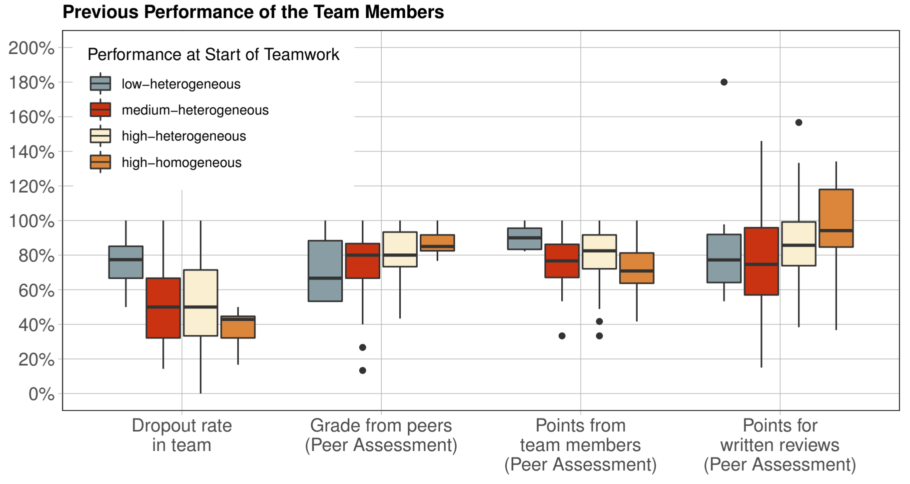

Figure 5.31: Team performance in correlation to the results of the team members in previous course assignments and exercises (PbT) in course *java1* (119 teams).

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, the categories *homogeneous* and *heterogenous* have been defined. In homogeneous teams, the difference

1. Points by the time of team building.

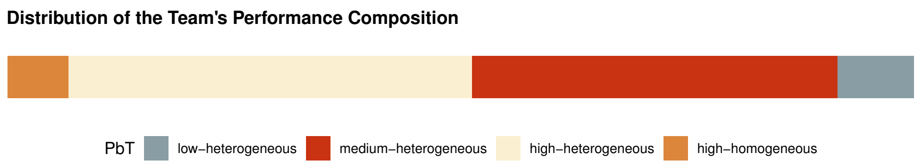

Figure 5.32: Distribution of the teams within the defined categories of previous results (*PbT*) in *java1*.

between the lowest and the highest PBT is max. 10%. Figure 5.32 shows the distribution of 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 5.31 shows the performance results for these categories in comparison. It is no big surprise that the high performers have significantly fewer 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.

5.11 Mentoring

A general difference between most of the courses on openSAP on the one hand, and openHPI/mooc.house on the other hand, is the format or duration of the team projects. Most of the projects in the examined courses on openSAP start right at the beginning of the course. The teams are matched immediately in Week 1, small collaborative assignments have to be handed-in weekly and are binary graded (submitted/not submitted). The teams mostly do not receive feedback for these intermediate submissions. At the end of the course, all these submissions serve as the basis for the team's peer-assessed final project. In the courses on openHPI, the projects either play a minor role within a course, such as e.g., the modeling excursus in the Java courses, or they play a very dominant role as in the workshops. The grading scheme of the workshops often completely relies on these projects. Other than the long-running projects on openSAP, both forms of projects on openHPI have a comparatively short duration (one or two weeks to work on the task, another week for the peer assessment, and yet another week to rate the reviews and digest the results).

Particularly in the courses with long-running team tasks on openSAP, the teams have been supported by mentors. Mostly, the mentors have supported the teams organizationally. In some courses, the mentors have also provided 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, it is interesting to learn how mentoring is perceived or appreciated by the participants and in measuring the effect of mentors on the teams' performance.

In the post-team-task surveys of *javaeinstieg2017*, *javawork2017*, and *bizmooc2018* about 60% of the participants stated that they do not need a mentor. Either they managed well, or teamwork did not happen, and—in their opinion—a mentor would not have improved the situation. About a third of the participants would appreciate pro-bono mentors—volunteering alumni from previous courses (Fig. 5.33). This number seems to be increasing with the level of difficulty of the task to be solved and the importance of the points to be gained in the team assignment for the overall course result. The last statement is mainly based on the survey results in *javawork2017*, but owed to the small number of participants in this survey it is not very reliable. The question was asked slightly

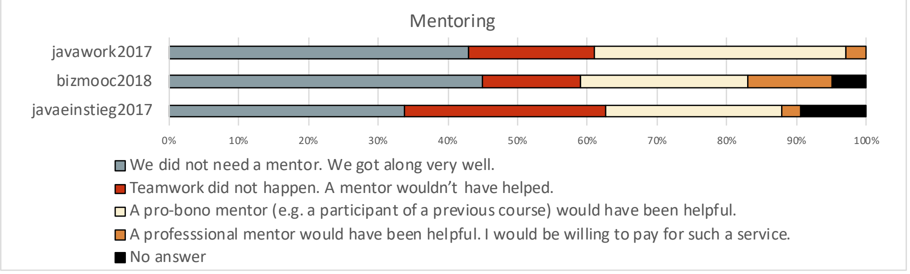

Figure 5.33: Post-teamwork surveys *javaeinstieg2017* (n=340), *bizmooc2018* (n=42), *javawork2017* (n=23) Which statement delineates best your opinion on mentoring for the team tasks?

differently in *java1* and *java-capstone-1* (see Figure 5.34). Again, about 30% of the participants would appreciate mentors, either for organizational support, or for guidance on the quality of their work, or both. About 25% of the participants would appreciate more announcements, particularly about upcoming deadlines. The "Other" option has been missing in *java-capstone-1*, in *java1* most of the participants, who had selected this option explained further that they had no need for any additional support and that everything was just fine. The numbers are analogous in both courses, whereas the team task in *java-capstone-1* had a much higher significance within the course—it was essential for a successful course completion—while in *java1* it was only a bonus exercise that provided a very small number of points.

Then, the differences in the answers about mentoring between users with previous teamwork experience in MOOCs, and those who have been new to this

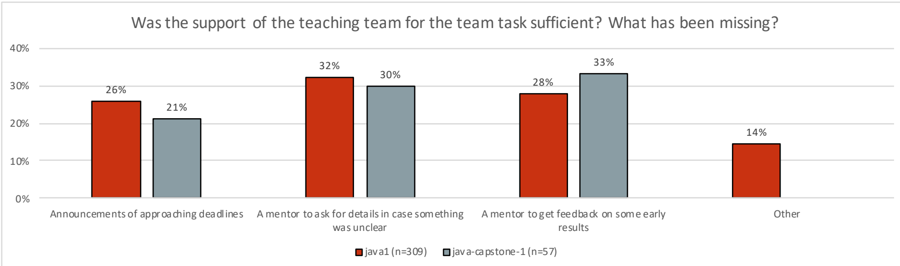

Figure 5.34: Post-teamwork survey *java1* (n=309) and *java-capstone-1* (n=57) Was the support of the teaching team for the team task sufficient? What has been missing? Multiple answers have been possible.

topic have been examined. As the pre-team-task survey did not exist in *java-capstone-1* this comparison has only been conducted on the *java1* data. The total number of participants is a little smaller, as only those who participated in pre- **and** post-surveys have been included (see Figure 5.35). It seems as if more experienced participants are vaguely more interested in additional announcements and mentor feedback. One explanation might be that the other courses in which they have experienced teamwork had also featured mentors to support them (and provided a positive experience). Part one of this explanation is quite probable as it can be assumed that the majority of the participants has gained their previous teamwork experience in an openSAP course—where the team tasks mostly have had mentor support. Part two can then be inferred from part one and the fact that the participants asked for more. However, there is no proof for that; so far it is just a hypothesis. In *javaeinstieg-mint-ec-2018*, we

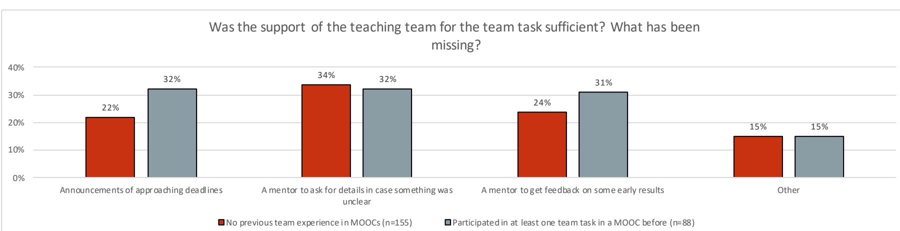

Figure 5.35: Pre-/Post-teamwork survey *java1* (n=243) Was the support of the teaching team for the team task sufficient? What has been missing? Comparing participants with previous experience (n=88) and without previous experience (n=155) in MOOC teams.

tested if more support from the teaching team, such as announcements with instructions how to approach the task or about approaching deadlines, etc., were perceived helpful by the teams. With the current possibilities of the platform, this approach is not scalable but as the number of teams was manageable in this particular course, it was a good opportunity to see whether it is worth the effort to work on a better platform support for the teaching teams here. 50% found these messages helpful and sufficient, another 20% found them helpful and would have liked more. 20% did not find them helpful, and the rest claimed that they never have received such messages. This result is encouraging enough to plan for a better platform support for team announcements, so that it will be possible to keep-up with this practice in large-scale courses.

For measuring the effects of mentors on the teams' performance, an experiment was conducted in *bizmooc2018*. In total twenty-eight teams worked on

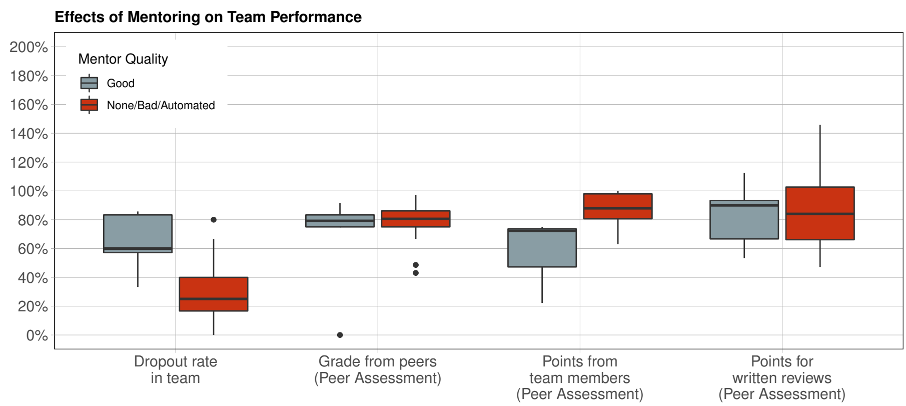

Figure 5.36: Is there a negative effect of mentoring on the teams' performance? (28 teams)

this course's team assignment. Each of the teams was randomly assigned to a different mentor quality. All mentors had an analogous level of knowledge about the task. Five of the teams were supported by active, motivated, human mentors. Another five teams were supported by a fake Robot[1]. Thirteen teams were supported by mentors that did not have time to provide proper support[2]. The remaining five teams did not receive any support at all. The fake Robot only sent regular announcements about upcoming deadlines to the team members and waited for their questions. The dialog was a one way street, however. None of the team members ever asked a question. The mentors who did not have

1. Fake Robot means that the author of this thesis pretended to be an automated tutor in the team forum.
2. This can be assumed to be a quite common situation among voluntary mentors. For example, in the *sbw1* survey, the time the mentors spent with their team has been considered not to be sufficient by many participants (see Figure 5.2).

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 5.36 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 whether 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.

In conclusion, we can state that the experiment somehow confirmed the opinion of the majority in the surveys that mentoring doesn't make much of a difference—at least for the shorter type of assignments. Either the teams managed well, without a mentor, or they didn't do anything and also ignored the mentors' efforts to activate them. However, the examined number of teams was low and the distribution of teams that received "good" mentoring versus teams that received "bad" mentoring was not optimal[1].

To complete the study, the participants have been asked whether they would be willing to serve as pro-bono mentors in future course iterations, and if they would be willing to pay for professional mentoring. 63 participants in *javaeinstieg2017*, 11 in *javawork2017* and 10 in *bizmooc2018* expressed their willingness to volunteer as mentors. From experience, only a fraction of this

1. Unfortunately, it is much easier to find colleagues who volunteer to be mentors with a very limited time contingent, than those that have sufficient time to provide a high quality mentoring. To make a proper statement, this experiment would need to be repeated in a larger course with a better distribution of well mentored and less well mentored teams

number will turn their willingness into action. This number will definitely not be sufficient to offer a dedicated mentor to each of the teams in one of the larger courses. It can be doubted that these mentors would be able to properly support more than one team. Very few participants in these courses have expressed to be willing to pay for a professional team mentor. Only in *bizmooc2018* this group was substantially larger than in the other courses. Mentoring as business model—as it has been established, e.g., at Udacity [148]—therefore, is not an option in the context of team tasks on our platforms. At least as long as the courses are offered for free and without “officially” recognized credentials[1].

5.12 Effects of Platform Modifications

Interviewing the participants of *javaeinstieg2017*, and particularly observing their interactions with the platform’s Collab Spaces, has inspired many ideas to improve the tool-set and the process. Some of the most requested features by the participants have been a text chat and improvements in the platform’s file-sharing abilities. To begin with, however, we have not added new functionalities to the Collab Spaces. Instead,

- ...we removed features that turned out to be rarely used and confusing. For example, TogetherJS[2] irritated the participants more than supporting them, so it was removed.
- ...we added a set of short videos to explain the Collab Spaces, the Peer Assessment and the Team Peer Assessment.
- ...we, generally, improved our communication strategy, and provided more detailed information at an earlier point of time.
- ...we rearranged the Collab Spaces’ navigation menu. Instead of displaying the name of the tools (e.g., Discussion Forum, Etherpad, or Google Hangout), we display a description of the possible activity (e.g., Discuss with your team, Collaborate on texts, Start a video chat.
- ...we added a detailed description to the process of starting a Hangout to keep people from starting Hangouts without scheduling a meeting and getting frustrated when none of the other members joins. Next to the technical side, the description explicitly explains that a meeting needs to be scheduled.

1. This might be either a formal degree, such as e.g., Georgia Tech’s online master’s or a degree that promises improved employability such as Udacity’s nanodegrees
2. https://togetherjs.com/

- ...we informed the participants about video chat alternatives, such as jitsi.org. (So far this information comes on a course by course basis and still needs to be added as a platform feature.)

The effect of these modifications was measured by comparing the team performance results of *javaeinstieg2017* (pre-mod) with the results of *java1* (post-mod1) and *bizmooc2018* (post-mod2). This set of courses was selected 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 roughly 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 analogous. 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. It would have been preferable 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 the examined selection is considered to be the lesser evil.

Instead, the DT and BI courses have been added to Table 5.7 as a reference. Table 5.7 shows that the amount of teams that passed has more than doubled. The amount of no-shows has been reduced significantly. The reference values from the DT and BI courses, however, show that other factors also have a strong influence:

- All 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 owed 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, the platform modifications have to be considered as a success. Figure 5.37 compares the dropout rates and the team performance

Courses	# Teams	Passed	Dysfunct.	Failed	No-shows
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 5.7: Pre-mod (*javaeinstieg2017*) and post-mod (*java1* and *bizmooc2018*) courses. Business innovation (BI) and Design Thinking (DT) courses have been added as a reference.

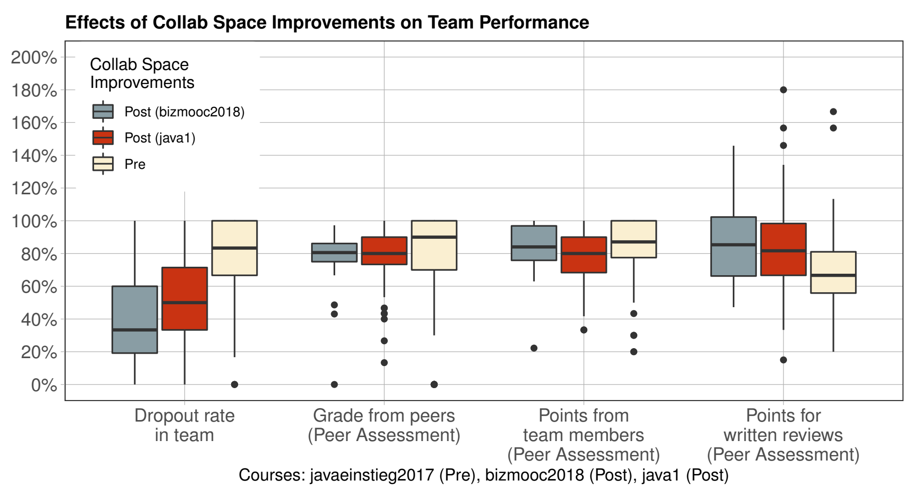

Figure 5.37: Improved team performance in *java1* and *bizmooc2018* (post-modification) compared to *javaeinstieg2017* (pre-modification) (399 teams)

.

in the courses within this 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 5.7 might imply that this is mainly resulting from the decreasing amount of no-shows. Therefore, the start and end sizes of those

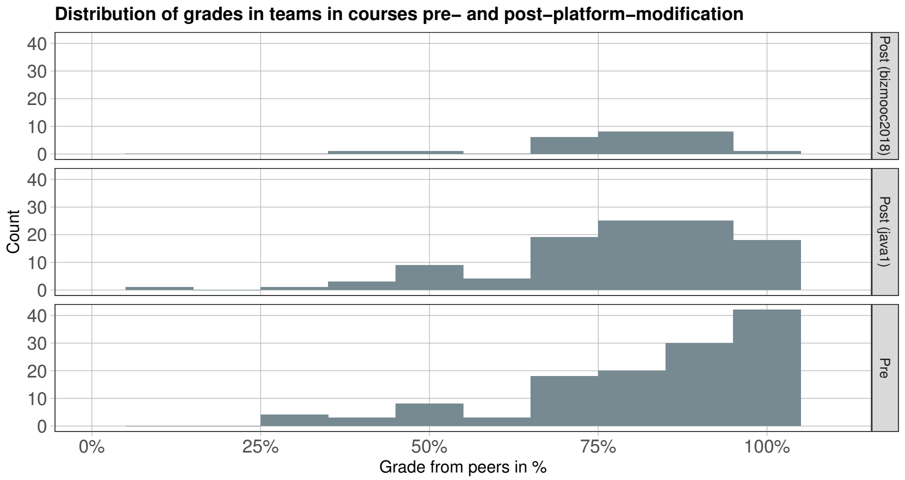

Figure 5.38: Grade distribution in pre- and post-platform-modification courses.

.

teams that have started to work on the task in the pre- and post-modification courses have been compared. While the teams in all courses had 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 received from the 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, cannot really be determined. Exams in schools are often designed and graded to fit the Gaussian normal distribution. Mortensen [149], has shown, however, that this model is not necessarily the best fit. In particular, the distribution of the participants' nature has to be considered. While some sort of bell-curve can be expected, its peak doesn't necessarily have to be in the center. As shown in Section 5.3, only the high performers of the total course population started to work in the teams. The grades for this task can, therefore, expected to be at the upper end of the scale. Figure 5.38 shows that the distribution of grades in the post-modification courses comes close to this expectation, while the grades in the pre-modification course seem to be overly good. 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[1]. While in the pre-mod course only 41%[2] of the

1. Rating a review is optional. Encouraging the participants to make use of this option is one of our goals. So this can be considered to be a success.
2. 41-46% if the business innovation and Design Thinking courses are also included.

participants have received a rating for their review, in the post-mod courses 53-63% of the participants have received a rating. The interviews that have been

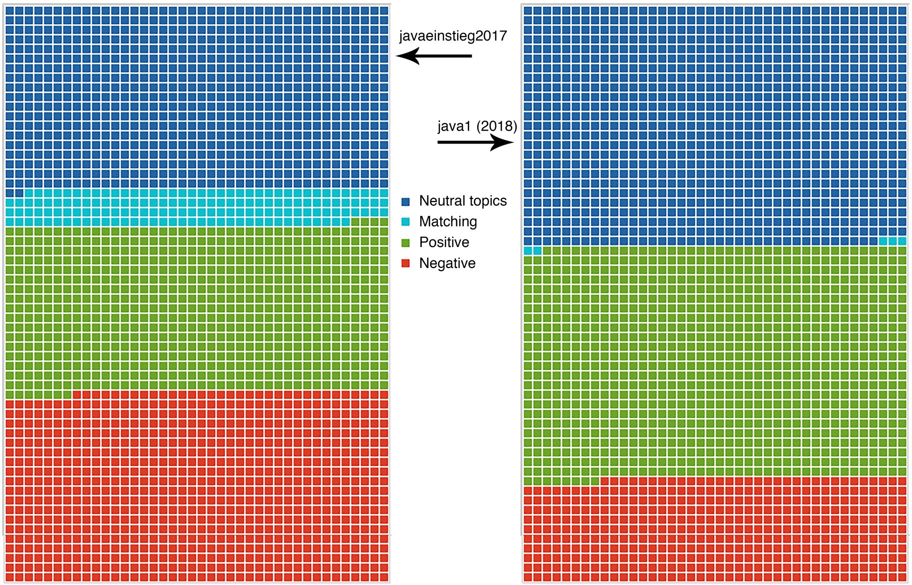

Figure 5.39: Interviews document portrait *javaeinstieg2017* vs. *java1*

.

conducted with the teamwork participants in *javaeinstieg2017* and *java1*, show that there also has been an improvement in how participants perceived the team assignment. The interview summaries have been coded by the author, and the tool MaxQDA was used to analyze these coded summaries. Figure 5.39 shows a comparison of the document portraits[1] of the *javaeinstieg2017* and the *java1* interview summaries. The blue and light blue areas represent neutral codes, such as feature suggestions, or preferences, etc. The green and red areas represent positive or negative perceptions about a list of topics within and surrounding the team assignments. Figure 5.39 shows that the interviewees' perception has changed significantly for the better in between the courses. The code clouds in Figure 5.40 visualize this shift in perception quite well. Unlike the word clouds that have been shown earlier, the code clouds are not created from the words in the actual summaries but from the codings that have been applied to these summaries. MaxQDA allows to add codings in every granularity. A coding can be attached to a complete answer, a single sentence, or even just a word. Multiple codings on the same, or even overlapping, text segments are also

1. In MaxQDA a so-called document portrait can be generated, that gives a very high level comparison of the codings that have been applied on the document.

Figure 5.40: Interviews code cloud *javaeinstieg2017* vs *java1*

.

possible. While these codings do not provide quantitative measurable data, they are very well suited to visualize the captured qualitative data, which is otherwise hard to grasp. While in *javaeinstieg2017* time issues still have been a very dominant topic, generally positive statements dominate in *java1*. The workload of both tasks has been comparable. The reason for this is less a miraculous increase in the participants' time resources, than the result of an improved communication about the time resources that are required to succeed in the task (what needs to be done, when are the deadlines, do the participants have to plan for extra time beyond the official end of the course to finish their work, etc.). To get this right: it is not that these topics have not been communicated in the pre-mod courses. It is more about the intensiveness, the visualization,

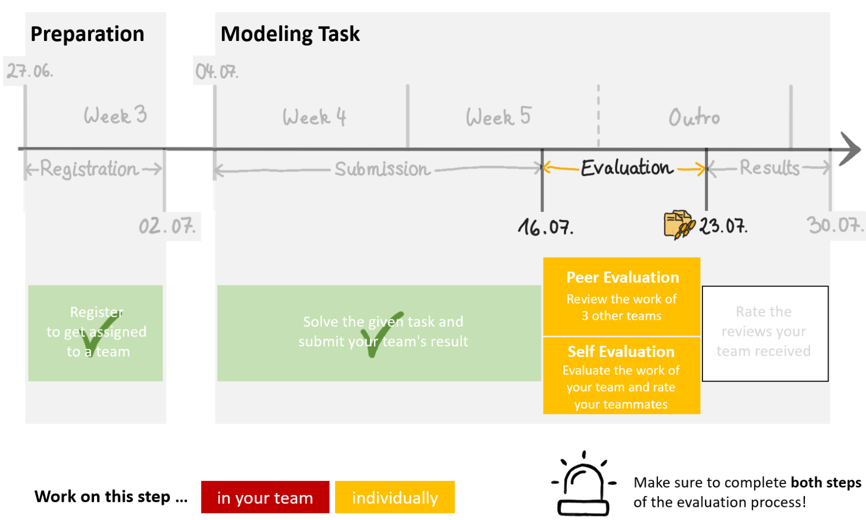

Figure 5.41: One element of the improved communication strategy. Weekly updated timelines to visualize the participant's current position in the workflow

.

the frequency, the visibility, etc. of this communication. Figure 5.41 shows an example for this. Additionally to sending out deadlines in the weekly course announcements, the whole process has been visualized in a timeline that was weekly updated, helping the participants to navigate through this process.

Figure 5.42 lists the codings with a particular positive or negative character in more detail. The percentages on the x-axis do not refer to the amount of interviewees, but to the total text corpus of the interview summaries. The generally positive perception of the team task has strongly increased. Particularly, the interviewees expressed more often explicitly that they did **not** have any issues with the user interface.

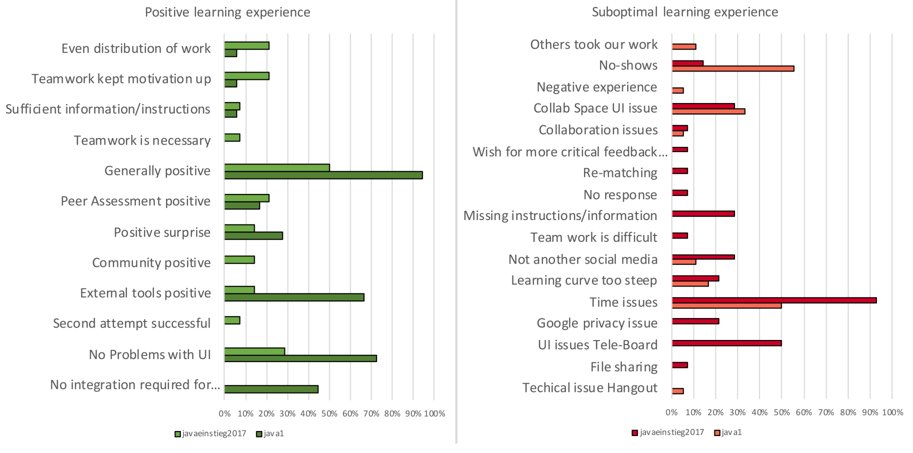

Figure 5.42: Interviews learning experience *javaeinstieg2017* vs *java1*

.

They have been very positive about the use of the external tools. In this case, the difference to *javaeinstieg2017* is not really interesting, as the use of external tools was not explicitely recommended there. We do not have the intention to add special purpose tools, such as an UML diagram editor, to the platform. We would not stand a chance to maintain a collection of such tools for each use case that might come up in the future. One of the interviewees nailed it with the statement “Don’t re-invent the wheel when you already have a running car.” Instead, we have decided to provide more recommendations about existing tools for those participants who are less familiar with finding such tools on their own. Many interviewed participants emphasized that they particularly enjoyed working with these tools, and that they are using them also for projects outside the course context. To some extent this explains the large number of statements that there is no need for a better integration of the external tools into the platform, e.g., via single-sign-on mechanisms. Those who use the tools also outside the course context, registered for them with a separate account anyway. Many of those who contemplated a single-sign-on mechanism to be comfortable, admitted that it would not be worth the effort and encouraged us to put our resources elsewhere.

It is interesting that the statements about no-shows have been discussed more often in *java1*, while they actually were more of an issue in *javaeinstieg2017*. This reveals one of the limitations of this approach. The questions in both interviews have been analogous, but not identical. During the year between the two interview sets many things happened, new questions waited for answers, others already had been answered. Decisions had to be made between comparable results and interviews that address the currently urgent questions. It has been decided averse to the comparability and for the greater benefit for the platform.

Next to the already tackled issues, the surveys and interviews resulted in a long list of improvements, which have been documented in the form of development tickets. Some elements of this list are already being implemented and will be added to the platform in the near future:

- A calendar tool to schedule meetings.
- Improvement of the file sharing possibilities.
- Replacing Etherpad with Libre Office.

The text chat that seemed to be one of the most urgent features to be implemented, after the first set of interviews, has undergone several changes in terms of conceptual design. We originally had planned to attach a separate third party tool that already comes with all the functionality that was requested by the participants (e.g., Rocket.chat). We revised this decision based on the second set of interviews. Overall, it turned out that the discussion forum was the most used communication tool in the majority of the teams. Also, many participants stated that we should keep things as simple as possible and not just add more and more features. Finally, it often turned out that they actually only need notifications on their mobile phones to inform them about new posts in the Collab Space forum and an easy way to reply to these posts from their phone. Therefore, the new plan is to "chatify" the existing discussion forum, e.g., by frequently updating the content to display posts by other participants close to real time. Furthermore, the existing mobile applications will be responsible for providing the notifications and the possibility to react.

6. Future Work

This chapter discusses the issues that still need to be addressed. It differentiates between platform features to be implemented, further research, and upcoming courses that will serve as vehicles for further evaluation.

6.1 Implementation

One of the results of this study is a list of suggestions to further modify the platform's collaborative features, particularly the Collab Spaces. This list has been discussed with openSAP and the HPI School of Design Thinking, the two major stakeholders[1] of these features, and was adjusted to also satisfy their needs. The results of this process are documented in six epics in openHPI's ticket system[2]. Each of these epics contains several feature requests that have been prioritized into three milestones based on their importance, effect, and complexity in terms of implementation. At the time of writing (August 2019), his work is in progress. The first tickets have already been implemented and are currently under review.

Additionally, a lightweight assessment system for intermediate submissions would be a big improvement for courses that feature long term team tasks. Currently, a somewhat clumsy workaround is used for this purpose. Teams upload their intermediate results in their Collab Space's file sharing area. The team mentors check if their team has submitted something useful and if that submission complies with the specified quality requirements. Then, they hand a code-word to the team members, which is the answer for a free-text question in a particular quiz, serving just this purpose. An assessment tool that allows instructors, mentors, or even peers to simply grade such submissions could substantially improve the situation. This tool is also be a candidate to apply at least a certain amount of automation, as a binary grade (passed/failed) would be sufficient and the criteria to make this decision are rather straight-forward.

1. The HPI School of Design Thinking recently started to use the platform to digitally support their on-campus courses.
2. In openHPI's ticket system, epics are the top-level items, mostly containers for a list of smaller tickets plus a description of the general purpose of the idea.

6.2 Further Research

Smaller teams

Currently, we start with a rather large initial team size to compensate for dropouts and no-shows in the teams. This could be a self-fulfilling prophecy, however, as the teams are too large, when all team members become active. Some members might then get the notion that they are actually not really needed and drop out for that reason. In smaller teams, the responsibility of each member is higher, as dropouts and no-shows can not be compensated so easily. To investigate this in more detail, a setup that compares small and large teams within one course is planned for one of the upcoming courses.

One-woman teams

So far, the research revealed that teams in which one woman is facing an otherwise all-male team, might have a higher dropout rate and perform worse than teams with a better gender distribution. The evidence is not very strong, but particularly in the programming courses, is worth a deeper investigation.

Matching based on assignment results

Except for one experiment in *java1*, the participant's course performance prior to the team registration are only used to filter low performers and deny them access to the team task. The results of this experiment suggest, however, that the participants course performance might also be a good matching criterion. A proper recipe, how exactly to match teams based on the participants' performance, still needs to be developed. The following considerations are based on the assumption that the participants' performance in the pre-matching assignments reflects their knowledge and skills in the course topic. Ideally, the difference between the team's top performer and the team's lowest performer should be measurable, but it should not be too big. This way there is a chance that the lower performers can be picked up by the top performers to deliver better results. If the gap is too big, however, the top performers will be stifled while the low performers will be overburdened, which in total is a lose-lose situation.

Mentoring

The mentoring experiment in *bizmooc2018* revealed that the teams that received the best mentoring, performed the worst. This result is very interesting, the data, however, are very weak as only few teams were involved. Repeating the experiment to check if the results can be confirmed will be worth the effort. Additionally, it will be interesting to run an experiment with voluntary team mentors, recruited from successful participants of previous course iterations. The surveys and experiments in this context have not provided consistent, reliable results yet. A field study with real voluntary team mentors in a controlled environment has the potential to show

- ...how reliable voluntary mentors are,
- ...how they can be motivated to provide the best support for their teams,
- ...how helpful they will actually be for the success of the teams.

Further exploration of the datasets

Finally, the datasets that have been generated for the purpose of this thesis still have more potential than what already has been exploited and will be a useful resource for future research.

6.3 Upcoming Courses

For the end of 2019, a course is scheduled that is particularly designed to make use of the collaboration features and the findings of the study in an optimized setting. The course is an "Introduction to Successful Remote Teamwork", so a general motivation to work with the features can be expected among the participants. The focus of this course's team assessment will be on the process of teamwork. The participants will define their goals, work on the task and then self-reflect on how it actually turned out to be. The course team has designed the assessment steps and later found out that they perfectly align with the suggestions of Willcoxson [130]. With this setup, we hope to make a step forward in learning about the inner mechanisms of these teams. Furthermore, two courses in the area of business innovation, both provided by a European research consortium on mooc.house, are currently in the production pipeline. One of them will be an iteration of *bizmooc2018*, which we hope will provide data that can be directly compared to its previous iteration. The other *corship2020* will take a very different teamwork approach; e.g., it will be working with very small teams of two or three members.

7. Conclusion

The central component of this thesis is a long term study that aims to enable, understand, and improve scalable, graded team assignments in MOOCs. A particular challenge in this context is the loose coupling of the participant to the course, as well as to the providing institution. This often results in high dropout rates within a course. Under the premise that a course is delivered following our general quality standards and communication recommendations, we consider the dropout rates, generally, rather as a natural process than as an issue that needs to be addressed. From the platform interaction data we know that many dropouts are actually no-shows. From surveys, help-desk requests, discussions with participants, interviews, and forum posts we know that the majority of the real dropouts have reasons way beyond the realm of our influence. New jobs or projects, family obligations, health issues, accidents and the likes are the main reasons to drop out and these issues cannot be solved by us. In the context of graded team assignments, however, dropouts are a more severe issue, as they can cause a considerable amount of frustration among the remaining team members. Some of the central research questions, therefore, address dropout-issues:

- Can teamwork reduce the attrition rate in MOOCs?
- How can we reduce the attrition rate in the teams?
- Can we predict who will drop out of the team assignment?

Further questions that have been addressed are:

- What is the participants' motivation to register for the team assignments?
- What is the best way to match the team members?
- How should the assignments be graded?

7.1 Setup

Since Summer 2016, graded team assignments have been offered in fifteen courses on openHPI, openSAP, and mooc.house. Course topics have been in the areas of software development, Design Thinking, and business innovation. The number of enrolled participants in these courses ranged from a few hundred in the Design Thinking pilots to ~20,000 in *java1*. The assignments have been problem-based and the team members worked on the tasks between two and six weeks. The study follows an iterative mixed-methods approach and combines user surveys, interviews, and the analysis of data collected in several contexts on the platform.

Three datasets containing the available, anonymized data from various sources within the platform (user profiles, course reports, peer assessment reports, TeamBuilder data, etc.) constitute the centerpiece of the study. Two of these datasets are on the user level—one for the team members and one for those who skipped the team tasks. The third dataset aggregates the user data on the team level. These data have been collected in a productive environment, and, therefore, are often not as clean and unambiguous as if they were created or collected under laboratory conditions. As a tradeoff, however, the data are authentic, and do not leave the human factors aside, which is extremely important in this particular context. To further strengthen this factor, the collected platform data have been enhanced with user surveys and interviews. The research effort took a holistic approach by combining the study with hands-on teaching experience in several examined courses.

7.2 Bias

The user surveys and even more the interviews contain a certain bias. The surveys are voluntary. Post-course surveys are primarily answered by participants who have not dropped out. The interviewees have been recruited from the subset of participants who have submitted the surveys. As it could be expected, mostly participants with a strong positive attitude towards the course, the team, and the platform have participated in the interviews. Nevertheless, the feedback was honest and made us aware of several issues that needed or still need to be addressed.

7.3 Results

To conclude the thesis, the following paragraphs will briefly recapitulate the research questions and the results of the study.

Can team assignments reduce the attrition rate in MOOCs?

For many readers, this will be the most urgent question. Unfortunately, the short and clear answer here is: **NO**. The words of the German playwright

Bertolt Brecht: “Food first, then morality”, summarize the situation quite well. Only participants with a generally good course performance have successfully participated in the team tasks. Participants who have been already struggling in the course, have either not applied for the team task, or dropped out. Depending on the significance of the team-task for the course result, they either just dropped out of the team task or they dropped the course completely.

In most cases, however, the reason for the dropout is neither the team character of an assignment, nor that it is graded by peer assessment, but the project-character of the task— particularly, the higher workload that comes with it and pulls the participants out of their comfort zones. In this context, it has to be emphasized, that the completion rate of a course cannot be the only performance indicator. Project-based tasks will always have a lower completion rate than courses that exclusively rely on multiple-choice exams. Next to the question “How many participants have learned something?”, the questions “How much have the successful participants learned?” and “At which level of Bloom’s taxonomy has this learning occurred?” have to be asked.

None of the surveys or interviews provided evidence for the often-heard claim, that the participants feel isolated in the courses and, therefore, drop out. Quite in contrary, many participants expressed that they are not interested in more social media like features. They enroll in the courses to learn, not to make new friends. An active teaching team with a visible presence in the course forum is absolutely sufficient to overcome any notion of isolation.

Nevertheless, there is a subset of the course population, which is very active and enthusiastic about working on a given task in teams. Team tasks are not a remedy to reduce the attrition rate in MOOCs, they are, however, a great tool to provide additional challenges for the more advanced course participants.

How can we reduce the attrition rate in the teams?

Can we predict who will drop out of the team assignment?

Basically, these questions have been partially answered already. Participants who are struggling with the rest of the course, will drop out of the team task. The most basic measures to reduce the attrition rate in the teams are therefore:

1. ...an open and clear communication strategy about what to expect in the team task, **before** the participants register.
2. ...a restrictive policy who will be allowed for the team task, e.g., based on an introductory test.

Finally, for those who have been admitted, communication is key. Particularly, approaching deadlines and due tasks have to be communicated properly. Further assistance by dedicated team mentors, however, has not shown a positive effect, so far. To make a proper statement here, however, further experiments have to be conducted. The surveys and interviews have shown that the participants are not very interested in extended mentor support. Particularly, there is no business case for paid mentoring options in the current course model on the HPI platforms.

Local teams perform as well as remote teams, the no-show rate in remote teams, however, seems to be higher. Encouraging the teams to get going, generally, is the biggest challenge that needs to be solved to increase the completion rate within the teams.

The data show, that most of the team dropouts are actually no-shows. They register for the task, but then neither get in contact with their teammates nor start working on the task. This observation was confirmed by the surveys and in the interviews. Conflicts within the teams only play a very marginal role in this context.

Generally, the provided collaboration tool-set sufficiently fulfills its purpose. The modifications that have been applied so far, have had a very positive effect. Further modifications are planned. The goal is to extend the functionality, while maintaining the simplicity and improving the ease of use. Special purpose tools to solve certain tasks will not be included in the tool-set. The participants prefer to work with tools that are available outside the course context and can be used for other projects as well.

What is the participants' motivation to register for the team assignments?

The socio-demographic and geographical background of the team workers, generally, mirrors the background of the course population. The basic difference between the team workers and the rest of the course population shows in the course performance. Mostly, the high performing participants apply-for and succeed-in the team tasks. Particularly for bonus assignments, this implies that an intrinsic interest in the task itself is more important than earning points. Most of the team workers already have earned the maximum available points in the course. The bonus points that they earn in the team task will be topped off in their course certificate anyway. There is nothing that binds them to complete the task, except for a genuine interest in the task itself, and maybe a certain responsibility towards their teammates. Increasing the external motivation, by raising the task's significance for the total course result, backfires in most cases. Instead of increasing the task completion, it will very probably increase the course's attrition rate.

What is the best way to match the team members?

Basically, there are three options:

1. A random interventionist approach,
2. A criteria-based interventionist approach,
3. A laissez-faire approach.

In an interventionist approach the instructors build the teams, generally with tool support, while in a laissez-faire approach the participants themselves select the members for their teams. A large body of literature recommends to use a criteria-based interventionist approach for best results. We are following these recommendations and a large majority of participants welcomes this decision. The reason, why they refrain from choosing their own team members, is that they just do not know anyone else in the course and do not want to ask complete strangers to join their team. Owed to the loose coupling of the participant to the course and the providing institution, it is not feasible to work with sophisticated matching criteria, e.g., Belbin tests, as they are used by many of the available team matching tools. The HPI's TeamBuilder is using very basic criteria, such as the users' timezone, background, gender, or age. One of the most important matching criteria is the participants' time commitment, which is also strongly appreciated by the participants.

The teams' dropout rate and performance have been analyzed depending on certain constellations of team composition. Some evidence has been shown that geographically heterogeneous teams perform slightly better than teams that are dominated by members with a certain geographical background. There are also hints in the data that—particularly in programming courses—teams in which one woman is facing an otherwise all-male team should be avoided. This still needs to be investigated in more detail, however.

How should the team assignments be graded?

Scalable grading can either be provided by an automated solution or by peer assessment. Peer assessment has been chosen as it comes with a set of advantages.

- It is general purpose and very flexible,
- It allows to grade complex tasks in the upper levels of Bloom's taxonomy,
- The tasks in these projects, generally, have the necessary complexity level to justify the overhead of a peer assessment,

- By adding the assessment of other solutions as a part of the assignment, peer assessment, basically lifts every assignment to the "Evaluating" level of Bloom's taxonomy,
- As reviews and rating are mandatory to receive points in our implementation of peer assessment, a built-in filter is provided to deny free-riders receiving points for the work of others.
- Literature suggests that team components should be combined with individual components to determine the final grade of a team member. Already in the HPI's original—single user—peer assessment feature, the grade consists of three elements: grade from peers, points for accurate self-assessment, and points for written reviews. For the team peer assessment, it was easy to adjust this. The grade from the peers forms the team component, all others are individual components. Additionally, an in-team-evaluation option has been added as a further individual component.

Generally, the whole process is perceived very well by the participants. The system is based on the assumption that participants who do not contribute to the team's solution will also not review the work of the other teams. Therefore, they do not receive any points and do not benefit from the work of the other team members. Albeit this system is not bullet proof, the comparison of platform data and interviews has shown that at least it worked quite well for a small, but statistically relevant, sample of teams. A further test to validate this assumption on a broader level is scheduled for the end of 2019.

Summary

Gradable Team Assignments in Large Scale Learning Environments
Collaborative Learning, Teamwork, and Peer Assessment in MOOCs

Thomas Staubitz
Hasso Plattner Institute
Internet Technologies and Systems

In times of constant change and ever new technologies, lifelong learning is becoming increasingly important. Massive Open Online Courses or MOOCs are an excellent tool to reach large parts of the population in a short time and with comparatively little effort. With its own openHPI platform and the openSAP, OpenWHO, and mooc.house platforms operated for various partners, HPI makes an important contribution to digital enlightenment both in German-speaking countries and internationally. In many areas, the platform is state of the art and at least on a par with the internationally more well-known platforms. Especially in the development and application of new teaching and learning methods and their technical support, openHPI is also internationally trend-setting. This dissertation deals with the possibilities of technical and didactic support of assessable team tasks in the context of MOOCs. Because of the size of the courses—usually a small teaching team is facing several thousand participants—a manual assessment of the participants' work by the teachers is not possible. Here, one of the alternative possibilities for the evaluation of tasks, the so-called peer assessment, is used and adapted for the special conditions of team assignments. Over the past five years, an iterative long-term study has been carried out, using different qualitative and quantitative methods of evaluation. The result of this research is a deep insight into the mechanisms of teamwork in scalable digital learning platforms, as well as a set of recommendations for improving the collaborative features of the HPI platforms, some of which have already been implemented or currently are being implemented.

Zusammenfassung (German)

Benotete Teamaufgaben in skalierenden E-Learning-Systemen
Kollaboratives Lernen, Teamarbeit und Peer Assessment in MOOCs

Thomas Staubitz
Hasso-Plattner-Institut
Internet-Technologien und Systeme

In einer Zeit stetigen Wandels und immer schneller wechselnder Technologien nimmt das lebenslange Lernen einen immer höheren Stellenwert ein. Massive Open Online Courses (MOOCs) sind ein hervorragendes Werkzeug, um in kurzer Zeit und mit vergleichsweise wenig Aufwand breite Teile der Bevölkerung zu erreichen. Das HPI leistet mit der eigenen Plattform openHPI und den für diverse Partner betriebenen Plattformen openSAP, OpenWHO und mooc.house sowohl im deutschsprachigen Raum als auch international einen wichtigen Beitrag zu digitalen Aufklärung. In vielen Bereichen ist die Plattform State of the Art und ist den international bekannteren Plattformen zumindest ebenbürtig. Gerade bei der Entwicklung und Anwendung von neuen Lehr- und Lernmethoden und deren technischer Unterstützung ist openHPI auch international richtungsweisend. Die vorliegende Dissertation befasst sich mit den Möglichkeiten der technischen und didaktischen Unterstützung von bewertbaren Aufgabenstellungen in MOOCs, die im Team zu bearbeiten sind. Durch die Größe der Kurse—in der Regel steht hier ein kleines Teaching Team mehreren tausend Teilnehmern gegenüber—ist eine manuelle Bewertung der Teilnehmenden durch die Lehrenden nicht möglich. Hier wird eine der alternativen Möglichkeiten zur Bewertung von Aufgaben, das sogenannte Peer Assessment, eingesetzt und für die speziellen Gegebenheiten der Bearbeitung von Aufgaben im Team angepasst. In den vergangenen fünf Jahren wurde eine iterative Langzeitstudie durchgeführt, bei der verschiedene qualitative und quantitative Methoden der Auswertung eingesetzt wurden. Das Ergebnis dieser Forschungsarbeit ist eine tiefgehende Einsicht in die Mechanismen der Teamarbeit in skalierenden digitalen Lernplattformen sowie eine Reihe von Empfehlungen zur weiteren Verbesserung der kollaborativen Eigenschaften der HPI-Plattformen, die zum Teil bereits umgesetzt wurden bzw. gerade umgesetzt werden.

Acknowledgements

I would like to thank...

- First of all Prof. Dr. Christoph Meinel, my supervisor and CEO of the Hasso Plattner Institute for making openHPI possible and always being open for new ideas.
- openHPI team: Christian Willems, Jan Renz, Franz Liedke, Malte Swart, Jan Graichen, Matthias Wiesner, Stefanie Schweiger, Catrina John, Tobias Rohloff, Max Bothe, and Dominic Sauer.
- openHPI Java courses: Ralf Teusner, Ann Katrin Kuessner, Christiane Hagedorn, Lea Gerneth, Alexander Kromer, Robert Stark, Arne Boockmeyer, Steffen Kötte, Stephan Haarmann, and Lukas Faber.
- openSAP Team: Clemens Link, Michaela Laemmler, Katrin Elk, Katarzyna Zakrzewska, Britta Radicke-Schlinsog, and many more.
- BizMOOC: Christian Friedl, Agnieszka Zur, and Thomas Neumann.
- tele-TASK: Matthias Bauer and Martin Malchow.
- DT@Scale: Karen von Schmieden, Mana Taheri, and Lena Mayer.
- openHPI teamwork course: Hanadi Traifeh and Salim Chujfi.
- Master students: Hauke Klement, Dominic Petrick, Tobias Pfeiffer, and Kirstin Heidler.
- My colleague Michaela Schmitz.
- Student assistants: Adrian Kleppat, Konstantin Dobler, Nele Noack, Jascha Beste, Johannes Jasper, and Sebastian Kliem.
- DT@SAP: Romana Oehmig, Sylvie Charpentier, and Andrea Anderson.
- Last but not least, my parents Peter and Hildegund Staubitz for being very supportive throughout my whole life.

References

[1] R.B. Fuller, P.H. Wagschal, and R.D. Kahn. *R. Buckminster Fuller on education*. University of Massachusetts Press, 1979. 1

[2] Stephen Downes. **The MOOC of One**. https://www.slideshare.net/Downes/2014-03-10-valencia/21-DESIGN_PRINCIPLESAutonomy_Choice_of_contents, 2014. [Online; accessed 20-03-2019]. 2

[3] Andrew Ng and Jennifer Widom. **Origins of the Modern MOOC (xMOOC)**. In Fiona M. Hollands and Devayani Tirthali, editors, *MOOCs: Expectations and Reality*, pages 34–48. Center for Benefit-Cost Studies of Education Teachers College, Columbia University, New York, NY, 2014. 2

[4] Anna Hansch, Christopher Newman, and Thomas Schildhauer. **Fostering Engagement with Gamification: Review of Current Practices on Online Learning Platforms**. *HIIG Discussion Paper Series*, **04**, 2015. 3

[5] Laura Pappano. **The Year of the MOOC**. https://www.nytimes.com/2012/11/04/education/edlife/massive-open-online-courses-are-multiplying-at-a-rapid-pace.html, 2012. [Online; accessed 20-03-2019]. 3

[6] Dawal Shah. **By The Numbers: MOOCs in 2018**. https://www.classcentral.com/report/mooc-stats-2018/, 2018. [Online; accessed 22-03-2019]. 3

[7] Justin Reich and José Ruiperez-Valiente. **The MOOC Pivot**. https://www.sciencemagazine digital.org/ sciencemagazine/11_january_2019/MobilePagedArticle.action?articleId=1455537& app=false#articleId1455537, 2019. [Online; accessed 22-03-2019]. 3

[8] Arizona State University. **Global Freshman Academy is now Earned Admission**. https://gfa.asu.edu/, 2019. [Online; accessed 22-03-2019]. 4

[9] Jan Renz, Florian Schwerer, and Christoph Meinel. **openSAP: Evaluating xMOOC Usage and Challenges for Scalable and Open Enterprise Education**. In *Proceedings of the Eighth International Conference on E-Learning in the Workplace*, 0 2016. 4

[10] Tracey Xiang. **Alibaba and Peking University Establish Chinese MOOC Platform**. https://technode.com/2015/02/28/alibaba-peking-university-establish-chinese-mooc-platform/, 2015. [Online; accessed 22-03-2019]. 5

[11] Franka Grünewald, Christoph Meinel, Michael Totschnig, and Christian Willems. **Designing MOOCs for the Support of Multiple Learning Styles**. In Davinia Hernandez-Leo, Tobias Ley, Ralf Klamma, and Andreas Harrer, editors, *Scaling up Learning for Sustained Impact*, **8095** of *Lecture Notes in Computer Science*, pages 371–382. Springer Berlin Heidelberg, 2013. 7, 11, 12, 15, 38

[12] Thomas Staubitz, Jan Renz, Christian Willems, and Christoph Meinel. **Supporting Social Interaction and Collaboration on an xMOOC Platform**. In *EDULEARN14 Proceedings*, 6th International Conference on Education and New Learning Technologies, pages 6667–6677. IATED, 7-9 July, 2014 2014. 11, 12, 13, 54

[13] Thomas Staubitz, Tobias Pfeiffer, Jan Renz, Christian Willems, and Christoph Meinel. **Collaborative Learning in a MOOC Environment**. In *ICERI2015 Proceedings*, 8th International Conference of Education, Research and Innovation, pages 8237–8246. IATED, 18-20 November, 2015 2015. 7, 54

[14] John Dewey. **My Pedagogic Creed**. *Journal of Education*, **101**(18):490–490, 1925. 7

[15] Jerome Bruner. *The Process of Education*. Harvard University Press, 1962. 7, 11

[16] Idit Harel and Seymour Papert. *Constructionism*. Praeger, 1991. 7

[17] Thomas Staubitz, Jan Renz, Christian Willems, Johannes Jasper, and Christoph Meinel. **Lightweight Ad Hoc Assessment of Practical Programming Skills at Scale**. In *Global Engineering Education Conference (EDUCON), 2014 IEEE*, pages 475–483. IEEE, 2014. 9

[18] Martin v. Löwis, Thomas Staubitz, Ralf Teusner, Jan Renz, Christoph Meinel, and Susanne Tannert. **Scaling youth development training in IT using an xMOOC platform**. In *Frontiers in Education Conference (FIE), 2015. 32614 2015. IEEE*, pages 1–9, Oct 2015.

[19] Thomas Staubitz, Hauke Klement, Jan Renz, Ralf Teusner, and Christoph Meinel. **Towards practical programming exercises and automated assessment in Massive Open Online Courses**. In *2015 IEEE International Conference on Teaching, Assessment, and Learning for Engineering (TALE)*, pages 23–30, Dec 2015. 34

[20] Thomas Staubitz, Hauke Klement, Ralf Teusner, Jan Renz, and Christoph Meinel. **CodeOcean - A versatile platform for practical programming exercises in online environments**. In *2016 IEEE Global Engineering Education Conference (EDUCON)*, pages 314–323, April 2016. 14, 34

[21] Thomas Staubitz, Ralf Teusner, Christoph Meinel, and Nishanth Prakash. **Cellular Automata as an Example for Advanced Beginners Level Coding Exercises in a MOOC on Test Driven Development**. *International Journal of Engineering Pedagogy (iJEP)*, **7**(2):125–141, 2017. 35

[22] Thomas Staubitz, Ralf Teusner, and Christoph Meinel. **Towards a repository for open autogradable programming exercises**. In *2017 IEEE 6th International Conference on Teaching, Assessment, and Learning for Engineering (TALE)*, pages 66–73, Dec 2017.

[23] Ralf Teusner, Thomas Hille, and Christiane Hagedorn. **Aspects on finding the optimal practical programming exercise for MOOCs**. In *2017 IEEE Frontiers in Education Conference (FIE)*, pages 1–8, Oct 2017.

[24] Ralf Teusner and Thomas Hille. **On the Impact of Programming Exercise Descriptions**. In *2018 Learning With MOOCS (LWMOOCS)*, pages 51–54, Sep. 2018.

[25] Ralf Teusner, Thomas Hille, and Thomas Staubitz. **Effects of Automated Interventions in Programming Assignments: Evidence from a Field Experiment**. In *Proceedings of the Fifth Annual ACM Conference on Learning at Scale*, L@S '18, pages 60:1–60:10, New York, NY, USA, 2018. ACM. 40

[26] Ralf Teusner, Christoph Matthies, and Thomas Staubitz. **What Stays in Mind? - Retention Rates in Programming MOOCs**. In *2018 IEEE Frontiers in Education Conference (FIE)*, pages 1–9, Oct 2018. 9

[27] Gregory R. Hancock. **Cognitive Complexity and the Comparability of Multiple-Choice and Constructed-Response Test Formats**. *The Journal of Experimental Education*, **62**(2):143–157, 1994. 9

[28] Linda Riebe, Antonia Girardi, and Craig Whitsed. **A Systematic Literature Review of Teamwork Pedagogy in Higher Education**. *Small Group Research*, **47**(6):619–664, 2016. 10, 46

[29] Richard L. Hughes and Steven K. Jones. **Developing and assessing college student teamwork skills**. *New Directions for Institutional Research*, **2011**(149):53–64. 10, 46

[30] Laura Helle, Päivi Tynjälä, and Erkki Olkinuora. **Project-Based Learning in Post-Secondary Education – Theory, Practice and Rubber Sling Shots**. *Higher Education*, **51**(2):287–314, Mar 2006. 10

[31] R. A. Ellis, P. Goodyear, M. Prosser, and A. O'Hara. **How and what university students learn through online and face-to-face discussion: conceptions, intentions and approaches**. *Journal of Computer Assisted Learning*, **22**(4):244–256. 11

[32] Christian Willems, Johannes Jasper, and Christoph Meinel. **Introducing Hands-On Experience to a Massive Open Online Course on openHPI**. In *Proceedings of IEEE International Conference on Teaching, Assessment and Learning for Engineering (TALE 2013)*, Kuta, Bali, Indonesia, 2013. IEEE Press. 11

[33] Hanan Khalil and Martin Ebner. **Interaction Possibilities in MOOCs - How Do They Actually Happen?** In *3rd International Conference on Higher Education Development "Future Visions for Higher Education Development"*, pages 1–24. ., 2013. 12

[34] Carlos Alario-Hoyos, Mar Pérez-Sanagustín, Carlos Delgado-Kloos, Hugo A. Parada G., Mario Muñoz-Organero, and Antonio Rodríguez-de-las Heras. **Analysing the Impact of Built-In and External Social Tools in a MOOC on Educational Technologies**. In Davinia Hernández-Leo, Tobias Ley, Ralf Klamma, and Andreas Harrer, editors, *Scaling up Learning for Sustained Impact*, pages 5–18, Berlin, Heidelberg, 2013. Springer Berlin Heidelberg. 12

[35] Yasmine Kotturi, Chinmay E. Kulkarni, Michael S. Bernstein, and Scott Klemmer. **Structure and Messaging Techniques for Online Peer Learning Systems That Increase Stickiness**. In *Proceedings of the Second (2015) ACM Conference on Learning @ Scale*, L@S '15, pages 31–38, New York, NY, USA, 2015. ACM. 12

[36] Etienne Wenger. *Communities of Practice: Learning, Meaning and Identity*. Cambridge University Press, Cambridge, MA, USA, 1999. 12

[37] Catrina John, Thomas Staubitz, and Christoph Meinel. **Took a MOOC. Got a Certificate. What now?** Accepted, 2019. 13, 15

[38] Bruno Poellhuber, Normand Roy, and Ibtihel Bouchoucha. **Understanding Participant's Behaviour in Massively Open Online Courses**. *The International Review of Research in Open and Distributed Learning*, **20**(1), Feb. 2019. 13, 14

[39] Thomas Staubitz and Christoph Meinel. **Collaborative Learning in MOOCs Approaches and Experiments**. In *2018 IEEE Frontiers in Education Conference (FIE)*, pages 1–9, Oct 2018. 14, 54

[40] Jan Renz, Daniel Hoffmann, Thomas Staubitz, and Christoph Meinel. **Using A/B Testing in MOOC Environments**. In *Proceedings of the Sixth International Conference on Learning Analytics & Knowledge*, LAK '16, pages 304–313, New York, NY, USA, 2016. ACM. 14

[41] C. Alario-Hoyos, M. Perez-Sanagustin, C. Delgado-Kloos, H. A. Parada G, and M. Munoz-Organero. **Delving into participants' profiles and use of social tools in MOOCs**. *IEEE Transactions on Learning Technologies*, **7**(03):260–266, jul 2014. 14

[42] Truong-Sinh An, Christopher Krauss, and Agathe Merceron. **Can Typical Behaviors Identified in MOOCs be Discovered in Other Courses?** In *Proceedings of the 10th International Conference on Educational Data Mining, EDM 2017, Wuhan, Hubei, China, June 25-28, 2017*, 2017. 14

[43] Katy Jordan. **"MOOC Completion Rates: The Data," The Katy Jordan MOOC Project**. http://www.irrodl.org/index.php/irrodl/article/view/2112/3340. [Online; accessed 26-04-2019]. 14

[44] Thomas Staubitz, Christian Willems, Christiane Hagedorn, and Christoph Meinel. **The gamification of a MOOC platform**. In *2017 IEEE Global Engineering Education Conference (EDUCON)*, pages 883–892, April 2017. 15

[45] Raja Gumienny, Lutz Gericke, Matthias Quasthoff, Christian Willems, and Christoph Meinel. **Tele-Board: Enabling efficient collaboration in digital design spaces**. In *Computer Supported Cooperative Work in Design (CSCWD), 2011 15th International Conference on*, pages 47–54, June 2011. 17

[46] Karel Kreijns, Paul A. Kirschner, and Wim Jochems. **Identifying the pitfalls for social interaction in computer-supported collaborative learning environments: a review of the research**. *Computers in Human Behavior*, **19**(3):335 – 353, 2003. 17, 39

[47] Z. Zheng, T. Vogelsang, and N. Pinkwart. **The Impact of Small Learning Group Composition on Student Engagement and Success in a MOOC**. In J. G. Boticario, O. C. Santos, C. Romero, M. Pechenizkiy, A. Merceron, P. Mitros, J. M. Luna, C. Mihaescu, P. Moreno, A. Hershkovitz, S. Ventura, and M. Desmarais, editors, *Proceedings of the 8th International Conference on Educational Data Mining (EDM 2015)*, pages 500–503, Madrid, Spain, 2015. 19, 20, 41, 44

[48] Thomas Staubitz, Dominic Petrick, Matthias Bauer, Jan Renz, and Christoph Meinel. **Improving the Peer Assessment Experience on MOOC Platforms**. In *Proceedings of the Third (2016) ACM Conference on Learning @ Scale*, L@S '16, pages 389–398, New York, NY, USA, 2016. ACM. 21, 54

[49] Felix Raczkowski. **Making points the point – towards a history of ideas of gamification**. https://research.gold.ac.uk/11539/1/9783957960016-rethinking-gamification.pdf. [Online; accessed 11-08-2019]. 34

[50] Kevin Werbach. **Gamification**. https://www.coursera.org/learn/gamification. [Online; accessed 23-10-2016]. 34

[51] Christopher Green. **Classics in the History of Psychology**. http://psychclassics.yorku.ca/Watson/intro.htm. [Online; accessed 11-08-2019]. 34

[52] Burrhus Frederic Skinner. *Walden Two*. Hackett Publishing Company, 1948. 34

[53] Noam Chomsky. **A Review of Skinner's Verbal Behavior**. In L.A. Jakobovits and M.S. Mirón, editors, *Readings in the psychology of language*, Prentice-Hall psychology series, pages 142–143. Prentice-Hall, 1967. 34

[54] Ludy T. Benjamin. **A history of teaching machines**. *American Psychologist*, **43**(9):703–712, 1988. 34

[55] Alex Y. Zheng, Janessa K. Lawhorn, Thomas Lumley, and Scott Freeman. **Application of Bloom's Taxonomy Debunks the "MCAT Myth"**. *Science*, **319**(5862):414–415, 2008. 34

[56] Hoi Suen. **Peer assessment for massive open online courses (MOOCs)**. *The International Review of Research in Open and Distributed Learning*, **15**(3), 2014. 35, 37

[57] Edward.F. Gehringer. **Strategies and mechanisms for electronic peer review**. In *Frontiers in Education Conference, 2000. FIE 2000. 30th Annual*, **1**, pages F1B/2–F1B/7 vol.1, 2000. 35

[58] Linda Knight and Theresa Steinbach. **The pedagogical foundations of massive open online courses**. *Journal of Information Technology Education*, **10**:81–100, 2011. 35, 36, 37

[59] James W. Pellegrino, Naomi Chudowsky, and Robert Glaser, editors. *Knowing What Students Know: The Science and Design of Educational Assessment*. The National Academies Press, Washington, DC, 2001. 35

[60] Stephan Trahasch. **From peer assessment towards collaborative learning**. In *Frontiers in Education, 2004. FIE 2004. 34th Annual*, pages F3F–16–20 Vol. 2, Oct 2004. 35

[61] John Hamer, Kenneth T. Ma, and Hugh H. Kwong. **A Method of Automatic Grade Calibration in Peer Assessment**. In *of Conferences in Research and Practice in Information Technology, Australian Computer Society*, pages 67–72, 2005. 35, 37

[62] Keith Topping. **Peer Assessment between Students in Colleges and Universities**. *Review of Educational Research*, **68**(3):pp. 249–276, 1998. 35

[63] John R. Baird and Jeff R. Northfield. *Learning from the PEEL experience*. School of Graduate Studies, Faculty of Education, Monash University, 1995.

[64] Jan Fermelis, Richard Tucker, and Stuart Palmer. **Online self and peer assessment in large, multi-campus, multi-cohort contexts**. In *Providing choices for learners and learning Proceedings ASCILITE Singapore 2007*, pages 271–281, 2007. 35, 36, 37

[65] Germaine L. Taggart, Sandra J. Jones, and Judy A. Nixon. *Rubrics: A Handbook for Construction and Use*. Rowman & Littlefield Education, 1999. 36

[66] Nancy Pickett and Bernie Dodge. **Rubrics for Web Lessons**. http://webquest.org/sdsu/rubrics/weblessons.htm, 2007. Online; accessed 14-September-2015. 36

[67] E.F. Redish, M. Vicentini, and Società italiana di fisica. *Research on Physics Education*. Number v. 156 in International School of Physics Enrico Fermi Series. IOS Press, 2004. 36

[68] Stephen Bostock. **Student peer assessment**. http://www.reading.ac.uk/web/FILES/engageinassessment /Student_peer_assessment_-_Stephen_Bostock.pdf, 2000. Online; accessed 11-08-2019. 36, 37

[69] David Boud. *Enhancing Learning Through Self-assessment*. Taylor & Francis, 1995. 36

[70] Ken Reily, Pam Ludford Finnerty, and Loren Terveen. **Two Peers Are Better Than One: Aggregating Peer Reviews for Computing Assignments is Surprisingly Accurate**. In *Proceedings of the ACM 2009 International Conference on Supporting Group Work*, GROUP '09, pages 115–124, New York, NY, USA, 2009. ACM. 36, 37

[71] Philip M. Sadler and Eddie Good. **The Impact of Self- and Peer-Grading on Student Learning**. *Educational Assessment*, **11**(1):1–31, 2006. 36, 37

[72] Chinmay Kulkarni, Koh Pang Wei, Huy Le, Daniel Chia, Kathryn Papadopoulos, Justin Cheng, Daphne Koller, and Scott R. Klemmer. **Peer and Self Assessment in Massive Online Classes**. *ACM Trans. Comput.-Hum. Interact.*, **20**(6):33:1–33:31, December 2013. 37, 38

[73] Stephen P. Balfour. **Assessing writing in MOOCS: Automated essay scoring and Calibrated Peer Review**. *Research & Practice in Assessment*, pages 40–48, 2013. 37

[74] David Glance, Martin Forsey, and Myles Riley. **The pedagogical foundations of massive open online courses**. *First Monday*, **18**(5), 2013. 37

[75] Ruiling Lu and Linda Bol. **A comparison of anonymous versus identifiable e-peer review on college student writing performance and the extent of critical feedback**. *Journal of Interactive Online Learning*, pages 100–115, 2007. 37

[76] Martyn Shuttleworth. **Double Blind Experiment**. https://explorable.com/double-blind-experiment, 2008. Online; accessed 11-08-2019. 37

[77] Dilrukshi Gamage, Mark E. Whiting, Thejan Rajapakshe, Haritha Thilakarathne, Indika Perera, and Shantha Fernando. **Improving Assessment on MOOCs Through Peer Identification and Aligned Incentives**. In *Proceedings of the Fourth (2017) ACM Conference on Learning @ Scale*, L@S '17, pages 315–318, New York, NY, USA, 2017. ACM. 37

[78] William Wu, Constantinos Daskalakis, Nicolaas Kaashoek, Christos Tzamos, and Matthew Weinberg. **Game Theory Based Peer Grading Mechanisms for MOOCs**. In *Proceedings of the Second (2015) ACM Conference on Learning @ Scale*, L@S '15, pages 281–286, New York, NY, USA, 2015. ACM. 37

[79] Catherine M. Hicks, C. Ailie Fraser, Purvi Desai, and Scott Klemmer. **Do Numeric Ratings Impact Peer Reviewers?** In *Proceedings of the Second (2015) ACM Conference on Learning @ Scale*, L@S '15, pages 359–362, New York, NY, USA, 2015. ACM. 38

[80] Mehdi S.M. Sajjadi, Morteza Alamgir, and Ulrike von Luxburg. **Peer Grading in a Course on Algorithms and Data Structures: Machine Learning Algorithms Do Not Improve over Simple Baselines**. In *Proceedings of the Third (2016) ACM Conference on Learning @ Scale*, L@S '16, pages 369–378, New York, NY, USA, 2016. ACM. 38

[81] Coursera. **Coursera Peer Assessment Student FAQs**. http://help.coursera.org/ customer/portal/topics/ 521177-peer-assessments, 2014. Online; accessed 3-November-2014. 38

[82] Iversity. **P2P Grading: Assessing Globally**. https: //iversity.org/blog/p2p-grading/, 2014. Online; accessed 4-April-2019. 38

[83] Karthik Raman and Thorsten Joachims. **Methods for Ordinal Peer Grading**. In *Proceedings of the 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD '14, pages 1037–1046, New York, NY, USA, 2014. ACM. 38

[84] Etienne Wenger, Richard McDermott, and William Snyder. *Cultivating Communities of Practice: A Guide to Managing Knowledge*. Harvard Business School Press, Boston, MA, USA, 2002. 38

[85] Etienne Wenger. **Communities of Practice: Learning as a Social System**. https://thesystemsthinker.com/communities-of-practice-learning-as-a-social-system/, 1998. [Online; accessed 11-08-2019]. 38

[86] Art Ellis, Aaron Brower, Ann Burgess, Tony Jacob, Jane Loftus, Cathy Middlecamp, Bob Mathieu, Elaine Seymour, and Leonhard Springer. **Collaborative Learning, Small Group Learning Page**. http://archive.wceruw.org/cl1/CL/default.asp, 1997. [Online; accessed 10-01-2018]. 39

[87] Marjan Laal and Seyed Mohammad Ghodsi. **Benefits of collaborative learning**. *Procedia - Social and Behavioral Sciences*, **31**(0):486 – 490, 2012. World Conference on Learning, Teaching and Administration - 2011. 39

[88] Starr Roxanne Hiltz, Nancy Coppola, Naomi Rotter, Murray Toroff, and Raquel Benbunan-Fich. **Measuring the Importance of Collaborative Learning for the Effectiveness of ALN: A Multi-Measure**. *Online Education: Learning effectiveness and faculty satisfaction*, **1**:101–119, 2000. 39

[89] L. S. Vygotsky. *Mind in society: The development of higher psychological processes*. Harvard University Press Cambridge, Mass., 1978. (Original manuscripts [ca. 1930-1934]). 39

[90] Sugata Mitra and Emma Crawley. **Effectiveness of Self-Organised Learning by Children: Gateshead Experiments**. *Journal of Education and Human Development*, **3**(3):79–88, September 2014. 39

[91] A. Sullivan Palincsar. **SOCIAL CONSTRUCTIVIST PERSPECTIVES ON TEACHING AND LEARNING**. *Annual Review of Psychology*, **49**(1):345–375, 1998. PMID: 15012472. 39

[92] Albert Bandura. *Social learning theory*. General Learning Press, 1971. 40

[93] John R. Anderson, C. Franklin Boyle, and Brian J. Reiser. **Intelligent Tutoring Systems**. *Science*, **228**(4698):456–462, 1985. 40

[94] Derek Sleeman and John Seely Brown. *Intelligent Tutoring Systems*. London : Academic Press, 1982.

[95] ALBERT T. CORBETT, KENNETH R. KOEDINGER, AND JOHN R. ANDERSON. **Chapter 37 - Intelligent Tutoring Systems**. In MARTING G. HELANDER, THOMAS K. LANDAUER, AND PRASAD V. PRABHU, editors, *Handbook of Human-Computer Interaction (Second Edition)*, pages 849 – 874. North-Holland, Amsterdam, second edition edition, 1997.

[96] AMY SOLLER, ALEJANDRA MARTÍNEZ, PATRICK JERMANN, AND MARTIN MUEHLENBROCK. **From Mirroring to Guiding: A Review of State of the Art Technology for Supporting Collaborative Learning**. *Int. J. Artif. Intell. Ed.*, **15**(4):261–290, December 2005. 40

[97] CAROLYN PENSTEIN ROSÉ, OLIVER FERSCHKE, GAURAV TOMAR, DIYI YANG, IRIS HOWLEY, VINCENT ALEVEN, GEORGE SIEMENS, MATTHEW CROSSLIN, DRAGAN GASEVIC, AND RYAN BAKER. **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*, pages 848–851, 2015. 40

[98] JONATHAN MCKINSEY. **Remote Pair Programming in a Visual Programming Language**. Technical Report UCB/EECS-2015-139, University of California Berkeley, Berkeley, 5 2015. 40

[99] JULIA CAMBRE, CHINMAY KULKARNI, MICHAEL S. BERNSTEIN, AND SCOTT R. KLEMMER. **Talkabout: Small-group Discussions in Massive Global Classes**. In *Proceedings of the First ACM Conference on Learning @ Scale Conference*, L@S '14, pages 161–162, New York, NY, USA, 2014. ACM. 40

[100] SEONGTAEK LIM, DERRICK COETZEE, BJOERN HARTMANN, ARMANDO FOX, AND MARTI A. HEARST. **Initial Experiences with Small Group Discussions in MOOCs**. In *Proceedings of the First ACM Conference on Learning @ Scale Conference*, L@S '14, pages 151–152, New York, NY, USA, 2014. ACM. 40

[101] SAIJING ZHENG, MARY BETH ROSSON, PATRICK C. SHIH, AND JOHN M. CARROLL. **Designing MOOCs As Interactive Places for Collaborative Learning**. In *Proceedings of the Second (2015) ACM Conference on Learning @ Scale*, L@S '15, pages 343–346, New York, NY, USA, 2015. ACM. 40

[102] MARGUERITE MCNEAL. **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 11-08-2019]. 41

[103] HAO FEI CHENG, BOWEN YU, YEONG HOON PARK, AND HAIYI ZHU. **ProjectLens: Supporting Project-based Collaborative Learning on MOOCs**. In *Proceedings of the Fourth (2017) ACM Conference on Learning @ Scale*, L@S '17, pages 253–256, New York, NY, USA, 2017. ACM. 41

[104] HAO-FEI CHENG, BOWEN YU, SIWEI FU, JIAN ZHAO, BRENT HECHT, JOSEPH KONSTAN, LOREN TERVEEN, SVETLANA YAROSH, AND HAIYI ZHU. **Teaching UI Design at Global Scales: A Case Study of the Design of Collaborative Capstone Projects for MOOCs**. In *Proceedings of the Sixth (2019) ACM Conference on Learning @ Scale*, L@S '19, pages 11:1–11:11, New York, NY, USA, 2019. ACM. 41

[105] ANNE. **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 11-08-2019]. 41

[106] JANE MANNING. **Comparing MOOC Platform Features**. https://teachingcommons.stanford.edu/teaching-talk/comparing-mooc-platform-features, 2014. [Online; accessed 04-01-2018]. 41

[107] SREECHARAN SANKARANARAYANAN, GAURAV SINGH TOMAR, MIAOMIAO WEN, AKASH BHARADWAJ, AND CAROLYN PENSTEIN ROSÉ. **From Insights to Interventions: Informed Design of Discussion Affordances for Natural Collaborative Exchange**. In *2016 AAAI Spring Symposia, Stanford University, Palo Alto, California, USA, March 21-23, 2016*, 2016. 41

[108] Miaomiao Wen, Diyi Yang, and Carolyn Penstein Rosé. **Virtual Teams in Massive Open Online Courses**. In Cristina Conati, Neil Heffernan, Antonija Mitrovic, and M. Felisa Verdejo, editors, *Artificial Intelligence in Education*, pages 820–824, Cham, 2015. Springer International Publishing. 41, 42

[109] An Ju, Elena Glassman, and Armando Fox. **Teamscope: Scalable Team Evaluation via Automated Metric Mining for Communication, Organization, Execution, and Evolution**. In *Proceedings of the Fourth (2017) ACM Conference on Learning @ Scale*, L@S '17, pages 249–252, New York, NY, USA, 2017. ACM. 42

[110] A. Ju, X. Fu, J. Zeitsoff, A. Hemani, Y. Dimitriadis, and A. Fox. **Scalable Team-Based Software Engineering Education via Automated Systems**. In *2018 Learning With MOOCS (LWMOOCS)*, pages 144–146, Sep. 2018. 42

[111] René F. Kizilcec. **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*, 2013. 42, 43, 94

[112] Seunghoo Chung, Jr. Robert B. Lount, Hee Man Park, and Ernest S. Park. **Friends With Performance Benefits: A Meta-Analysis on the Relationship Between Friendship and Group Performance**. *Personality and Social Psychology Bulletin*, **44**(1):63–79, 2018. 43

[113] Susan Brown Feichtner and Elaine Actis Davis. **Why Some Groups Fail: a Survey of Students' Experiences with Learning Groups**. *Organizational Behavior Teaching Review*, **9**(4):58–73, 1984. 43

[114] Junko Shimazoe and Howard Aldrich. **Group Work Can Be Gratifying: Understanding & Overcoming Resistance to Cooperative Learning**. *College Teaching*, **58**(2):52–57, 2010. 43

[115] Richard A. Layton, Misty L. Loughry, Matthew W. Ohland, and George Dante Ricco. **Design and Validation of a Web-Based System for Assigning Members to Teams Using Instructor-Specified Criteria**. 2010. 43

[116] Miaomiao Wen. *Investigating Virtual Teams in Massive Open Online Courses: Deliberation-based Virtual Team Formation, Discussion Mining and Support*. PhD thesis, Carnegie Mellon University, 8 2016. 43, 87

[117] Rakesh Agrawal, Behzad Golshan, and Evimaria Terzi. **Forming Beneficial Teams of Students in Massive Online Classes**. In *Proceedings of the First ACM Conference on Learning @ Scale Conference*, L@S '14, pages 155–156, New York, NY, USA, 2014. ACM. 43

[118] Zhilin Zheng and Niels Pinkwart. **Dynamic Re-Composition of Learning Groups Using PSO-Based Algorithms**, 2014. 44

[119] René Röpke, Eduard Gallwas, Johannes Konert, and Henrik Bellhäuser. **MoodlePeers: Automatisierte Lerngruppenbildung auf Grundlage psychologischer Merkmalsausprägungen in E-Learning Systemen**. In *Proceedings der 14. E-Learning Fachtagung Informatik der G.I. (DeLFI 2016)*, pages 233–245, Potsdam, Germany, 2016. Köllen Druck+Verlag GmbH. 44

[120] Johannes Konert, Dimitrij Burlak, Stefan Göbel, and Ralf Steinmetz. **GroupAl: ein Algorithmus zur Formation und Qualitätsbewertung von Lerngruppen in E-Learning- Szenarien mittels n-dimensionaler Gütekriterien**. In *Proceedings der Deutschen E-Learning Fachtagung (DELFI 2013)*, pages 71–82. Köllen Verlag, September 2013. 44

[121] Erkan Er, Eduardo Gómez-Sánchez, Miguel Bote-Lorenzo, and Asensio-Pérez. 44

[122] R. Meredith Belbin. *Team Roles at Work*. Routledge, 2 edition, May 2012. 44, 45

[123] Bruce Wayne Tuckman. **Developmental sequence in small groups.** *Psychological bulletin*, **63**(6):384–399, 1965. 45

[124] Bruce W. Tuckman and Mary Ann C. Jensen. **Stages of Small-Group Development Revisited**. *Group & Organization Studies*, **2**(4):419–427, 1977. 45

[125] Tudor Rickards and Susan Moger. **Creative Leadership Processes in Project Team Development: An Alternative to Tuckman's Stage Model**. *British Journal of Management*, (4):273–283, December 2000. 45

[126] Connie JG Gersick. **Time and transition in work teams: Toward a new model of group development**. *Academy of Management journal*, **31**(1):9–41, 1988. 45

[127] Scott D. Johnson, Chanidprapa Suriya, Seung Won Yoon, Jared V. Berrett, and Jason La Fleur. **Team Development and Group Processes of Virtual Learning Teams**. *Computers & Education*, **39**(4):379, 2002. 45

[128] Eberly Center for Teaching Excellence & Educational Innovation. **How can I assess group work?** https://www.cmu.edu/teaching/ designteach/ design/instructionalstrategies/ groupprojects/ assess.html, 2008. [Online; accessed 11-08-2019]. 46

[129] Geoffrey J. Kennedy. **Peer-assessment in Group Projects: Is It Worth It?** In *Proceedings of the 7th Australasian Conference on Computing Education - Volume 42*, ACE '05, pages 59–65, Darlinghurst, Australia, Australia, 2005. Australian Computer Society, Inc. 46

[130] Lesley E. Willcoxson. **"It's not Fair!": Assessing the Dynamics and Resourcing of Teamwork**. *Journal of Management Education*, **30**(6):798–808, 2006. 46, 47, 80, 115

[131] Kali Prasad Nepal. **An approach to assign individual marks from a team mark: the case of Australian grading system at universities**. *Assessment & Evaluation in Higher Education*, **37**(5):555–562, 2012. 46

[132] Edward Gehringer. **Expertiza: Rubric-Based Peer Assessment of Team Projects**. http://slideplayer.com/slide/7773567/, 2015. [Online; accessed 11-08-2019]. 47

[133] Matthew W. Ohland, Misty L. Loughry, David J. Woehr, Lisa G. Bullard, Richard M. Felder, Cynthia J. Finelli, Richard A. Layton, Hal R. Pomeranz, and Douglas G. Schmucker. **The Comprehensive Assessment of Team Member Effectiveness: Development of a Behaviorally Anchored Rating Scale for Self- and Peer Evaluation**. *Academy of Management Learning & Education*, **11**(4):609–630, 2012. 47

[134] Jeremy Roschelle and Stephanie D. Teasley. *The Construction of Shared Knowledge in Collaborative Problem Solving*, pages 69–97. Springer Berlin Heidelberg, Berlin, Heidelberg, 1995. 50

[135] Lynn Power. **Collaboration vs. Cooperation. There Is a Difference.** https://www.huffingtonpost.com/lynn-power/collaboration-vs-cooperat_b_10324418.html, 2017. [Online; accessed 11-08-2019]. 51

[136] Thomas Staubitz and Christoph Meinel. **Collaboration and Teamwork on a MOOC Platform: A Toolset**. In *Proceedings of the Fourth (2017) ACM Conference on Learning @ Scale*, L@S '17, pages 165–168, New York, NY, USA, 2017. ACM. 54

[137] Thomas Staubitz and Christoph Meinel. **Team Based Assignments in MOOCs: Results and Observations**. In *Proceedings of the Fifth Annual ACM Conference on Learning at Scale*, L@S '18, pages 47:1–47:4, New York, NY, USA, 2018. ACM. 54, 55, 56

[138] Thomas Staubitz, Hanadi Traifeh, and Christoph Meinel. **Team-Based Assignments in MOOCs - User Feedback**. In *2018 Learning With MOOCS (LWMOOCS)*, pages 39–42, Sep. 2018. 54, 58

[139] Thomas Staubitz, Ralf Teusner, and Christoph Meinel. **MOOCs in Secondary Education - Experiments and Observations from German Classrooms**. In *2019 IEEE Global Engineering Education Conference (EDUCON)*, 2019. 54

[140] Thomas Staubitz and Christoph Meinel. **Graded Team Assignments in MOOCs: Effects of Team Composition and Further Factors on Team Dropout Rates and Performance**. In *Proceedings of the Sixth (2019) ACM Conference on Learning @ Scale*, L@S '19, pages 5:1–5:10, New York, NY, USA, 2019. ACM. 54, 55, 56

[141] T. Brown. *Change by Design: How Design Thinking Transforms Organizations and Inspires Innovation*. HarperCollins, 2009. 57

[142] Roger L Martin. *The design of business: Why design thinking is the next competitive advantage*. Harvard Business Press, 2009. 57

[143] Jolie Kennedy. **Characteristics of Massive Open Online Courses (MOOCs): A Research Review, 2009-2012**. *Journal of Interactive Online Learning*, **13**(1), 2014. 73

[144] Monica Bulger, Jonathan Bright, and Cristóbal Cobo. **The real component of virtual learning: motivations for face-to-face MOOC meetings in developing and industrialised countries**. *Information, Communication & Society*, **18**(10):1200–1216, 2015. 73

[145] Marianne Krasny, Bryce DuBois, Mechthild Adameit, Ronnie Atiogbe, Lukman Baih, Tergel Bolderdene, Zahra Golshani, Rodrigo González-González, Ishmael Kimirei, Yamme Leung, Lo Shianyun, and Yue Yao. **Small Groups in a Social Learning MOOC (slMOOC): Strategies for Fostering Learning and Knowledge Creation**. *Online Learning*, **22**(2), 2018. 73

[146] Chinmay Kulkarni, Julia Cambre, Yasmine Kotturi, Michael S. Bernstein, and Scott R. Klemmer. **Talkabout: Making Distance Matter with Small Groups in Massive Classes**. In *Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work & Social Computing*, CSCW '15, pages 1116–1128, New York, NY, USA, 2015. ACM. 73

[147] Catrina John, Thomas Staubitz, and Christoph Meinel. **Performance of Men and Women in Graded Team Assignments in MOOCs**. Accepted, 2019. 96, 97

[148] johirbuet. **How I earn more as a mentor in udacity than I was earning before as freelance Software Engineer?** https://steemit.com/education/@johirbuet/how-i-earn-more-as-a-mentor-in-udacity-than-i-was-earning-before-as-freelance-software-engineer. [Online; accessed 30-04-2019]. 105

[149] Uwe Mortensen. **Verteilte Schulnoten sind normal – aber nicht notwendig normalverteilt**. http://www.uwe-mortensen.de/MythosNormalVerteilungC.pdf, 2016. Online; accessed 8-November-2019. 108

Appendices

Appendix I - Courses

The thesis refers to several courses on the HPI platforms by their *course code*. The following list provides details for the courses, alphabetically ordered by these course codes. Additionally, a set of tags was created to map the listed courses to the steps of the study in which they have been involved.

- *sgc-team1*: The courses that have been included in the first iteration of the course data analysis.
- *sgc-team2*: These courses included some sort of survey related to teamwork or peer assessment.
- *sgc-team3*: The second iteration of the course data analysis. At that point, fifteen courses containing team assignments had been delivered on the platforms. Eight of these courses have been examined in this set.
- *sgc-team3-not*: The other seven courses that have been removed from the analysis in *sgc-team3* as they introduced too much noise in the data.

The following terms and abbreviations are used in the list:

- CoP: Confirmation of Participation. To receive a CoP, a participant needs to have seen at least 50% of the course material.
- RoA: Record of Achievement. To receive a RoA, a participant needs to have achieved at least 50% of the available points in the relevant exercises and exams of a course.
- Shows: The listed number is retrieved at **course middle**. Participants who have seen at least one course item. This number serves as the basis for the calculation of the completion rate. Participants who have not joined the course before course middle, generally, stand a lower chance of completing the course as the deadlines of some assignments have already passed. Generally, the show-rate is roughly 50-80% of the enrollments.
- Enrollments: The listed number is retrieved at **course end**. The number of enrollments tends to be growing after course end as the courses are still available in archived mode.
- Completion: The number of issued certificates. The completion rate is based on the shows at course middle.

- Project work:
  - Task: Description of the team task.
  - Deliverables: The artefacts to be handed in by the students.
  - Prerequisites: Some team tasks only required registration, others had some sort of limitations, such as a min. amount of achieved points in previous exams.
  - Grading: The relevance of the team task's grade for the course in total.
  - Participants: Number of participants who registered for the team task. In case the admission was limited, additionally the number of admitted participants.
  - % of total: Percentage of admitted team members from shows at course middle.
  - Milestones: required intermediate submissions

Referenced Courses A–E

***bizmooc2018* (mooc.house)**	
URL	`https://mooc.house/courses/bizmooc2018`
Title	Intrapreneurship - Make your Business great again :-)
Language	English
Term	Feb 26, 2018 to Mar 26, 2018
Shows	1,898 (at course middle)
Enrollments	2,857 (at course end)
Completion	381 participants received a RoA (27% of shows at course middle, 91% of team members.)
Tags	*sgc-team2*, *sgc-team3*
Description	The course introduces the participants to entrepreneurial thinking within a company. The course offered a fast track with videos, quizzes, and interactive elements (e-tivities). It also offered a full track, which added a hands-on, team-based, and peer-assessed project and offered a separate certificate.

Project work:	
Team-based	Yes
Peer-assessed	Yes
Task	The participants had to pitch a business idea within a fictitious company, taking care of business model, stakeholders, sponsors, and the company's profile.
Deliverables	A video or a slide deck presenting the business idea.
Prerequisites	Registration for the team assignment.
Matching	Timezone, selected task, commitment, diverse age, gender and expertise.
Grading	The team-based, peer-assessed task was mandatory for the full track. Full track participants received an additional certificate. The team task accounted for 100% of the points for this certificate.
# Teams	29
Team size	4–6
Participants	156 (8% of shows at course middle)
Duration	The full track and with it the team task started after the fast track has ended in week 4 and went for another two weeks.
Milestones	No
Mentors	Yes. The teams have received mentoring in different quality to explore the effects of mentoring.

Table 1: Course properties *bizmooc2018*

***bmi1-1* (openSAP)**	
URL	`https://open.sap.com/courses/bmi1-1`
Title	Designing Business Models for the Digital Economy (Edition Q3/2017)
Language	English
Term	Sep 05, 2017 to Nov 08, 2017
Shows	2,486 (at course middle)
Enrollments	6,576 (at course end)
Completion	158 participants received a RoA (6% of shows at course middle, 24% of team members.)
Tags	*sgc-team3*

Description	Explores new business models that have emerged in the wake of the digital economy and shows differences to former, well-established business models. Introduces the Business Model Innovation (BMI) approach, to design and iteratively improve business models.
Project work:	
Team-based	Yes
Peer-assessed	Yes
Task	An end-to-end BMI case
Deliverables	A pdf including all the intermediate steps that lead to the final business model.
Prerequisites	Registration for the team assignment.
Matching	Timezone
Grading	50% of the regular points. To obtain a RoA the participation was mandatory.
# Teams	89
Team size	7–8
Participants	654 (26% of shows at course middle)
Duration	6 weeks
Milestones	Weekly. Mandatory but not enforced.
Mentors	No

Table 2: Course properties *bmi1-1*

***bmi1-pilot1* (openSAP)**	
URL	`https://open.sap.com/courses/bmi1-pilot1`
Title	Designing Business Models for the Digital Economy (Pilot Q2/2016)
Language	English
Term	Jun 14, 2016 to Aug 12, 2016
Shows	167 (at course middle)
Enrollments	297 (at course end)
Completion	42 participants received a RoA (25% of shows at course middle, 51% of team members.)
Tags	*sgc-team3-not*
Description	Pilot course for *bmi1-1*
Project work:	
Team-based	Yes
Peer-assessed	No

Task	Several collaborative exercises to define an innovative digital business model.
Deliverables	No final deliverable
Prerequisites	Registration for the team assignment.
Matching	Timezone, Language
Grading	50% of the regular points. To obtain a RoA the participation was mandatory.
# Teams	18
Team size	4–6
Participants	77 (46% of shows at course middle)
Duration	6 weeks
Milestones	Weekly, graded (submitted/not submitted) by team mentors.
Mentors	Yes

Table 3: Course properties *bmi1-pilot1*

***bmi1-pilot2* (openSAP)**	
URL	`https://open.sap.com/courses/bmi1-pilot2`
Title	Designing Business Models for the Digital Economy (Pilot Q1/2017)
Language	English
Term	Jan 17, 2017 to Mar 08, 2017
Shows	92 (at course middle)
Enrollments	156 (at course end)
Completion	29 participants received a RoA (32% of shows at course middle, 66% of team members.)
Tags	*sgc-team3-not*
Description	Pilot course for *bmi1-1*
Project work: Same as bmi1-pilot1. Differences listed below.	
# Teams	9
Team size	4–6
Participants	44 (48% of shows at course middle)

Table 4: Course properties *bmi1-pilot2*

dt1 (openSAP)	
URL	`https://open.sap.com/courses/dt1`
Title	Developing Software Using Design Thinking (Edition Q1/2017)
Language	English
Term	Feb 28, 2017 to Apr 26, 2017
Shows	3,169 (at course middle)
Enrollments	8,155 (at course end)
Completion	258 participants received a RoA (8% of shows at course middle, 54% of team members.)
Tags	*sgc-team1*, *sgc-team3*
Description	The course offers the participants to experience Design Thinking. It is focussing particularly on its application to software development.
Project work:	
Team-based	Yes
Peer-assessed	Yes
Task	Use the Design Thinking process to develop a software application prototype.
Deliverables	Low-resolution prototype and process documentation.
Prerequisites	Only a limited number of those who had achieved 100% of the points in the initial test have been admitted to the team assignment.
Matching	14 local and 47 distributed teams have been created. The distributed teams were located within a certain range of timezones to allow synchronous collaboration.
Grading	55% of the regular points. Mandatory for RoA.
# Teams	61
Team size	6-8
Participants	874 applications, 376 admitted (12% of shows at course middle)
Duration	Started in week 1 and went for 6 weeks.
Milestones	Weekly. Mandatory.
Mentors	Yes

Table 5: Course properties *dt1*

dt1-1 (openSAP)	
URL	`https://open.sap.com/courses/dt1-1`
Title	Developing Software Using Design Thinking (Edition Q2/2017)

Language	English
Term	Jun 13, 2017 to Aug 09, 2017
Shows	2,142 (at course middle)
Enrollments	6,286 (at course end)
Completion	265 participants received a RoA (12% of shows at course middle, 51% of team members.)
Tags	*sgc-team1*, *sgc-team3*
Description	Revised version of *dt1*
Project work: Same as in *dt1*. Differences listed below.	
Prerequisites	Registration for the team assignment. Participation in pre-team-task exam.
Matching	Timezone only.
# Teams	66
Team size	7–8
Participants	559 applications, 512 admitted (24% of shows at course middle)

Table 6: Course properties *dt1-1*

***dt1-2* (openSAP)**	
URL	`https://open.sap.com/courses/dt1-2`
Title	Developing Software Using Design Thinking (Edition Q4/2017)
Language	English
Term	Oct 17, 2017 to Dec 13, 2017
Shows	1,475 (at course middle)
Enrollments	4,866 (at course end)
Completion	179 participants received a RoA (12% of shows at course middle, 53% of team members.)
Tags	*sgc-team3*
Description	Revised version of *dt1-1*
Project work: Same as in *dt1-1*. Differences listed below.	
Prerequisites	Registration for the team assignment.
# Teams	48
Team size	6–8
Participants	333 (23% of shows at course middle)

Table 7: Course properties *dt1-2*

dt1-pilot3 (openSAP)	
URL	`https://open.sap.com/courses/dt1-pilot3`
Title	Developing Software Using Design Thinking (SAP internal)
Language	English
Term	Apr 05, 2016 to Jun 03, 2016
Shows	160 (at course middle)
Enrollments	251 (at course end)
Completion	74 participants received a RoA (46% of shows at course middle, 81% of team members.)
Tags	*sgc-team3-not*
Description	Pilot for the Design Thinking courses. Similar to *dt1*.
Project work: Same as in *dt1*. Differences listed below.	
Team-based	Yes. First course that employed the TeamBuilder.
Peer-assessed	No. The teamwork was assessed by the mentors.
Prerequisites	Registration for the team assignment.
Matching	6 local teams. 13 distributed teams matched by timezone and language.
Grading	50% of the regular points. Mandatory for RoA.
# Teams	19
Team size	4–6
Participants	88 (55% of shows at course middle)

Table 8: Course properties *dt1-pilot3*

dt1-pilot4 (openSAP)	
URL	`https://open.sap.com/courses/dt1-pilot4`
Title	Developing Software Using Design Thinking (Edition Q4/2016)
Language	English
Term	Oct 18, 2016 to Dec 14, 2016
Shows	197 (at course middle)
Enrollments	305 (at course end)
Completion	62 participants received a RoA (31% of shows at course middle, 54% of team members.)
Tags	*sgc-team1*, *sgc-team3-not*
Description	Revised version of *dt1-pilot3*

Project work: Same as *dt1*. Differences listed below.	
Matching	10 local teams. 10 distributed matched by timezone, language, and heterogeneous expertise.
# Teams	20
Team size	3–6
Participants	94 (48% of shows at course middle)

Table 9: Course properties *dt1-pilot4*

Referenced Courses F–J

***java1* (openSAP)**	
URL	`https://open.sap.com/courses/java1`
Title	Object-Oriented Programming in Java
Language	English
Term	Jun 13, 2018 to Jul 26, 2018
Shows	9,504 (at course middle)
Enrollments	21,693 (at course end)
Completion	2,318 participants received a RoA (24% of shows at course middle, 89% of team members.)
Tags	*sgc-team2*, *sgc-team3*
Description	Learn the basics of Object-Oriented-Programming in Java, including an excursus on Object-Oriented Modeling.
Project work:	
Team-based	Yes
Peer-assessed	Yes
Task	Create a UML model for a given setting.
Deliverables	UML model, a glossary of the used identifiers, and a code scaffold that implemented the model.
Prerequisites	Only participants who have achieved ˜80% of the points in the course's previous assignments have been admitted.
Matching	Time zone, commitment, community, diverse age, background and gender.

Grading	The task was an optional bonus assignment and covered 8% of the regular points.
# Teams	119
Team size	6-8
Participants	810 applications, 745 admitted (8% of shows at course middle)
Duration	Started in week 4 and had a duration of 2 weeks.
Milestones	No
Mentors	No

Table 10: Course properties *java1*

***java-capstone-1* (openHPI)**	
URL	`https://open.hpi.de/courses/java-capstone-1`
Title	Java Capstone Series Pt. 1
Language	English
Term	Sep 10, 2018 to Sep 24, 2018
Shows	1,786 (at course middle)
Enrollments	2,327 (at course end)
Completion	66 participants received a RoA (4% of shows at course middle, 18% of team members.)
Tags	*sgc-team3-not*
Description	Follow-up for *java1* (on openSAP). The participants learned some basics about design patterns and git. The main part of the course was to work in teams or alone on a given software project.
Project work:	
Team-based	Yes. Alternatively, the participants had the option to work on the same task alone.
Peer-assessed	Yes
Task:	Develop a game in Java, use github to collaborate on the code.
Deliverables	Detailed documentation of the development process (lab-report) and a documentation of the delivered game.
Prerequisites	Registration for the team assignment.
Matching	Timezone, selected task, commitment, diversity of age, gender, and expertise.
Grading	100% of the regular points. To obtain a RoA the participation was mandatory.
# Teams	32
Team size	5–9

Participants	222 (12% of shows at course middle)
Duration	Started in week 1 and had a duration of 2 weeks.
Milestones	Encouraged but not mandatory.
Mentors	No. But the teams received basic support from the teaching team via particular announcements.

Table 11: Course properties *java-capstone-1*

***javaeinstieg2017* (openHPI)**	
URL	`https://open.hpi.de/courses/javaeinstieg2017`
Title	Objektorientierte Programmierung in Java (An Introduction to Object-Oriented Programming in Java)
Language	German
Term	Mar 27, 2017 to May 14, 2017
Shows	6,610 (at course middle)
Enrollments	10,402 (at course end)
Completion	2,124 participants received a RoA (32% of shows at course middle, 73% of team members.)
Tags	*sgc-team1*, *sgc-team2*, *sgc-team3*
Description	The participants learn the basics of object-oriented-programming in Java. The course follows an objects-first approach. After learning about the absolute basics of the Java syntax, the participants start to write classes and instantiating them. Then the basic programming mechanisms such as conditions and loops are discussed. After that, more advanced OOP topics, such as encapsulation and inheritance are discussed.
Project work:	
Team-based	Yes
Peer-assessed	Yes
Task	Model a small Java application in an UML-like structure.
Deliverables	A class-diagram and a glossary explaining the model's most important elements.
Prerequisites	Registration for the team assignment.
Matching:	Timezone, time commitment, diversity of age, gender, and expertise. Re-arranging dysfunctional teams has been attempted in a few cases.
Grading	The task was an optional bonus assignment and covered 6% of the regular points.
# Teams	251

Team size	6-7
Participants	1514 (23% of shows at course middle)
Duration	The team-based assignment started towards the end of the course and had a duration of about two weeks.
Milestones	No
Mentors	No

Table 12: Course properties *javaeinstieg2017*

javaeinstieg-mint-ec-2018 **(openHPI)**	
URL	`https://open.hpi.de/courses/javaeinstieg-mint-ec-2018`
Title	Objektorientierte Programmierung in Java - Schul-Cloud-Edition 2018
Language	German
Term	Feb 26, 2018 to Jun 11, 2018
Shows	373 (at course middle)
Enrollments	462 (at course end)
Completion	83 participants received a RoA (22% of shows at course middle, 91% of team members.)
Tags	*sgc-team2*, *sgc-team3-not*
Description	In terms of content the same as *javaeinstieg2017*. The course has been stretched from 4 weeks to 3 months to reduce the weekly workload and fit better in the school context.
Project work: Same as *javaeinstieg2017*	
Matching	Teams have been formed locally–within certain schools (5)–or distributed–with members in schools all over Germany (5). The participants had the choice which of both they preferred. Three separate teams for teachers.
# Teams	13
Team size	4–6
Participants	62 (17% of shows at course middle)
Mentors	No. The teams received particular support via additional announcements from the instructors.

Table 13: Course properties *javaeinstieg-mint-ec-2018*

javawork2017 (openHPI)	
URL	`https://open.hpi.de/courses/javawork2017`
Title	Java Workshop: Einführung in eine Java-Programmierumgebung (IDE)(An Introduction to a Java Programming IDE)
Language	German
Term	May 01, 2017 to May 15, 2017
Shows	1,481 (at course middle)
Enrollments	4,336 (at course end)
Completion	194 participants received a RoA (13% of shows at course middle, 65% of team members.)
Tags	*sgc-team1*, *sgc-team2*, *sgc-team3-not*
Description	This course extended the course *javaeinstieg2017* with a two-week workshop to apply the previously learned competences in a programming project.
Project work:	
Team-based	Yes. Alternatively, the participants had the option to work on the same task alone.
Peer-assessed	Yes
Matching	The participants have been asked to form teams on their own.
Participants	824 applications, 267 admitted (5% of shows at course middle)
Task	The participants had to complete the Java code of a small game.
Deliverables	Project with compiling and running Java code. Short documentation.
Grading	100.00% of the regular points. To obtain a RoA the participation was mandatory.
# Teams	22
Team size	2
Participants	43 (3% of shows at course middle)
Duration	Started in week two and had a duration of 2 weeks.
Notes	188 out of 463 participants (40%), who started to work on the assignment alone submitted a solution. 34 of 40 (85%) of those who worked in a team submitted a solution.

Table 14: Course properties *javawork2017*

Referenced Courses P–T

***sbw1* (openSAP)**	
URL	`https://open.sap.com/courses/sbw1`
Title	Enabling Entrepreneurs to Shape a Better World
Language	English
Term	May 30, 2016 to Aug 05, 2016
Shows	5,088 (at course middle)
Enrollments	11,664 (at course end)
Completion	967 participants received a RoA (19% of shows at course middle, 85% of team members.)
Tags	*sgc-team1*, *sgc-team2*, *sgc-team3*
Description	Enables the participants to develop a new mindset and to use field-tested tools and techniques for designing viable, impactful solutions for a better world, based on self-sustaining business models.
Project work:	
Team-based	Yes
Peer-assessed	Yes. First graded peer-assessed, team-based assignment on any of the HPI platforms.
Task	Develop an idea for a social innovation in the context of migration in Mexico, in China, or in Europe.
Deliverables	Pdf document containing a pitch for the innovation including a business model.
Prerequisites	Registration for the team assignment. Participants had to have earn a minimum of 90% of the points in the previous tasks. Only a limited number of those have been admitted.
Matching	Timezone, selected topic.
Grading	33.33% (optional Special Track)
# Teams	39
Team size	6-7
Participants	824 applications, 267 admitted (5% of shows at course middle)
Duration	Started in week 1 and had a length of 7.5 weeks.
Milestones	The teams had the chance to earn bonus points by submitting weekly intermediate steps.
Mentors	Yes

Table 15: Course properties *sbw1*

Appendix II - Interview Summaries

Tables 16 and 17 give an overview on the interviewed participants in *javaeinstieg2017*. All names have been pseudonymized. The complete structured interview summaries are available as a CSV file (data/interviews.csv), and can be found on the attached CD.

Interviews 2017 (*javaeinstieg2017*)

Name	Gender	Age	Interview Type	Date	Country	Team Success
Bert	male	40-49	V-chat	25.07.17	Germany	passed
Martha	female	10-19	V-chat	31.07.17	Germany	passed
Stefan	male	50-59	V-chat	23.08.17	Germany	passed
Norbert	male	10-19	F-to-F	24.08.17	Germany	passed
Gerhard	male	50-59	V-chat	24.08.17	Germany	passed
Markus	male	10-19	F-to-F	30.08.17	Germany	passed
Uwe	male	40-49	F-to-F	30.08.17	Germany	passed
Dorian	male	30-39	V-chat	31.08.17	Germany	dysfun.
Nino	male	50-59	F-to-F	12.09.17	Germany	passed
Vilmos	male	50-59	V-chat	13.09.17	Germany	dysfun.
Elias	male	40-49	V-chat	13.09.17	Germany	passed
Claus	male	40-49	V-chat	18.09.17	Germany	passed
Doris	female	30-39	V-chat	21.09.17	Germany	passed
Nicole	female	50-59	V-chat	22.09.17	Germany	passed

Table 16: Interview participants 2017 (pseudonymized). F-to-F: The participant has visited us at the HPI for a face-to-face meeting. V-chat: A video conference via Google Hangouts or Skype.

Interviews 2018 (*java1*)

Name	Gender	Age	Interv. Type	Date	Country	Team Success
Gennaro	male	50-59	V-chat	29.10.18	Chile	passed
Berta	female	40-49	V-chat	29.10.18	Germany	passed
Nina	female	40-49	V-chat	30.10.18	Germany	passed
Holden	male	60-69	V-chat	30.10.18	Ireland	dysfun. (frd)
Klaas	male	40-49	V-chat	30.10.18	Netherlnd.	passed
Pratyush	male	20-29	V-chat	30.10.18	India	passed
Tanvi	female	30-39	V-chat	01.11.18	India	passed
Nestor	male	20-29	V-chat	01.11.18	Peru	passed
Norbert	male	50-59	V-chat	02.11.18	Germany	passed
Iago	male	20-29	V-chat	02.11.18	Brazil	passed
Riya	female	40-49	V-chat	02.11.18	India	passed
Kaelen	male	50-59	V-chat	06.11.18	Brazil	passed
Daria	female	50-59	V-chat	06.11.18	Romania	dysfun.
Dieter	male	50-59	V-chat	06.11.18	Germany	passed
Albert	male	40-49	V-chat	08.11.18	Netherlnd.	passed
Sidell	male	40-49	Email	19.11.18	Venezuela	passed
Domen	male	50-59	V-chat	19.11.18	Slovenia	passed
Konstantin	male	40-49	V-chat	30.11.18	Germany	passed

Table 17: Interview participants 2018 (pseudonymized). V-chat: Video conference, Google Hangouts or Skype. Email: A video chat was not possible for technical reasons. Questions have been sent and answered by email. Dysfun.: dysfunctional (only 1 member left in the team at the end of the task). frd: free-riders detected.

Appendix III - Surveys

The tables in this appendix list all survey questions that are relevant for this thesis in their original wording. Questions that allowed multiple answers are marked as (MA).Questions that allowed open answers are marked as (OA). All other questions allow only one answer option. The questions have been asked in the in the following surveys:

- *sbw1*, Post course survey
- *javaeinstieg2017*, Post-team-task survey
- *javawork2017*, Post-course survey
- *bizmooc2018*, Pre-team-task survey
- *bizmooc2018*, Post-team-task survey
- *java-mint-ec-2018*, Post-team-task survey
- *java1*, Pre-team-task survey
- *java1*, Post-team-task survey
- *java-capstone-1* Post-team-task survey

***sbw1*: 2016 Post-Course Survey** (n=57)	
Q1:	Generally, how satisfied have you been with the Team Challenge?
A1:	7-point Likert scale (1-7)
Q2:	To what extent do you agree with the following statements?
A1	The Team Challenge tasks were relevant. (7-point Likert)
A2	The Team Challenge tasks were manageable. (7-point Likert)
A3	The Team Challenge tasks were suitable for virtual team work. (7-point Likert)
A4	I enjoyed working in my team. (7-point Likert)
A5	The support our team received from our mentor was valuable for our team project. (7-point Likert)
A6	The time the mentor spent with our team was sufficient. (7-point Likert)

A7	I was able to acquire important know-how through the Team Challenge tasks. (7-point Likert)

Table 18: Survey questions *sbw1* post-course survey

javaeinstieg2017: **2017 Post-Team-Task Survey** (n=340, German)	
Q1:	Waren die angebotenen Tools für die gemeinsame Bearbeitung von Artefakten ausreichend (Etherpad, Tele-Board)?
A1	Wir haben ausschließlich die in den Lerngruppen angebotenen Kollaborationstools benutzt und sind damit gut zurechtgekommen.
A2	Wir haben ausschließlich die in den Lerngruppen angebotenen Kollaborationstools benutzt, diese reichen aber bei weitem nicht aus.
A3	Wir haben zusätzlich zu den in den Lerngruppen angebotenen Kollaborationstools die Tools anderer Anbieter genutzt bzw. mit analogen Methoden (z.B. Papier und Stifte) gearbeitet.
A4	Wir haben in unserem Team die in den Lerngruppen angebotenen Kollaborationstools nicht genutzt und hauptsächlich mit den Tools anderer Anbieter oder analogen Methoden gearbeitet.
Q2:	Waren die angebotenen Tools für die Kommunikation im Team ausreichend (Teamforum, Hang-outs, TogetherJS, Chat im Etherpad)?
A1	Wir haben ausschließlich die in den Lerngruppen angebotenen Kommunikationskanäle benutzt und sind damit gut zurechtgekommen.
A2	Wir haben ausschließlich die in den Lerngruppen angebotenen Kommunikationskanäle benutzt, diese reichen aber bei weitem nicht aus.
A3	Wir haben zusätzlich zu den in den Lerngruppen angebotenen Kommunikationskanälen die Tools anderer Anbieter genutzt.
	(Weitere Erläuterungen (z.B. welche Tools habt ihr benutzt) oder Wünsche zu der vorhergehenden Frage.) (OA)
Q3:	Welche Aussagen bezüglich eines möglichen Mentors für die Teamarbeit beschreibt eure Meinung am besten?
A1	Ein “professioneller” Mentor, wäre sehr hilfreich gewesen. Ich wäre bereit für einen solchen Service zu bezahlen.
A2	Einen Mentor haben wir nicht gebraucht, wir kamen sehr gut zurecht.
A3	Ein “ehrenamtlicher” Mentor, z.B. ein Teilnehmer eines früheren Kurses wäre sehr hilfreich gewesen.
A4	Eine Teamarbeit fand überhaupt nicht statt. Ein Mentor hätte auch nichts genutzt.

Q4:	Hättest du dir dein Team lieber selbst gesucht anstatt einem Team zugeteilt zu werden?
A1	Ja
A2	Nein
	(Wenn du die vorhergehende Frage mit **Ja** beantwortet hast, wie wärst du dabei vorgegangen, nach welchen Kriterien hättest du deine Teammitglieder ausgewählt?) (OA)

Table 19: Survey questions *javaeinstieg2017* post-team-task survey

javawork2017: **2017 Post-course Survey** (n=23, German)	
Q1:	Teamarbeit: Waren die angebotenen Tools für die Kommunikation im Team ausreichend (Teamforum, Hang-outs, TogetherJS, Chat im Etherpad)?
A1	Wir haben ausschließlich die in den Lerngruppen angebotenen Kommunikationskanäle benutzt und sind damit gut zurechtgekommen.
A2	Wir haben ausschließlich die in den Lerngruppen angebotenen Kommunikationskanäle benutzt, diese reichen aber bei weitem nicht aus.
A3	Wir haben zusätzlich zu den in den Lerngruppen angebotenen Kommunikationskanälen die Tools anderer Anbieter genutzt.
A4	Ich habe die Aufgabenstellung alleine bearbeitet.
	(Weitere Erläuterungen (z.B. welche Tools habt ihr benutzt) oder Wünsche zu der vorhergehenden Frage) (OA)
Q2:	Teamarbeit: Welche Aussagen bezüglich eines möglichen Mentors für die Teamarbeit beschreibt eure Meinung am besten?
A1	Einen Mentor haben wir nicht gebraucht, wir kamen sehr gut zurecht.
A2	Eine Teamarbeit fand überhaupt nicht statt. Ein Mentor hätte auch nichts genutzt.
A3	Ein "ehrenamtlicher" Mentor, z.B. ein Teilnehmer eines früheren Kurses wäre sehr hilfreich gewesen.
A4	Ein "professioneller" Mentor, wäre sehr hilfreich gewesen. Ich wäre bereit für einen solchen Service zu bezahlen.
Q3:	Teamarbeit: Hättest du eher im Team gearbeitet wenn wir die anderen Teammitglieder für dich ausgesucht hätten?
A1	Ja
A2	Nein

	(Wenn du die vorhergehende Frage mit Ja beantwortet hast, welches wären für dich die wichtigsten Kriterien für ein erfolgreiches Team?) (OA)

Table 20: Survey questions *javawork2017* post-course survey

***bizmooc2018*: 2018 Pre-team-task Survey** (n=158)	
Q1:	How comfortable are you when collaborating with others on a task?
A1	Very comfortable
A2	Comfortable
A3	Neutral (OK)
A4	More or less comfortable
A5	I prefer to work alone
Q2:	How many times did you work in a team when attending a MOOC?
A1	Never
A2	1-3 times
A3	4-6 times
Q3:	Please rank your previous experience when you've worked with a team in a MOOC:
A1	I haven't worked in a team before
A2	very good
A3	good
A4	neutral
A5	disappointing
A6	very disappointing
Q4:	What do you expect from your teammates? Please choose the most important statement among the following.
A1	a high level of commitment
A2	an equal contribution of time and effort
A3	respecting my time
A4	being able to communicate my ideas/challenges in a safe environment
A5	being able to have fun with my teammates

Table 21: Survey questions *bizmooc2018* pre-team-task survey

bizmooc2018: **2018 Post-team-task Survey** (n=42)	
Q1:	Have the provided collaboration tools (text collaboration tool, brainstorming tool) been sufficient to jointly work on the task?
A1	We only used the tools that were offered in the Collab Spaces and managed to work on the task very well.
A2	We only used the tools that were offered in the Collab Spaces, these are in no way sufficient, however.
A3	Additionally to the tools that were offered In the Collab Spaces, we used the tools of other providers or worked with offline tools, such as paper and pen.
A4	We haven't used the offered tools at all and mostly worked with the tools of other providers or offline.
	(Further explanations (e.g. which tools did you use) or requests concerning the previous question.) (OA)
Q2:	Have the provided communication tools (team forum, video conference, chat within the text collaboration tool)? been sufficient for a proper communication in the team?
A1	Additionally to the existing tools, we used the tools of other providers.
A2	We only used the communication channels that were offered in the Collab Spaces.
A3	This worked well for us. We only used the communication channels that were offered in the Collab Spaces.
A4	These communication channels are by no means sufficient.
	(Further explanations (e.g. which tools did you use) or requests concerning the previous question.) (OA)
Q3:	Which statement in terms of a mentor for the teams meets your opinion best?
A1	A professional mentor would have been very helpful.
A2	I would be willing to pay for such a service.
A3	We didn't need a mentor, we got along very well.
A4	A pro-bono mentor, e.g. a participant of a previous iteration of the course would have been very helpful.
A5	Team work did not happen at all. A mentor wouldn't have helped either.
Q4:	Would you have preferred to select a team on your own instead of being assigned to a team by us?
A1	Yes

A2	No
	(If you have answered the previous question with yes, what would have been your criteria to select team members. How would you have done that?)(OA)

Table 22: Survey questions *bizmooc2018* post-team-task survey

***javaeinstieg-mint-ec-2018*: 2018 Post-team-task Survey** (n=26, German)	
Q1:	Wie habt ihr die Aufgabe bearbeitet? (Welche Aussage trifft am ehesten zu.)
A1	Wir haben alles gemeinsam bearbeitet.
A2	Wir haben uns besprochen und dann die Aufgaben aufgeteilt. (Glossar, Diagramm, Code, etc.)
A3	Wir haben uns besprochen und dann die Aufgaben aufgeteilt. (Bad, Küche, Bibliothek, etc.)
A4	Ich habe die Aufgabe größtenteils alleine bearbeitet. Die anderen haben mir (teilweise) Feedback gegeben.
A5	Ich habe die Aufgabe ganz alleine bearbeitet.
A6	Jemand anderes hat die Aufgabe hauptsächlich bearbeitet. Ich habe nur kleine Teile beigetragen.

Table 23: Survey questions *javaeinstieg-mint-ec-2018* post-team-task survey

***java1*: 2018 Pre-team-task Survey** (n=833)	
Q1:	How comfortable are you when collaborating with others on a task?
A1	very comfortable
A2	comfortable
A3	neutral (OK)
A4	more or less comfortable
A5	I prefer to work alone
Q2:	How many times did you work in a team when attending a MOOC?
A1	Never
A2	1-3 times
A3	4-6 times

Q3:	What do you expect from your teammates? Please choose the most important statement among the following.
A1	a high level of commitment
A2	an equal contribution of time and effort
A3	respecting my time
A4	being able to communicate my ideas/challenges in a safe environment
A5	being able to have fun with my teammates

Table 24: Survey questions *java1* pre-team-task survey

java1: 2018 Post-team-task Survey (n=309)	
Q1:	Which of the following tools did you miss the most to support the work in your team? (You can select as much as you want. You should restrict yourself, however, to those that would solve a real need in your team rather than just being a nice to have.) (MA)
A1	Video-chat (other than Google Hangouts)
A2	Meeting planner (Doodle or similar)
A3	Collaborative text-editor with better formatting options.
A4	Collaborative brainstorming tool
A5	A tool that shows which of the team members are currently online.
A6	A team calendar Announcements about upcoming deadlines, etc.
A7	Other
	(If you selected "Other", you can add more details here.)(OA)
Q2:	What would be your preferred matching criterion for the teams?
A1	Similar time commitment. (How much time is a participant willing to spent on the given team task)
A2	Diversity of age, professional background, gender.
A3	Homogeneous age, background, gender.
A4	Local teams that can meet face to face.
A5	Previous interaction in the discussion forum.
A6	Same mother tongue
A7	Other
	(If you selected "Other", you can add more details here.)(OA)

Q3:	If you did submit the task, how many team members have still been there in the end?
A1	I finished the task on my own.
A2	Two of us finished the task.
A3	Three of us finished the task.
A4	Four of us of us finished the task.
A5	Five of us of us finished the task.
A6	Six of us of us finished the task.
A7	Seven (or more) of us finished the task.
Q4:	How did the team size diminish?
A1	The others never showed up.
A2	One by one they lost interest or lacked the time.
A3	All of a sudden the others were gone.
A4	We had many conflicts within the team.
A5	Other
	(If you selected "Other", you can add more details here.) (OA)
Q5:	How did you work on the task?
A1	We mostly collaborated. We discussed and did most of it together.
A2	We mostly cooperated (divide and conquer). One of us did the diagram, another one the glossary, etc.
A3	We mostly cooperated (divide and conquer). One of us did the printer, another one the furniture, etc.
A4	I did most of the work alone. The others gave feedback. (Or someone else did most of the work and you only gave feedback.)
A5	I did all the work alone.
Q6:	Which of the following was your main communication tool?
A1	Offline face-to-face meetings
A2	Discussion forum in the Collab Space
A3	Google Hangout in the Collab Space
A4	Alternative video chat, such as Jitsi or similar
A5	Private Chat / Whatsapp / Messenger / Skype, etc.
A6	No communication
A7	Other

	(If you selected "Other", you can add more details here.) (OA)
Q7:	Was the support of the teaching team for the team task sufficient? What has been missing?
A1	Announcements of approaching deadlines
A2	A mentor to ask for details in case something was unclear
A3	A mentor to get feedback on some early results
A4	Other
	(If you selected "Other", you can add more details here.) (OA)

Table 25: Survey questions *java1* post-team-task survey

java-capstone-1: 2018 Post-team-task Survey (n=57)	
Q1:	Which of the following tools did you miss the most to support the work in your team? (You can select as much as you want. You should restrict yourself, however, to those that would solve a real need in your team rather than just being a nice to have.) (MA)
A1	Text-based chat or messenger
A2	Video-chat (other than Google Hangouts)
A3	Meeting planner (Doodle or similar)
A4	Collaborative text-editor with better formatting options.
A5	Collaborative brainstorming tool
A6	A tool that shows which of the team members are currently online.
A7	A team calendar.
A8	Announcements about upcoming deadlines, etc.
A9	Other
	(If you've selected "Other" in the previous question, you can add details here.) (OA)
Q2:	What would be your preferred matching criterion for the teams?
A1	Similar time commitment. (How much time is a participant willing to spent on the given team task)
A2	Diversity of age, professional background, gender.
A3	Homogeneous age, background, gender.
A4	Local teams that can meet face to face.
A5	Previous interaction in the discussion forum.

A6	Same mothertongue
A7	Other
	(If you've selected "Other" in the previous question, you can add details here.) (OA)
Q3:	If you did submit the task, how many team members have still been there in the end?
A1	I finished the task on my own.
A2	Two of us finished the task.
A3	Three of us finished the task.
A4	Four of us of us finished the task.
A5	Five of us of us finished the task.
A6	Six of us of us finished the task.
A7	Seven (or more) of us finished the task.
Q4:	How did the team size diminish?
A1	The others never showed up.
A2	One by one they lost interest or lacked the time.
A3	All of a sudden the others were gone.
A4	We had many conflicts within the team.
A5	Other
	(If you've selected "Other" in the previous question, you can add details here.) (OA)
Q5:	How did you work on the task?
A1	We mostly collaborated. We discussed and did most of it together.
A2	We mostly cooperated (divide and conquer). One of us did the diagram, another one the glossary, etc.
A3	We mostly cooperated (divide and conquer). One of us did the printer, another one the furniture, etc.
A4	I did most of the work alone.
A5	The others gave feedback. (Or someone else did most of the work and you only gave feedback.)
A6	I did all of the work alone.
Q6:	Which of the following was your main communication tool?
A1	Offline face-to-face meetings

A2	Discussion forum in the Collab Space
A3	Google Hangout in the Collab Space
A4	Alternative video chat, such as Jitsi or similar
A5	Private Chat / Whatsapp / Messenger / Skype, etc.
A6	No communication
A7	Other
	(If you've selected "Other" in the previous question, you can add details here.) (OA)
Q7:	Was the support of the teaching team for the team task sufficient? What has been missing?
A1	Announcements of approaching deadlines
A2	A mentor to ask for details in case something was unclear
A3	A mentor to get feedback on some early results

Table 26: Survey questions *java-capstone-1* post-team-task survey

Appendix IV - Aggregated Team Data

The following list delineates each column of the aggregated team data in detail. It includes the method that has been used to aggregate the data for each column. If no particular method is listed, the value has been identical for each team member. The complete dataset is available as a CSV file and can be found on the attached CD (data/teams_aggregated.csv).

Column	**Description**	**Possible values or *Example***
Platform	The platform offering the course (openHPI, openSAP, mooc.house)	HPI, SAP, MOC
CLNG	Course language	en, de
Course	Course code as on platform	*java1*
Team	Team name as on platform	*Team #44*
SizeStart	Number of team members when the teamwork started	2-10
SizeEnd	Number of team members who have received points for the teamwork (→assessed the work of other teams)	0-10
Mentor	Identifier for the mentor (pseudonymized)	*Robert the Robot*
FE	First enrollment. The percentage of platform newbies in a team.	0-1
EDD	Average enrollment delta in days. Positive values: the number of days that a participant has enrolled after course start. Negative values: the number of days that a participant has enrolled before course start.	*-120, 28*
LNG	The percentage of team members who have set the course language as their default language (very rough approximation of native speakers)	0-1

AgeComposition	The composition of the team members in terms of age. • unknown: age data are available for less than half of the team members • homogeneous: max. two adjacent age groups • heterogeneous: three or more different age groups, or non-adjacent age groups • < 40: the majority is less than 40 years old • > 40: the majority is more than 40 years old • even: the age groups are evenly distributed Possible age groups are: 10-19, 20-29, 30-39, 40-49, 50-59, 60-69, >70.	*heterogeneous<40*, *heterogeneous even*, *homogeneous>40*
Timezone	The composition of team members in terms of time zone. • heterogeneous: difference between max. and min. <= 3h • homogeneous: difference between max. and min. >= 3h	heterogeneous, homogeneous
TimezoneRange	“smallest” and “largest” time zone in team (JSON)	*[[-5,"min"],[-3,"max"]]*
Expertise	The distribution of expertise in the team. • homogeneous: if team size is > 2 and # expertise < 3 or if team size equals 2 and # expertise < 2 • heterogeneous: if not homogeneous • unknown: if info is available for less than half of the team members	heterogeneous, homogeneous, unknown
Expertise Summary	The detailed numbers for each expertise in the team (JSON)	*[[0,"Other"], ..., [3,"Media"]]*
Gender	Percentage of female team members (unknown: if info is available for less than half of the team members)	*0-1*

Commitment	The majority of selected commitments (unknown: if info is available for less than half of the team members)	*3-4*
Commitment Summary	The detailed numbers, which time commitments have been made how often (JSON)	*[[0,"1-2"],[0,"3-4"],[7,"5-6"]]*
Career	The majority of career stages within a team. (unknown: if info is available for less than half of the team members)	teacher, student, professional, academic researcher, heterogeneous, unknown
CareerSummary	The detailed numbers, which career stages are represented within a team. (JSON)	*[[1,"professional"], ..., [2,"student"]]*
Degree	The majority of (highest achieved) degrees in a team (unknown: if info is available for less than half of the team members).	high-school, bachelor, master, professional, academic researcher, heterogeneous, unknown
DegreeSummary	The detailed numbers, which career stages are represented within a team (JSON)	*[[2,"professional"], ..., [0,"other"]]*
BIT	Background in IT. Majority in team (unknown: if info is available for less than half of the team members).	unknown, beginner, advanced, expert
BITSummary	The detailed numbers, which backgrounds in IT are represented within a team.	*[[1,"beginner"], ..., [0,"expert"]]*
Experience	Work experience in years. Majority in team (unknown: if info is available for less than half of the team members).	up_to_five_years, up_to_10_years, more_than_10_years, unknown
ExperienceSummary	The detailed numbers, which work experience is represented within a team.	*[[0,"up_to_5_years"], ..., [1,"up_to_10_years"]]*
Position	Position in job. Majority in team (unknown: if info is available for less than half of the team members).	technician, team_leader, project_manager, department_head, intern, other
PositionSummary	The detailed numbers, which positions are represented within a team.	*[[1,"technician"], ..., [0,"other"]]*

Countries	Geographical background. Majority in team (unknown: if info is available for less than half of the team members).	Two letter ISO country codes
CountriesSummary	The detailed numbers, which countries are represented within a team.	*[[3,"US"], ..., [0,"(Other)"]*
PbT	Points before team building. The points that the participants have received in all assignments where the deadline has passed before the team building process was started. Median points (percentage) in team (very low(<30%) low(<50%), medium(50-80%), high(80-100%)) (unknown: if info is available for less than half of the team members).	high, medium, low, very low
PbTDist	The distribution of the grades within the team. • homogeneous: difference between max. and min. <= 10% • heterogeneous: difference between max. and min. > 10%	homogeneous, heterogeneous
Posts	The average amount of forum posts per team member.	*7.42857*
Threads	The average amount of forum threads started per team member.	*1.75*
FActivity	The average forum activity (active and passive) per team member.	*117.25*
Daily_Forum_Activity	The average forum activity (active and passive) per team member and day.	*2.5581*
IVP	Average percentage of visited items.	*0.839*
PP	Average percentage of achieved points.	*0.69875*
Sessions	Average amount of sessions.	*67.25*
ASD	Team average of average session duration.	*663.25*
TSD	Team average of total session duration.	*6274527.7556*
QP	Average quiz performance (Percentage).	*0.80285714*
Att_min	Course attendance. When has the last item been accessed by a team member. Percentage of available course items. Smallest value in team.	*0.441860465*

Att_max	When has the last item been accessed by a team member. Percentage of available course items. Largest value in team.	*1*
Att_mean	When has the last item been accessed by a team member. Percentage of available course items. Average value in team.	*0.9102990*
Att_Summary	Complete list of course attendance values of team members.	*[0.7906976744, 1, 1, 1, 1, 1, 1]*
CoP	Average percentage of issued Confirmations of Participation (CoP) in team.	*0.875*
RoA	Average percentage of issued Records of Achievement (RoA) in team.	*0.75*
CC	Average percentage of team members who completed the course. Either by getting a RoA or by clicking the "Course completed" button.	*0.857142857*
Quantile	Average quantile of team members. Members with a quantile of 0.95 or more are among the Top 5% of the course participants.	*0.857142857*
Top	The maximum position (Top5, Top10, Top20, None) that a team member has achieved.	Top5, Top10, Top20, None
State	The state of the team at the end of the task. • passed: at least two team members have finished the team task including reviewing the work of other teams. • dysfunctional: only one team member has finished the team task including reviewing the work of other teams. • failed: one or more team members have started to work on the task, but none of them finished • not started: none of the team members has started to work on the task.	passed, dysfunctional, failed, not started
GfP	Grade from peers. The grade that the team has received for its submission by the reviewing peers. Percentage.	0-1

BPRev	Bonus points for reviews. The average amount of bonus points that the team members have received from the teams they have reviewed. Percentage of those who have successfully completed the task.	0-1
BPSA	Bonus points for self-assessment. The average amount of bonus points that the team members have received for their self-assessment. Percentage of those who have successfully completed the task.	1, NA
BPT	Bonus points for team assessment. The average amount of bonus points that the team members have received from their teammates for their contribution, social, and organizational skills. Percentage of those who have successfully completed the task.	0-1
SR	Star rating from expert reviewers. The sum of all stars that the team members have received from the expert reviewers.	24
Nmn	The average amount of nominations for the gallery that a team has received from their peers.	24
Phase	The peer assessment phase in which the majority of the team members have arrived. • Results: the participant has submitted a solution and has assessed the work of other teams. • PeerGrading: the participant has submitted a solution (or was dragged along by her teammates). She has not, however, assessed the work of other teams and, therefore, does not receive any points. • unknown: the participant has not started the peer assessment.	Results, PeerGrading, unknown, (AssignmentSubmission, SelfAssessment)
PhaseDist	The complete list of final peer assessment phases for the team members (JSON)	*[[1,"PeerGrading"], ..., [4,"Results"]]*
IVPSec1-12	Percentage of visited items per course section. NA if Section did not exist in course.	0-1

PPSec1-12	Percentage of achieved points per course section. NA if section did not exist or did not have graded quizzes.	0-1
local	Local teams have been able to meet face to face.	Local, Remote, unknown

Table 27: Aggregated team data

Appendix V - Team Member Data

The following list delineates each column of the team members data in detail. The complete dataset is available as a CSV file and can be found on the attached CD (data/teams_participant_data.csv).

Column	**Description**	**Possible values or *Example***
Platform	The platform that offered the course (openHPI, openSAP, mooc.house)	HPI, SAP, MOC
CLNG	Course language	en, de
Course	Course code as on platform	*java1*
UserHash	Anonymized user ID	alphanumeric, lowercase, length: 64
ED	Enrollment date. When did the participant enroll for this course?	Time stamp
FE	First enrollment. Was it the participant's first course on this platform?	TRUE, FALSE
UserCreated	When did the participant register on the platform?	Time stamp
LNG	The user's preferred language. TeamBuilder data if available, otherwise platform settings	*en*
Age	The user's age. Data from user profile.	*75*
AgeMerged	The user's age. Profile and TeamBuilder data merged.	10-19, 20-29, ..., 70-110
EDD	Enrollment delta in days. Positive values: the number of days that a participant has enrolled after course start. Negative values: the number of days that a participant has enrolled before course start.	*-120, 28*

Posts	The number of the participant's forum posts in the course.	*10*
Threads	The number of forum threads that the participant has started in the course.	*5*
CoP	Confirmation of participation. Did the participant earn one.	TRUE, FALSE
RoA	Record of achievement. Did the participant earn one.	TRUE, FALSE
CC	Course completed. Participant defines if she completed a course.	TRUE, FALSE
Quantile	Quantile of team members. Members with a quantile of 0.95 or more are among the Top 5% of the course participants	0-1
Top	The position (Top5, Top10, Top20, None) that the participant has achieved.	Top5, Top10, Top20, None
IVP	Items visited percentage. Percentage of visited items, compared to available items.	0-1
PP	Percentage of achieved points	0-1
Team	Team name as on platform	*Team #44*
Mentor	Pseudonym for team mentor	*Robert the Robot*
PbT	Percentage of points before team building. The points that the participant has received in all assignments where the deadline has passed before the team building process was started.	0-1
Timezone	The timezone in which the participant is located (TeamBuilder)	*GMT+01:00*
Task	The task that was selected by the participant (TeamBuilder)	*Value Consult - Compensation for Success.*
Commitment	The time that a participant has committed to spent on the team task per week. (TeamBuilder)	1-2, 3-4, 5-6

Expertise	The participant's background. (TeamBuilder)	Business and Management, Computer Science and Engineering, Creative Design, Humanities, Life Science, Media, Social Science, Other, NA
Gender	The participant's gender. (TeamBuilder and profile)	male, female, unknown
TeamSuccess	The state of the participant's team at the end of the task. • passed: at least two team members have finished the team task including reviewing the work of other teams. • dysfunctional: only one team member has finished the team task including reviewing the work of other teams. • failed: one or more team members have started to work on the task, but none of them finished • not started: none of the team members has started to work on the task.	passed, dysfunctional, failed, not started
GfP	Grade from peers. Percentage. The same for all members of a team, if they have also reviewed other teams. Otherwise: zero.	0-1
BPRev	Bonus points for reviews. Percentage. If three reviews are mandatory, then six reviews are possible. Receiving best ratings on all mandatory reviews: 100% points. Max. possible amount of points: 200%	0-2
BPSA	Bonus points for self-assessment. Percentage.	0-1
BPT	Bonus points for in-team assessment. Percentage.	0-1
SR	Star rating. Rating from expert reviewers.	0, 1, 2, 3
Nmn	Nominations for gallery	0-x

Phase	The phase in which the participant left the peer assessment	Assigment-Submission, PeerGrading, Self-Assessment, Results
IVSec1-x	Items visited in section.	0-x
IVPSec1-x	Items visited in section. Percentage.	0-1
PSec1-x	Points achieved in section.	0-x
PPSec1-x	Points achieved in section. Percentage.	0-1
Q1-x	Points achieved in Quiz.	0-x
Career	The participants career stage.	teacher, student, professional, academic researcher, NA
Degree	The participant's highest achieved degree.	high-school, bachelor, master, professional, academic researcher, NA
BIT	The participant's background in IT	beginner, advanced, expert, NA
Experience	The participants work experience in years.	up_to_five_years, up_to_10_years, more_than_10_years, NA
Position	Position in job.	technician, team_leader, project_manager, department_head, intern, other, NA
TCC	Top country code. Country by IP. The country from which the participant has most often visited the course	DE
TCN	Top country name. Country by IP. The country from which the participant has most often visited the course	Germany
City	Top city name. City by IP. The city from which the participant has most often visited the course	Berlin

DWA	Desktop web activity.	0-x
MWA	Mobile web activity.	0-x
MAA	Mobile application activity.	0-x
LVI	Last visited item. The last item a user has visited on the platform.	UUID
LVIT	The time when the user has visited the last item.	Timestamp
Sessions	The number of sessions, the participants has started in the course. A new session is started whenever there is no activity in the course for more than 30 minutes (and only then).	0-x
ASD	Average session duration in seconds. Rounded.	0-x
TSD	Total session duration in seconds. $Sessions * ASD \approx TSD$	0-x
VPAP	Video play activity. Percentage. Tracks unique video plays, not if the video has been played till the end.	0-1
VDAP	Video download activity. Percentage.	0-1
SDAP	Slide download activity. Percentage.	0-1
FA	Forum activity. Total. Read, write, upvote, downvote, accept answer, etc.	0-x
FAD	Forum activity. Per Day (takes length of course into account). Read, write, upvote, downvote, accept answer, etc.	0-x

Table 28: Team member data

Appendix VI - Non-Team Participant Data

The following list delineates each column of the participants' data that have not opted to work in a team in detail. The complete dataset is available as a CSV file and can be found on the attached CD (data/courses_non_team_data.csv).

Column	Description	Possible values or *Example*
Platform	The platform that offered the course (openHPI, openSAP, mooc.house)	HPI, SAP, MOC
CLNG	Course language	en, de
Course	Course code as on platform	*java1*
UserHash	Anonymized user ID	alphanumeric, lowercase, length: 64
ED	Enrollment date. When did the participant enroll for this course?	Time stamp
FE	First enrollment. Was it the participant's first course on this platform?	TRUE, FALSE
UserCreated	When did the user register on the platform?	Time stamp
LNG	The user's preferred language. Platform settings	*en*
Age	The user's age. Data from user profile.	*75*
AgeMerged	The user's age group.	10-19, 20-29, ..., 70-110
EDD	Enrollment delta in days. Positive values: the number of days that a participant has enrolled after course start. Negative values: the number of days that a participant has enrolled before course start.	*-120, 28*

Posts	The number of the participant's forum posts in the course.	*10*
Threads	The number of forum threads that the participant has started in the course.	*5*
CoP	Confirmation of participation. Did the participant earn one.	TRUE, FALSE
RoA	Record of achievement. Did the participant earn one.	TRUE, FALSE
CC	Course completed. Participant defines if she completed a course.	TRUE, FALSE
Quantile	Quantile of team members. Members with a quantile of 0.95 or more are among the Top 5% of the course participants	0-1
Top	The position (Top5, Top10, Top20, None) that the participant has achieved.	Top5, Top10, Top20, None
IV	Items visited. Amount of visited items.	0-x
IVP	Items visited percentage. Percentage of visited items, compared to available items.	0-1
Points	Amount of achieved points.	0-x
PP	Percentage of achieved points	0-1
IVSec1-x	Items visited in section.	0-x
IVPSec1-x	Items visited in section. Percentage.	0-1
PSec1-x	Points achieved in section.	0-x
PPSec1-x	Points achieved in section. Percentage.	0-1
Q1-x	Points achieved in Quiz.	0-x
Career	The participants career stage.	teacher, student, professional, academic researcher, NA
Degree	The participant's highest achieved degree.	high-school, bachelor, master, professional, academic researcher, NA
BIT	The participant's background in IT	beginner, advanced, expert, NA

Experience	The participants work experience in years.	up_to_five_years, up_to_10_years, more_than_10_years, NA
Position	Position in job.	technician, team_leader, project_manager, department_head, intern, other, NA
TCC	Top country code. Country by IP. The country from which the participant has most often visited the course	DE
TCN	Top country name. Country by IP. The country from which the participant has most often visited the course	Germany
City	Top city name. City by IP. The city from which the participant has most often visited the course	Berlin
DWA	Desktop web activity.	0-x
MWA	Mobile web activity.	0-x
MAA	Mobile application activity.	0-x
LVI	Last visited item. The last item a user has visited on the platform.	UUID
LVIT	The time when the user has visited the last item.	Timestamp
Sessions	The number of sessions, the participants has started in the course. A new session is started whenever there is no activity in the course for more than 30 minutes (and only then).	0-x
ASD	Average session duration in seconds. Rounded.	0-x
TSD	Total session duration in seconds. $Sessions * ASD \approx TSD$	0-x
VPAP	Video play activity. Percentage. Tracks unique video plays, not if the video has been played till the end.	0-1
VDAP	Video download activity. Percentage.	0-1

SDAP	Slide download activity. Percentage.	0-1
FA	Forum activity. Total. Read, write, upvote, downvote, accept answer, etc.	0-x
FAD	Forum activity. Per Day (takes length of course into account). Read, write, upvote, downvote, accept answer, etc.	0-x

Table 29: Non-team course participants data