Supporting Virtual Team Formation through Community-Wide Deliberation

MIAOMIAO WEN, Coursera
KEITH MAKI, Carnegie Mellon University
STEVEN P DOW, UC San Diego
JAMES HERBSLEB, Carnegie Mellon University
CAROLYN ROSE, Carnegie Mellon University

Team-based learning is a structured, small-group learning method that has been associated with many positive outcomes in traditional classroom settings. However, relatively little research has focused on how to form and support teams within online learning platforms, such as Massive Open Online Courses (MOOCs). A number of challenges arise for team formation in voluntary online classes: students may drop out and leave their team, and even if they do persist with the course, the team may not work together effectively. In this paper, we introduce a team-formation strategy that incorporates a deliberation process, where participants hold discussions in preparation for the collaboration task. First, we present a crowdsourced experiment that compares teams that are formed before or after a community deliberation process. Results demonstrate that teams engaging in a larger community deliberative process prior to team formation exhibit better team performance–as measured by team collaboration product quality–than pre-discussion teams. In a second crowdsourced experiment, we further explore the benefits of community-wide processes by automatically assigning teams based on participants' transactive interaction during deliberation. The results demonstrate advantages in terms of team performance for teams formed based on observed interactions during the community-level deliberation, compared to randomly formed teams. Finally, in a case study, we demonstrate how we successfully adapted the team formation strategy for use in a small MOOC.

Additional Key Words and Phrases: Team collaboration; team formation; transactivity; online learning

ACM Reference Format:
Miaomiao Wen, Keith Maki, Steven P Dow, James Herbsleb, and Carolyn Rose. 2017. Supporting Virtual Team Formation through Community-Wide Deliberation. Proc. ACM Hum.-Comput. Interact. 1, CSCW, Article 109 (November 2017), 20 pages. https://doi.org/10.1145/3134744

1 INTRODUCTION

In traditional classrooms, team-based learning is widely adopted to help learners foster the competencies of collaboration and communication in practice [36]. Team-based learning is one type of peer learning, where students form small teams to complete a course project together. There has been interest in incorporating a collaborative team-based learning component into MOOCs ever since the beginning [48]. However, currently most MOOC learners experience MOOCs as solitary learning experiences. Dillahunt et al. showed that learners requested more tangible benefits

2573-0142/2017/11-ART109 $15.00
https://doi.org/10.1145/3134744

from MOOCs such as more project-based experiences and the ability to interact with others [12]. On the other hand, simply placing online learners in small groups has never guaranteed that collaboration will occur, especially in light of well known process losses in groups [25]. One reason is that online learners come to courses with a wide range of purposes and levels of motivation [33]. They are typically distributed and do not have existing relationships [34]. One strategy for team formation occurs through personal messaging early in a course and is typically based on limited learner profiles, e.g. demographics and prior knowledge [48]. After team formation, students mostly interact with their team and lose contact with the larger course community. Many students drop out before engaging in team project collaboration. There has been limited success in self-selected or randomly assigned MOOC teams [47, 48, 50]. The key challenges of forming effective teams in MOOCs include: (1) students depend on their team though it may become an ineffective source of support due to attrition, and (2) team composition is frequently suboptimal due to scant information regarding which learners will likely work well together.

To address these challenges, the key idea of our team formation method is that learners should have the opportunity to interact meaningfully with the community before assignment into teams. We hypothesize that this community discussion may also provide evidence of who might work well with each other in a team collaboration: we identify which students are showing interest in each other's work, and especially, which students are offering each other meaningful feedback during community discussion. In particular, our team formation approach matches students with others with whom they have exchanged what is referred to in the Learning Sciences as transactive discussion during community deliberation. A transactive exchange is one where participants "elaborate, build upon, question or argue against previously presented ideas" [6]. It has long been established in learning science that transactive discussion is an important process that reflects good social dynamics in a group [44] and results in collaborative knowledge integration [9]. Similar concepts have been referred to under different names, such as information elaboration in organizational communication research. In that literature, this form of discussion has been shown to be positively related to team performance [39].

Many earlier research studies have utilized A/B testing in real online courses [23, 26]. However, instructors may hesitate to try out untested designs in live courses. Inspired by Coetzee et al. [8], we first prototyped and validated our proposed approach using a crowdsourcing service, Amazon Mechanical Turk (MTurk). Based on the positive results, we then did a team formation case study in a live team-based MOOC. The team formation process begins with individual work where participants learned about one of four learning materials, then participated in an open discussion, and then worked with three other teammates (using a Jigsaw learning configuration where each team member has finished one of the four learning materials [3]) to solve a challenge. We adopted team performance as the key metric of this study. In the MTurk experiments, we assessed team performance based on the quality of the produced team proposal.

MTurk Experiment 1 tests the extent to which teams that engage in a large community forum deliberation process prior to team formation achieve better team task outcomes than teams that instead perform an analogous deliberation within their team. We hypothesize that the community wide discussion provides students with a wealth of insight into alternative task-relevant perspectives in the collaboration [7]. Simply stated, our first research question is:

RQ1. Will exposure to large community discussions lead to more successful small team collaborations?

To address the disadvantage that teams in online courses have frequently been formed with limited evidence of who will work effectively together, in MTurk Experiment 2, we look for evidence of participants transactively reasoning with each other during community-wide deliberation and use it as input into a team formation algorithm. Simply stated, our second research question is:

RQ2. Can evidence of transactive exchange between participants during deliberation inform the composition of more successful teams compared to randomly formed teams?

Since crowd workers likely have some different motivations from actual online learners, there remains an open question about the extent to which findings from MTurk would make predictions about what we would see in a real online course. In the third experiment, we adapted and successfully deployed our team formation strategy in a team-based MOOC where students form teams of four to design a superhero team story together. We formed 52 teams, all of which submitted their team project at the end of the course. The positive team formation effects that we observed in the MTurk environment were also observed in the noisy, externally valid environment of a real MOOC.

This paper contributes: (1) a community-wide deliberation-based approach for team formation; (2) evidence that this approach leads to better team performance than teams formed before community-wide deliberation; (3) an algorithm for optimizing team formation based on detected transactive exchanges that occurred during deliberation, and evidence that this leads to better team processes and better outcomes; (4) a case study in a team-based MOOC where we deployed these techniques with 52 teams, which suggests the utility of the methods in practice as verified in the MTurk environment.

2 RELATED WORK

2.1 Strategies for Team Formation

Group formation methods are an important component for enhancing team member participation in small groups [20]. Three common approaches to team formation in courses include self-selection, random, and criteria-based [10]. Self selection, the most prevalent form of grouping for team-based MOOCs [48], is considered better for interaction but difficult to implement in a short time since in this case participants typically do not know each other and lack the face-to-face contact to "feel out" potential group members [38]. For this reason, student teams in classrooms and E-learning contexts are usually assigned by instructors, often randomly [11]. Most previous work in team formation has focused on team composition rather than the socialization process through which participants enter into their teams. CATME is a well-known software that assign teams based on survey information, e.g. schedule and gender [28]. Zheng et al. shows that forming MOOC teams based on demographic features, e.g. gender and time zone, does not significantly improve teams' engagement and success in MOOCs [50]. Instead of focusing on group composition, we design a practical team formation procedure through which participants are organized into small teams. Research in distributed team formation shows that prior collaborations and familiarity are good predictors for how to form productive virtual teams [31, 41]. We design a deliberation task that provides participants an opportunity to get to know each other's work and interact with each other prior to team work. The resulting data about how people interact in the forum informs our automated approach to team formation. We design the tasks and the team formation method together to support effective collaboration.

2.2 Community vs. Small Team Deliberation

Deliberation, or rational discourse that marshals evidence and arguments that bear on a decision, can be effective for engaging groups of diverse individuals to challenge one another to think in new ways [7]. Effective deliberation involves participants with diverse intellectual resources sharing their insights with the group so they can be integrated to form complex solutions that would not be possible with more limited access to these resources. These benefits are associated with both small groups discussions and community-wide deliberation.

A larger community has access to a wider variety of resources compared to a small group. Sunstein [42] points out that owing to the increased social and political pressure within a small group, deliberation often fails to elicit all the relevant information that group members are aware of. Interaction with others of differing views is essential in order to comprehend and appreciate other perspectives [5]. Contact with opposing views is more likely to happen during a community-wide deliberation [37]. On the other hand, small groups often benefit from agility in the form of a high degree of goal-setting autonomy and freedom to shape their practices to their own needs. Informed by this research, we study whether exposing members to broader discussion prior to group work will be more valuable as preparation for teamwork than participation in deliberation within the team itself.

2.3 Transactivity as Evidence for Team Formation

Transactivity is akin to discourse strategies identified within the organizational communication literature for solidarity building in work settings [40] as well as rhetorical strategies associated with showing openness [35]. The idea is part of the neo-Piagetian perspective on learning where it is understood that optimal learning between students occurs when students respect both their own ideas and those of the peers that they interact with, which is grounded in a balance of power within a social setting. Transactivity is known to be higher within teams where there is mutual respect [4] and a desire to build common ground [17]. High transactivity teams are associated with higher learning [21, 44], higher knowledge integration [9], and effective collaborative problem solving [4].

The concept of participants building on other people's idea is also related to the idea of information elaboration. Information elaboration is a complex form of communication that involves “the exchange of information and perspectives, the process of feeding back the results of this individual-level processing into the group, and discussion and integration of its implications" [46]. These processes extend beyond information sharing to capture the extent to which team members contribute detailed explanations of their ideas, and spend time constructively discussing each other's perspectives, integrating information, and determining how to apply their knowledge resources to the problem at hand [19]. Transactive discussion also shows that students are paying attention and connecting with each other, which can enhance group cohesion [43]. Information elaboration has been shown to be positively related to team performance [39]. Informed by this research, we study whether teams composed of individuals with a history of engaging in more transactive communication during a pre-collaboration deliberation achieve more effective collaboration in their teams.

3 METHOD

3.1 Experimental Paradigm

3.1.1 Collaboration Task Description. We designed a highly-interdependent collaboration task that requires negotiation in order to create a context where effective team collaboration would be necessary for team success. We used a Jigsaw paradigm, which has been demonstrated as an effective way to achieve a positive team composition and is associated with positive team outcomes [3]. Following the Jigsaw paradigm, each member of the team was given special knowledge of one of the four energy sources (coal, wind, nuclear and hydro power), and was instructed to represent the values associated with their energy source in contrast to the rest, e.g. coal energy was paired with an economical energy perspective. The collaboration task asked teams to consider municipal energy plan alternatives that involved combinations of four energy sources each paired with specific advantages and disadvantages. The team collaborative task was to select a single energy plan and write a proposal arguing in favor of their decision with respect to the trade-offs they have made. The

In this final step, you will work together with other Turkers to recommend a way of distributing resources across energy types for the administration of City B. City B requires 12,000,000 MWh electricity a year from four types of energy sources: coal power, wind power, nuclear power and hydro power. Your team needs to negotiate which plan is the best way of meeting your assigned goals, given the city's requirements and information below.
(City B's 8 requirements were listed here, e.g. "The city is concerned with water pollution.")

	Energy plan				**Cost**	**Waste disposal cost**	**Carbon tax credit**	**Total**
	Coal	Wind	Nuclear	Hydro				
Plan 1	40%	20%	20%	20%	$840,000K	$14,400K	$48,000K	$892,800K
Plan 2	20%	40%	20%	20%	$912,000K	$0	$0	$901,000K
Plan 3	20%	20%	40%	20%	$984,000K	$14,400K	$0	$988,800K
Plan 4	20%	20%	20%	40%	$984,000K	$0	$0	$973,600K

Fig. 1. Collaboration task description.

set of potential energy plans was constructed to reflect different trade-offs among the requirements, with no plan satisfying all of them perfectly. This ambiguity created an opportunity for intensive exchange of perspectives. The collaboration task is shown in Figure 1.

3.1.2 Experimental Procedure. We designed a four-step MTurk experiment:

Step 1. Preparation (~5 minutes)

In this step, each participant was asked to provide a nickname, to be used in later steps. To prepare for the Jigsaw task, each worker was randomly assigned to read an instructional article (~500 words) about the pros and cons of one of four energy sources (coal, wind, nuclear and hydro power). We asked them to complete a quiz (8 single-choice questions) reinforcing the content of their assigned article. The quiz was not used as a filter for later steps, it was designed to strengthen learning. Feedback including correct answers and explanations was provided along with the quiz.

Step 2. Individual Task (~5 minutes)

In this step, we asked each worker to write a proposal to recommend one of the four energy sources (coal, wind, nuclear and hydro power) for a city given its five requirements, e.g. "The city prefers a stable energy". After each worker finished this step, their proposal was automatically posted in a forum as the start of a thread with the title "[Nickname]'s Proposal".

Step 3. Deliberation (10-15 minutes)

In this step, workers joined a threaded forum discussion akin to those available in many MOOCs. Each proposal written by the workers in the Individual Task (Step 2) was displayed for workers to read and comment on. Each worker was required to write at least five replies to the proposals posted by the other workers. To encourage the workers to discuss transactively, the task instruction included "when replying to a post, please elaborate, build upon, question or argue against the ideas presented in that post, drawing from the argumentation in your own proposal where appropriate." The workers were not aware that they will be grouped based on their discussion.

Step 4. Collaboration (~15 minutes)

In the collaboration step, team members were first synchronized and then directed to a shared Etherpad[1] to write a proposal together to recommend one of four suggested energy plans based on a city's eight requirements (Figure 1). Etherpad-lite is an open-source collaborative editor [51], meaning workers in the same team were able to see each other's edits in real-time. They were able to communicate with each other using a chat utility on the sidebar. The collaborative task was

[1]http://etherpad.org/

designed to contain richer information than the individual proposal writing task in the Individual Task (Step 2). Workers were also required to fill out a survey measuring their perceived group outcomes after collaboration.

Outcome Measures

Both of our research questions made claims about team success. We evaluated this success using two types of outcomes, namely objective success through quantitative task performance and process measures as well as subjective success through a group satisfaction survey.

The quantitative task performance measure was an evaluation of the quality of the proposal produced by the team. In particular, a scoring rubric defined how to identify the following elements from a proposal:

1. Which requirements were considered (e.g., *"Windfarms may be negative for the bird population"*)
2. Which comparisons or trade-offs were made (e.g., *"It is much more expensive to build a hydro plant than it is to run a windfarm"*)
3. Which additional valid desiderata were considered beyond stated requirements (e.g., *"Hydro plants also require a large body of water with a strong current"*)
4. Which incorrect statements were made about requirements (e.g., *"Hydro does not affect animal life around it"*)

Positive points were awarded to each proposal for correct requirements considered, comparisons made, and additional valid desiderata. Negative points were awarded for incorrect statements. We measured **Team Performance** by the total points assigned to the team proposal. Two PhD students who were blind to the conditions applied the rubric to five proposals (a total of 78 sentences) and the inter-rater reliability was good (Kappa = 0.74). The two raters then coded all the proposals. We used the number of transactive contributions during team collaboration discussion in the Collaboration step as a measure of **Team Process**.

Group Experience Satisfaction was measured using a three item group experience survey administered to each participant after the Collaboration step. The survey was based on items used in prior work [8, 16, 30]. The survey instrument included item (measured on a 1-7 Likert scale) related to: (1) Satisfaction with team experience; (2) Satisfaction with proposal quality; (3) Satisfaction with the communication within the group.

Control Variables

Intuitively, workers who display more effort in the Individual Task might perform better in the collaboration task. We used the average group member's Individual Task proposal length as a control variable for group performance.

Transactivity Annotation, Prediction and Measurement

To enable us to use counts of transactive exchanges as evidence to inform an automated team assignment procedure, we needed to automatically judge whether a reply post in the Deliberation step was transactive or not using machine learning. Using a validated and reliable coding manual for transactivity from prior work [17, 21], an annotator previously trained to apply that coding manual annotated 426 reply posts collected in pilot studies we conducted in preparation for the studies reported in this paper. Each of those posts was annotated as either "transactive" or "non-transactive". 70% of them were transactive.

A transactive contribution displays the author's reasoning and connects that reasoning to material communicated earlier. Two example posts illustrating the contrast are shown below:

Transactive: *"Nuclear energy, as it is efficient, it is not sustainable. Also, think of the disaster probabilities"*.

Non-transactive: *"I agree that nuclear power would be the best solution"*.

Automatic annotation of transactivity has been reported in the Computer Supported Collaborative Learning literature. For example, researchers have applied machine learning using text, such as

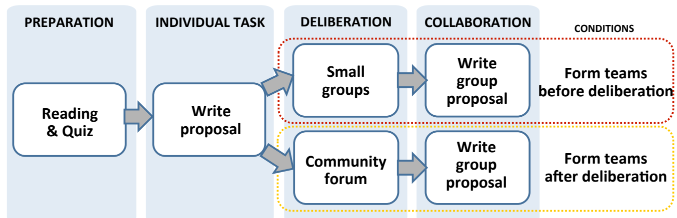

Fig. 2. Workflow diagrams for Experiment 1.

chat data [21] and transcripts of whole group discussions [2]. We trained a Logistic Regression classifier with L2 regularization using a set of features, which included unigrams (i.e., single word features) as well as a feature indicating the post length [13]. We evaluated our classifier with a 10-fold cross validation and achieved an accuracy of 0.843 and a 0.615 Kappa. Given the adequate performance of the classifier, we used it to predict whether each reply post in the Deliberation step was transactive or not.

To measure **the amount of transactive communication** between two participants in the Deliberation step, we counted the number of times both their posts in a same discussion thread were transactive; or one of them was a thread starter and the other one's reply was transactive.

4 MTURK EXPERIMENT 1. GROUP TRANSITION TIMING: BEFORE DELIBERATION VS. AFTER DELIBERATION

This experiment (Figure 2) assessed the extent to which a measurable improvement occurs in team performance when team members transition into teams after they experience community-level deliberation. Thus, what we manipulated was the step when workers began to work within their small team. To control for timing of synchronization and grouping, in both conditions, workers were synchronized and assigned into small teams based on a Jigsaw paradigm after the Individual Task. The only difference was at which step the workers move into their small teams. For the **After Deliberation** condition, in the Deliberation step workers could potentially interact with workers both inside and outside their group since they had the discussion in a large group (40-50 workers). Workers were not told that they had been assigned into teams until the Collaboration step (Step 4). In the **Before Deliberation** condition, each team was given a separate forum in which to interact with their teammates. The Before Deliberation condition is similar to the current team-based MOOCs where teams are formed early in the course and only interact with their teammates. By comparing these two conditions, we test our hypothesis that exposure to deliberation within a larger community will improve team performance.

4.1 Synchronizing Workers

MTurk does not provide a mechanism to bring several workers to a collaboration task at the same time. We built on earlier investigations that described procedures for assembling multiple crowd workers on online platforms to form synchronous on-demand teams [8, 27]. Our approach was to start the synchronous step at fixed times, announcing them ahead of time in the task description and allowing workers to wait before the synchronous step. A countdown timer in the task window displayed the remaining time until the synchronous step began, and a pop-up window notification was used to alert all participants when the waiting period had elapsed.

4.2 Participants

Participants were recruited on MTurk with the qualifications of having a 95% acceptance rate on 1000 tasks or more. Each worker was only allowed to participate once. A total of 252 workers participated in Experiment 1, the workers who were not assigned into teams or did not successfully complete the group satisfaction survey were excluded from our analysis. Worker sessions lasted 34.8 minutes on average. Each worker was paid $4. To motivate participation during the Collaboration step [8], workers were awarded a bonus based on their level of interaction with their team ($0.1 - $0.5), while an extra bonus was given to workers whose team submitted a high quality proposal ($0.5). We included only teams of 4 workers in our analysis, there were in total 22 Before Deliberation teams and 20 After Deliberation teams.

A chi-squared test revealed no significant difference in worker attrition between the two conditions. We considered a worker as having "dropped out" from their team if they were assigned into a team but did not edit the proposal in the Collaboration step. There was no significant difference in the dropout rate of workers between the two conditions ($\chi^2(1) = 0.08$, $p = 0.78$). The dropout rate for workers in Before Deliberation teams was 30%. The dropout rate for workers in After Deliberation teams was 28%.

4.3 Results

4.3.1 Teams exposed to community deliberation prior to group work demonstrate better team performance. We built an ANOVA model with Group Transition Timing (Before Deliberation, After Deliberation) as the independent variable and Team Performance as the dependent variable. In order to control for differences in average verbosity across teams, we included the Individual Task proposal length averaged across team members as a covariate for each group. There was a significant main effect of Group Transition Timing on Team Performance ($F(1,40) = 5.16$, $p < 0.05$) such that After Deliberation groups had a significantly better performance ($M = 9.25$, $SD = 0.87$) than the Before Deliberation groups ($M = 6.68$, $SD = 0.83$), with an effect size of 2.95 standard deviations. We think the effect on team performance comes from workers being exposed to more posts and comments during community deliberation compared to the deliberation in small team.

We also tested whether the differences in teamwork process between conditions was visible in the extent of the number of transactive turns during collaboration discussion. We built an ANOVA model with Group Transition Timing (Before Deliberation, After Deliberation) as the independent variable. In this case, there was no significant effect ($p = 0.28$). Thus, teams in the After Deliberation condition were able to achieve better performance in their team product without requiring more transactive discussion.

4.3.2 Survey results. In addition to assessing group using our scoring rubric, we assessed the subjective experience of workers using the group experience survey discussed earlier. For each of the four aspects of the survey, we built an ANOVA model with Group Transition Timing (Before Deliberation, After Deliberation) as the independent variable and the survey outcome as the dependent variable. TeamID and assigned energy condition (Coal, Wind, Hydro, Nuclear) were included as control variables nested within condition. There were no significant effects on any of the subjective measures in this experiment.

5 MTURK EXPERIMENT 2. GROUPING CRITERIA: RANDOM VS. TRANSACTIVITY MAXIMIZATION

While in Experiment 1 we investigated the impact of exposure to community resources prior to teamwork on team performance, in Experiment 2 we investigated how the discussion trace during community deliberation may inform effective team formation. This time teams in both conditions

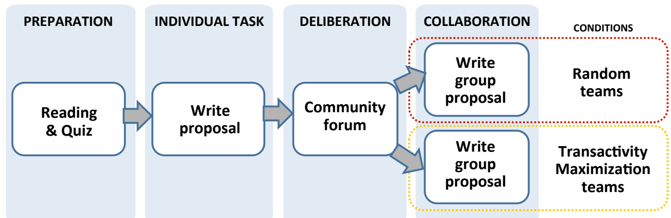

Fig. 3. Workflow diagrams for Experiment 2.

were grouped after experiencing the Deliberation step in the community context. In Experiment 2 (Figure 3), we again made use of the Jigsaw paradigm, but in the experimental condition, which we termed the **Transactivity Maximization** condition, we additionally applied a constraint that preferred to maximize the extent to which workers assigned to the same team had participated in transactive exchanges in the deliberation. We are not manipulating the total number of transactive exchanges during the community discussion. The maximization assigns participants into teams with the people they had transactive exchanges with rather than others, so that the total number of transactive exchanges between team members is maximized. In the control condition, which we termed the **Random** condition, teams were formed randomly apart from enforcing the Jigsaw constraint. It is a strong baseline since Jigsaw is a best practice in team formation. As previous work has shown that self-selection does not work well in team-based MOOCs [47], we did not compare with self-selected team formation. In this way we tested the hypothesis that observed transactivity is an indicator of potential effective team collaboration.

From a technical perspective in Experiment 2 we used a constraint satisfaction algorithm to manipulate how the teams were assigned. For the Transactivity Maximization condition, teams were formed so that the amount of transactive discussion among the team members was maximized. We first compute the number of transactive communications between each pair of participants. A Minimal Cost Max Network Flow algorithm greedily assigns participants into teams to increase the total amount of measured transactive communication that had occurred among the team members during deliberation [1][2]. This standard network flow algorithm tackles resource allocation problems with constraints. In our case, the constraint was that each group should contain four people who have read about different energies (i.e. a Jigsaw group). At the same time, the minimal cost part of the algorithm maximized the transactive communication that was observed among the group members during Deliberation step. The algorithm finds an approximately optimal grouping within $O(N^3)$ (N = number of workers) time complexity. A brute force search algorithm, which has an $O(N!)$ time complexity, would take too long (over a week) to finish since the algorithm needs to operate in real time. Except for the grouping algorithm, all the steps and instructions were identical for the two conditions.

5.1 Participants

A total of 246 workers participated in Experiment 2, the workers who were not assigned into teams or did not complete the group satisfaction survey were excluded from our analysis. The compensation was the same as in MTurk experiment 1. Worker sessions lasted on average 35.9 minutes. We included only teams of 4 workers in our analysis. There were in total 27 Transactivity Maximization teams and 27 Random teams, with no significant difference in attrition between

[2]Source code of the Minimal Cost Max Network Flow algorithm can be found here: https://tinyurl.com/mincostmaxflow

conditions ($\chi^2(1) = 1.46$, $p = 0.23$). The dropout rate of workers in Random teams was 27%. The dropout rate of workers in Transactivity Maximization teams was 19%.

5.2 Results

As a manipulation check, we compared the average number of transactive exchanges observed among team members during the deliberation between the two conditions using a t-test. The teams in the Transactivity Maximization condition (M = 12.85, SD = 1.34) were observed to have had significantly more transactive communication during the deliberation than those in the Random condition (M = 7.00, SD = 1.52) ($p < 0.01$), with an effect size of 3.85 standard deviations, demonstrating that the maximization was successful in manipulating the average experienced transactive exchange within teams between conditions.

5.2.1 Teams that experienced more transactive communication during deliberation demonstrate better team performance. To assess whether the Transactivity Maximization condition resulted in more effective teams, we tested for a difference between team formation conditions on Team Performance. We built an ANOVA model with Grouping Criteria (Random, Transactivity Maximization) as the independent variable and Team Performance as the dependent variable. Average team member Individual Task proposal length was again the covariate. There was a significant main effect of Grouping Criteria (F(1,52) = 6.13, $p < 0.05$) on Team Performance such that Transactivity Maximization teams (M = 11.74, SD = 0.67) demonstrated significantly better performance than the Random groups (M = 9.37, SD = 0.67) ($p < 0.05$), with an effect size of 3.54 standard deviations, which is a large effect.

Across the two conditions, observed transactive communication during deliberation was significantly correlated with Team Performance (r = 0.26, $p < 0.05$). This also indicated teams that experienced more transactive communication during deliberation demonstrated better team performance.

5.2.2 Teams that experienced more transactive communication during deliberation demonstrate more transactive interaction within their teams. In Experiment 2, workers were assigned to teams based on observed transactive communication during the deliberation step. Assuming that individuals that were able to engage in positive collaborative behaviors together during the deliberation would continue to do so once in their teams, we would expect to see evidence of this reflected in their observed team process, whereas we did not see such an effect in Experiment 1 where teams were assigned randomly in all conditions. Group processes have been demonstrated to be strongly related to group outcomes in face-to-face problem solving settings [49]. Thus, we should consider evidence of a positive effect on group processes as an additional positive outcome of the experimental manipulation.

In order to test whether such an effect occurred in Experiment 2, we built an ANOVA model with Grouping Criteria (Random, Transactivity Maximization) as the independent variable and number of transactive contributions during teamwork as the dependent variable. There was a significant effect of Grouping Criteria on the number of transactive discussion (F(1,45) = 5.02, $p < 0.05$). Random teams (M = 14.51, SD = 3.58) demonstrated significantly fewer transactive contributions than the Transactivity Maximization teams (M = 18.30, SD = 3.16), with an effect size of 1.10 standard deviations.

Table 1 shows one transactive and one non-transactive collaboration discussion. The transactive discussion contained reasoning about the pros and cons of the energy plans, which can easily translate into the team proposal. The non-transactive collaborative discussion came to a quick consensus without discussing each participant's rationale behind choosing an energy plan. Then team members need to generate and organize their reasons for choosing the plan. For participants

Transactive	Non-transactive
A: based on plan 1 and 2 I am thinking 2 only because it reduces greenhouse gases B: Yeah so if we go with 2, we will need to trade off the water pollution and greenhouse gas C: BUT we run into the issue of budget... so where do we say the extra almost $100k comes from?	A: My two picks are Plan 1 and Plan 2 B: Alright, lets take a vote. Type either Plan 1 or Plan 2 in chat. B: Plan 2 C: Plan 1 D: plan 2. B: That settles it, itâĂŹs plan 2.

Table 1. Example of Transactive vs. Non-transactive Discussions during Team Collaboration.

who initially did not pick the chosen energy plan, without the transactive discussion process, it might be difficult for them to integrate their knowledge and perspective into the team proposal.

5.2.3 Transactivity maximization teams reported higher communication satisfaction. For each of the four aspects of the group experience survey, we built an ANOVA model with Grouping Criteria (Random, Transactivity Maximization) as the independent variable and the survey outcome as the dependent variable. TeamID and assigned energy condition (Coal, Wind, Hydro, Nuclear) were included as control variables nested within condition. There were no significant effects on Satisfaction with team experience or with proposal quality. However, there was a significant effect of condition on Satisfaction with communication within the team ($F(1,112) = 4.83$, $p < 0.05$), such that workers in the Random teams ($M = 5.12$, $SD = 1.7$) rated the communication significantly lower than those in the Transactivity Maximization teams ($M = 5.69$, $SD = 1.51$), with effect size .38 standard deviations. Thus, with respect to subjective experience, we see advantages for the Transactivity Maximization condition, but the results are weaker than those observed for the objective measures. Nevertheless, these results are consistent with prior work where objectively measured learning benefits are observed in high transactivity teams [9].

5.3 Discussion

The preliminary results from the crowdsourced experiments were in accordance with our hypotheses when tested in a high internal validity context. In order to study whether our team formation paradigm will play out in an externally valid context, we deployed our team formation paradigm in a MOOC case study. MOOC students likely have different motivations from crowd workers. Our MTurk task typically lasts less than an hour; while a virtual team collaboration in a team-based MOOC can be several weeks long. Only a deployment study can answer questions like: how many students in a MOOC will actually participate and enjoy the team collaboration?

6 CASE STUDY: TEAM COLLABORATION AS AN EXTENSION OF A MOOC

We collaborated with the Smithsonian Institution on their edX MOOC, "The Rise of Superheroes and Their Impact On Pop Culture". In the previous three offerings of this MOOC, students learn about how comic book creators build a superhero character. To finish a superhero MOOC, students either design a superhero of their own or write a biography of an existing superhero as their final project submission. We designed a team-based MOOC, "Rise of the Superheroes and the Heroes of the Future"[3], as an extension of the superhero MOOCs. We advertised the team-based MOOC to all the superhero MOOC alumni who had previously created or researched a superhero.

Based on the collaboration task in the MTurk experiments, we designed an interdependent team project that requires knowledge integration and intensive discussion. In the collaboration task,

[3]https://www.edx.org/course/rise-superheroes-heroes-future-smithsonianx-popx2-1

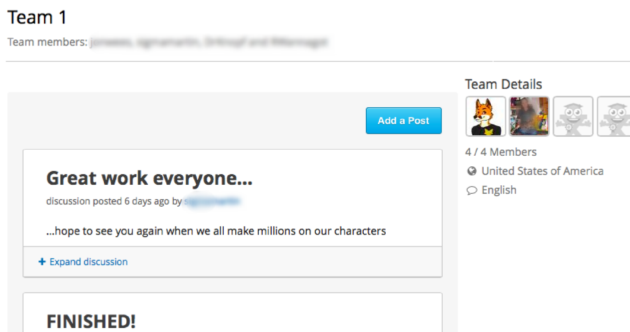

Fig. 4. An edX's Team space example. Students can view the team members and posts.

team members design a superhero team story together using the superheroes they have designed or researched in previous offerings of this MOOC. They will discuss what social issue the team is fighting for and whether the team supports individual freedom or government control. Each team was required to submit one superhero team story together as their final team project. In the team project instructions, we informed learners that their team score will be based on how complete their project is. The required parts of the project included "Choose a social issue for your original superhero team to take on"; "Create a supervillain who is so devious that your entire team needs to battle him or her"; and "An epic battle between your original superhero team and the supervillain you've created", etc. An optional survey was administered after the course finished.

We adapted the team formation process in our MTurk experiments to this team-based MOOC. Since all the students had previously finished a superhero design or analysis that was part of the original course, there was no need to designate course weeks for individual work. The MOOC was three weeks long. Students went through two steps to finish the course: Course community deliberation (Week 1) and Team collaboration (Week 2 and 3). In the deliberation step, students need to first post their hero design or analysis from the original course as a discussion thread starter, and then comment on at least three other people's heroes. To encourage students to provide feedback transactively, we suggested them to comment on one element of the hero that was successful and on one element where an improvement could be made in the instructions. In the Team collaboration step, the team members collaborated on designing a superhero team. Each team was assigned an edX's Team space (beta version, as shown in Figure 4) where only team members can post and comment.

Although we did not run an A/B comparisons where some teams were formed randomly while other teams were formed based on the algorithm, there was nevertheless natural variation in the level of transactive discussion among team members during deliberation. Therefore, the correlation between the level of transactive discussion during deliberation and the team performance can be considered an indication that the team formation method was successful. Based on the results of our crowdsourced team formation studies, we hypothesized that teams that had more transactive discussions during the community deliberation would have better team process and performance. To measure team performance, we use the team score which was based on how complete the team project was. To evaluate team participation and process, we measure (1) how many students participated in the team project? (2) the extent to which the superhero team stories are integrated. In particular, we check if all the heroes interact with each other in the story.

6.1 Results and Discussion

6.1.1 Transactivity maximization team formation in the MOOC. In total, 770 students enrolled in our team-based MOOC. 106 of them paid for the verified certificate ("verified students"). By the end of Week 1, there were in total 208 students who had posted their previous superhero story or analysis and commented on other students' stories or analysis. We manually annotated 300 comment posts that were randomly sampled from all the discussion posts. 61% of them were transactive, which was comparable to the percentage in our MTurk experiments. Then we trained a similar logistic regression classifier to predict whether a comment post is transactive or not. The model was applied to the rest of the reply posts. We planned to use whether a student did a superhero design or analysis in previous superhero MOOC as the Jigsaw grouping conditions. However, there were only 10 students who did a superhero analysis. We decided not to do the Jigsaw grouping. Paying for the verified certificate may indicate that the student is more motivated to finish the MOOC. To balance for motivation level, we did not group verified and unverified students into the same teams. In total, we formed 38 unverified teams and 14 verified teams. In a manipulation check, we verified that the maximization successfully increased the average within team transactive exchange over what would have been present from random selection. Table 2 shows the average number of transactive exchanges in the formed teams and what we would observe in randomly formed teams. The verified teams had more transactive exchanges during course community deliberation compared to unverified teams.

	Unverified Teams	**Verified Teams**
Transactivity Maximization	7.81	9.92
Random	0.35	2.08

Table 2. Average number of transactive exchanges within Transactivity maximized teams and what we would observe in randomly formed teams in the course community deliberation.

6.1.2 High completion rate in the team-based MOOC. All 52 teams submitted their team project. Out of all the 208 students who were assigned to teams, 182 students (87.5%) collaborated in their teams and finished the course.

The completion rate in a typical MOOC is around 5% [24]. The completion rate of the previous three superhero MOOCs is 6%. We think several factors contributed to the high completion rate that was observed in our team-based MOOC. In the case study, all the enrolled students were alumni from previous offerings of this MOOC, who had already demonstrated effort to finish a MOOC on a similar topic. Our team formation process only groups students who have posted their individual work in the course discussion forum. This screening process ensured that students who were assigned to teams had already demonstrated serious intention of working in a virtual team. Therefore, it was not surprising that the retention rate in the formed teams was much higher than for typical MOOC students to finish their MOOC. Compared to a typical five or six-week MOOC, our MOOC was short, only three weeks long. The carefully designed team-based learning experience, may have also contributed to students' commitment to the MOOC. It will require further experimentation to fully understand the effect of team-based learning on student commitment in MOOCs.

6.1.3 Evidence of Teams that experienced more transactive communication during deliberation had more complete team projects. A SmithsonianX staff evaluated all the team projects with a 4 point scale where 4 = Finished all the components, 3 = Only missing panels, 2 = Only missing story and panels and 1 = Missing more than story or panels. 40 teams (77%) finished all the required components. The average team project score is 3.60 (SD = 0.80).

The number of transactive contributions among team members during the community deliberation had a close to significant association with the team project score (F(1, 50) = 2.97, p = 0.08). The verified status of the team and whether there was one student in the team who did a superhero analysis had no significant association with the team project score. Since our team formation maximized teams' average amount of transactive communication, this suggestive evidence indicates that our team formation method may improve teams' overall performance, even in the noisy real MOOC context.

6.1.4 Teams that experienced more transactive communication during deliberation demonstrated better collaboration participation. To examine the effect of grouping on team collaboration participation, we counted how many students actually participated in the team project. For a student to count as having participated, their hero needs to have shown up in the story. The number of transactive exchanges among team members during the community deliberation had a significant association with how many students participated in the collaboration (F(1,50) = 5.85, $p < 0.05$). Whether team members were verified or not and whether there was one student in the team who did a superhero analysis had no significant association with how many students participated in the collaboration.

As a reflection of whether students in the team interacted with each other, we checked whether all of the superheroes have interacted with each other in the team authored stories. Overall, in 44 out of 52 superhero team stories, there was at least one scene where all four superheroes interacted with each other. Controlling for whether the team members were verified students, the number of transactive exchanges during the community deliberation had a close to significant effect on whether all the superheroes ended up interacting or not (F(2,49) = 4.38, p = 0.08). Superheroes in a verified team were significantly more likely to have interacted with each other in the team story compared to unverified team ($p < 0.05$). To sum up, there is an indication that teams that experienced more transactive discussion has better team collaboration participation.

6.1.5 Observations on how transactivity maximization team formation affected the teams. We read all the posts in the team spaces. Since the transactivity maximization team formation tends to assign students with a history of transactive discussion into teams, and we saw that many students recognized their team members with whom they have interacted with during community deliberation: *"I've already read your story in week 1"*, *"I am happy to see that my team members had characters that I was familiar with." "Sup Ron, first of all; thanks for your comments on Soldier Zeta"*. This created a friendly start for the team. Some students indicated that they already had idea about how to collaborate: *"I can already see Osseus and the Soul Rider bonding over the fact your character had a serious illness and The Soul Rider brother was mentally handicapped"*, *"We've already exchanged some ideas last week, I think we can have some really fun dynamics with our crew of heroes!"*. Since transactive discussion is a collaborative knowledge building activity that naturally increases familiarity among team members, these observations are in accordance with prior research that shows that prior collaborations and familiarity may improve the performance of distributed teams [41].

6.1.6 Satisfaction with the team experience. There were in total 138 students who responded to our optional, anonymous post-course survey. Satisfaction with the team experience was rated on a scale from 1-5 with 5 being very satisfied. On average, the satisfaction with the team experience is 4.20 out of 5 (SD = 1.22). Satisfaction with the project is 3.96 out of 5 (SD = 1.06). Overall, students reported being satisfied with their team experience and project submission. On the survey question

“What was the biggest benefit of participation in the team?", the top three most frequently chosen benefits were “The MOOC is more social"(41%), “Get feedback or help" (25%) and “Take on a more challenging project"(24%).

7 GENERAL DISCUSSION

In this paper we address two team formation related research questions. The first is whether participation in deliberation within a course community is more valuable as preparation for teamwork than participation in deliberation within the team only. Here we found that moving into small teams late after course community deliberation has advantages in terms of the quality of the product produced by the teams. We see no effect on team collaboration process measures.

The second related question is the extent to which additional benefit from participation in the deliberation in a community context to teams could be achieved by using evidence of potential successful team interactions from observed transactive exchanges during the deliberation. Here we found that teams formed such that observed transactive interactions between team members in the deliberation was maximized produced objectively higher quality team products than teams assigned randomly. On subjective measures we see a significant positive impact of transactivity maximization on perceived communication quality, which is consistent with what we would expect from the literature on transactivity where high transactivity teams have been demonstrated to produce higher quality outcomes and greater learning [21, 44].

These results provide positive evidence in favor of a team formation strategy in two stages: Individuals first participate in a pre-teamwork deliberation activity where they explore the space of issues in a context that provides beneficial exposure to a wide range of perspectives. Individuals are then grouped automatically through a transactivity detection and maximization procedure that uses communication patterns arising naturally from community processes to inform group formation with an aim for successful collaboration.

7.1 Benefits of Community Deliberation

One benefit of course community deliberation is that students can potentially get feedback and support from all the students in the course. During week 1 of the superhero MOOC, 208 students posted 1302 posts and comments about their superhero designs. Although the discussion instructions did not ask students to discuss transactively, 60% of the contributions to the discussion were transactive. The discussion itself is a valuable learning experience for MOOC students, which prepared students for their later collaboration in small teams. At the end of the course, 44 of the 52 teams voluntarily posted their superhero team projects to the forum for peer feedback. The discussion and feedback in the course forum continued even after the course has ended. This demonstrates the value of community discussion.

7.2 Challenges for Supporting Online Team Collaboration Communication

Since there was no synchronous communication or personal messaging functions in edX, most of the teams communicated asynchronously with posts in their team space. In the post-course survey, many students said that the team discussion space was difficult to use in the sense that messages got buried and students were not notified when there was a new message in their team space. This made keeping on top of the discussion challenging unless students remembered to check each day and put in the effort to sort through the messages. 10 (19%) teams scheduled sychronous meeting over Skype/Google Hangout in their team space. 6 (12%) teams switched to communicate using Facebook groups. In the post course survey, 27 (20%) students indicated that they mainly used email to communicate with their team members. We think a well-integrated team synchronous communication and a messaging tool will be helpful to include in future team space designs [14, 26].

One of the survey question asked "What was most difficult about working in your team?" 30 (21.9%) students mentioned that it was difficult to communicate with their teammates because of times zones [29]. They had trouble finding time to chat live and also found it difficult to agree on changes or make progress since either changes were made while some team members were offline, or it took so long to make a decision the project felt rushed. Further research is needed to study how to support team collaboration and communication.

7.3 Effective groups and effective learning

Much prior work associates Transactivity with learning [21, 44]. This raises the question of whether we can also increase learning with our team formation method. However, in the learning sciences, there has often been a tension observed between emphasizing performance and emphasizing learning [15]. A group that effectively produces a successful outcome, might not optimize for learning for the team members at the same time. Too much group cohesion might lead to groupthink. In project courses, students sometimes take up roles where they can use the knowledge they already have in order to achieve a high quality product, which undercuts the learning that could take place. Often, learning requires focusing on skills that are just beyond a person's ability level. Thus, engagement that leads to learning may frequently appear less successful in terms of performance. We cannot assume that a manipulation that supports a high quality product will necessarily support higher learning. In future work, we will study the effects of our team formation method on students' learning gains.

7.4 Implications for Crowd Work

Most commercial crowdsourcing applications nowadays are based on micro-tasks, which are given to independent workers and do not require cooperation. Recent research explores using crowdsourcing for more complex tasks, which are often interdependent of subjective nature and based on worker cooperation (e.g. [22, 30, 32, 51]). This paper explores an interdependent task that involves workers interacting with each other, each representing a different piece of knowledge and perspective. Our research confirms prior findings that prior collaborations and familiarity are predicative of team performance [31, 41], and most importantly, we propose a practical way of forming efficient crowd worker groups to perform interdependent tasks.

7.5 Reflections on the use of MTurk as a proxy for online courses

Under the course or the instructor's constraints, it can be difficult to do A/B testing in real online courses. Crowdsourced environments are appropriate for controlled experiments that test the internal validity of the hypothesis. In our work, we offer a successful proof of concept where a manipulation from the MTurk environment was adapted for use in an online learning environment.

Online learners and crowd workers are similar in their working environment, education and demographics, while they are different in their context and participation motivation [8, 18]. Crowdsourced experiments may not represent how MOOC students will adopt or enjoy designs. Moreover, the task dropout rate observed in MTurk experiments might not be indicative of the dropout rate in MOOCs. In our work, since the crowd workers did the task mostly for compensation, the main factor that affected crowd workers' intention to participate in or finish the task was the amount of compensation. This is obviously not the case for MOOC students. It is crucial to understand how many students will actually adopt or enjoy the designs by doing deployment studies. This paper examined the effects of the grouping process within these diverse populations in order to test whether the findings from the MTurk environment have some generality to online collaborative learning teams.

7.6 Limitations and future work

One limitation of this work stems from the fact that we carefully designed the sequence of activities –including the individual task, the deliberation, and the team task– to be closely related. Each task provided directly relevant input to the next task. In this carefully designed series of tasks, we see advantages in terms of team performance from the experience of deliberation in the community context. The results of the study leave open the question of whether forming teams after a community-wide deliberation process would be advantageous even in the case where the tasks were not as closely related. If not, the advantages of this post-deliberation team assignment might prove to be short lived. In that case, community level deliberation experiences may need to be interleaved with small group teamwork in order to maintain the advantage over time. Our work indicates that course instructors could design both team tasks and team formation methods that together support effective online team collaboration.

Secondly, we compared transactivity maximization with a random team assignment baseline in MTurk Experiment 2. This leaves open questions for how self-selection would differ from our approach (e.g., would individuals self-select into teams where they had experienced transactive exchanges), and how the associated benefits would compare. MTurk Experiment 2 also leaves open questions for how transactivity maximization teams would compare with a simpler "interaction maximization" team assignment, where total communication among team members during community deliberation are maximized. We have observed that the number of times team members were in the same discussion threads were not significantly correlated with our team success measures ($p > 0.1$) in the MOOC case study. Further experiment is necessary to decide whether "transactive communication" is a stronger indicator of team success than exchanges in general.

Another limitation of the team formation process is the required number of participants. We noticed in our experimentation that if there were too few students (i.e., less than 16), the maximization process did not successfully produce teams that met the criteria better than random teams. While more is better in this sense, there may be limitations on how much bigger of a student population can be supported. For example, when there are large number of participants in the community deliberation (e.g., more than 500), navigability may become another challenge for participants' online community deliberation [45].

8 CONCLUSIONS

In this paper we conducted three experiments in order to obtain an empirical basis for the design of a team formation procedure that can be applied in project-based online courses. The studies were motivated by two specific hypotheses: (1) teams with the benefit of exposure to the community during a pre-teamwork deliberation process will demonstrate advantages in terms of team performance; (2) teams that experienced more transactive communication during a pre-teamwork deliberation process will demonstrate advantages in terms of team performance. The results of the experiments provide evidence in support of both hypotheses.

The first two studies reported in this paper provide evidence of a result when the variables of interest are isolated in a setting in which it is possible to achieve high internal validity. The final case study in a MOOC demonstrates the external validity of these findings as we see consistent findings in a different, more externally valid setting. The correlational results in the deployment study combined with the causal results from the controlled MTurk studies together offer firm empirical support for the effectiveness of our team formation process. This research could provide guidance for team formation in MOOCs, as well as online labor platforms.

9 ACKNOWLEDGEMENTS

This research was funded in part by NSF grants IIS-1302522, ACI-1443068, IIS-1546393, IIS-1122206, Cyberlearning-1122320 and funding from Google.

REFERENCES

[1] Ravindra K Ahuja and James B Orlin. 1989. A fast and simple algorithm for the maximum flow problem. *Operations Research* 37, 5 (1989), 748–759.

[2] Hua Ai, Rohit Kumar, Dong Nguyen, Amrut Nagasunder, and Carolyn P Rosé. 2010. Exploring the effectiveness of social capabilities and goal alignment in computer supported collaborative learning. In *Intelligent Tutoring Systems*. Springer, 134–143.

[3] Elliot Aronson. 1978. *The jigsaw classroom.* Sage.

[4] Margarita Azmitia and Ryan Montgomery. 1993. Friendship, transactive dialogues, and the development of scientific reasoning. *Social development* 2, 3 (1993), 202–221.

[5] Seyla Benhabib. 1992. *Situating the self: Gender, community, and postmodernism in contemporary ethics.* Psychology Press.

[6] Marvin W Berkowitz and John C Gibbs. 1983. Measuring the developmental features of moral discussion. *Merrill-Palmer Quarterly (1982-)* (1983), 399–410.

[7] Robert J Cavalier. 2011. *Approaching deliberative democracy: Theory and practice.* Carnegie Mellon University Press.

[8] Derrick Coetzee, Seongtaek Lim, Armando Fox, Bjorn Hartmann, and Marti A Hearst. 2015. Structuring interactions for large-scale synchronous peer learning. In *Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work & Social Computing*. ACM, 1139–1152.

[9] Richard De Lisi and Susan L Golbeck. 1999. Implications of Piagetian theory for peer learning. (1999).

[10] Ronald Decker. 1995. Management team formation for large scale simulations. *Developments in Business Simulation and Experiential Learning* 22 (1995).

[11] Katherine Deibel. 2005. Team formation methods for increasing interaction during in-class group work. In *ACM SIGCSE Bulletin*, Vol. 37. ACM, 291–295.

[12] Tawanna R Dillahunt, Sandy Ng, Michelle Fiesta, and Zengguang Wang. 2016. Do Massive Open Online Course Platforms Support Employability?. In *Proceedings of the 19th ACM Conference on Computer-Supported Cooperative Work & Social Computing*. ACM, 233–244.

[13] Rong-En Fan, Kai-Wei Chang, Cho-Jui Hsieh, Xiang-Rui Wang, and Chih-Jen Lin. 2008. LIBLINEAR: A library for large linear classification. *The Journal of Machine Learning Research* 9 (2008), 1871–1874.

[14] Oliver Ferschke, Diyi Yang, Gaurav Tomar, and Carolyn Penstein Rosé. 2015. Positive impact of collaborative chat participation in an edx mooc. In *International Conference on Artificial Intelligence in Education*. Springer, 115–124.

[15] Barbara Gabbert, David W Johnson, and Roger T Johnson. 1986. Cooperative learning, group-to-individual transfer, process gain, and the acquisition of cognitive reasoning strategies. *The Journal of Psychology* 120, 3 (1986), 265–278.

[16] Elizabeth Avery Gomez, Dezhi Wu, and Katia Passerini. 2010. Computer-supported team-based learning: The impact of motivation, enjoyment and team contributions on learning outcomes. *Computers & Education* 55, 1 (2010), 378–390.

[17] Gahgene Gweon, Mahaveer Jain, John McDonough, Bhiksha Raj, and Carolyn P Rosé. 2013. Measuring prevalence of other-oriented transactive contributions using an automated measure of speech style accommodation. *International Journal of Computer-Supported Collaborative Learning* 8, 2 (2013), 245–265.

[18] Catherine M Hicks, Vineet Pandey, C Ailie Fraser, and Scott Klemmer. 2016. Framing Feedback: Choosing Review Environment Features that Support High Quality Peer Assessment. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*. ACM, 458–469.

[19] Inga J Hoever, Daan Van Knippenberg, Wendy P van Ginkel, and Harry G Barkema. 2012. Fostering team creativity: perspective taking as key to unlocking diversity's potential. *Journal of Applied Psychology* 97, 5 (2012), 982.

[20] Albert L Ingram and Lesley G Hathorn. 2004. Methods for analyzing collaboration in online. *Online collaborative learning: Theory and practice* (2004), 215–241.

[21] Mahesh Joshi and Carolyn Penstein Rosé. 2007. Using transactivity in conversation for summarization of educational dialogue.. In *SLaTE*. 53–56.

[22] Aniket Kittur. 2010. Crowdsourcing, collaboration and creativity. *ACM Crossroads* 17, 2 (2010), 22–26.

[23] René F Kizilcec. 2016. How Much Information?: Effects of Transparency on Trust in an Algorithmic Interface. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*. ACM, 2390–2395.

[24] Daphne Koller, Andrew Ng, Chuong Do, and Zhenghao Chen. 2013. Retention and intention in massive open online courses: In depth. *Educause Review* 48, 3 (2013), 62–63.

[25] Yasmine Kotturi, Chinmay Kulkarni, Michael S Bernstein, and Scott Klemmer. 2015. Structure and messaging techniques for online peer learning systems that increase stickiness. In *Proceedings of the Second (2015) ACM Conference on*

Learning@ Scale. ACM, 31–38.

[26] Chinmay Kulkarni, Julia Cambre, Yasmine Kotturi, Michael S Bernstein, and Scott Klemmer. 2015. Talkabout: Making Distance Matter with Small Groups in Massive Classes. In *CSCW*.

[27] Walter S Lasecki, Young Chol Song, Henry Kautz, and Jeffrey P Bigham. 2013. Real-time crowd labeling for deployable activity recognition. In *Proceedings of the 2013 conference on Computer supported cooperative work*. ACM, 1203–1212.

[28] RA Layton, MW Ohland, and H Pomeranz. 2007. Software for student team formation and peer evaluation: CATME incorporates Team-Maker. In *2007 ASEE Annual Conference*.

[29] Yi-Chieh Lee, Wen-Chieh Lin, Fu-Yin Cherng, Hao-Chuan Wang, Ching-Ying Sung, and Jung-Tai King. 2015. Using Time-Anchored Peer Comments to Enhance Social Interaction in Online Educational Videos. In *SIGCHI*. ACM, 689–698.

[30] Ioanna Lykourentzou, Angeliki Antoniou, Yannick Naudet, and Steven P Dow. 2016. Personality Matters: Balancing for Personality Types Leads to Better Outcomes for Crowd Teams. In *Proceedings of the 19th ACM Conference on Computer-Supported Cooperative Work & Social Computing*. ACM, 260–273.

[31] Ioanna Lykourentzou, Robert Kraut, and Steven P Dow. 2017. Team Dating Leads to Better Online Ad Hoc Collaborations. In *Proceedings of the ACM Conference on Computer Supported Cooperative Work & Social Computing*.

[32] Ioanna Lykourentzou, Shannon Wang, Robert E Kraut, and Steven P Dow. 2016. Team Dating: A Self-Organized Team Formation Strategy for Collaborative Crowdsourcing. In *Proceedings of the 2016 CHI Conference Extended Abstracts on Human Factors in Computing Systems*. ACM, 1243–1249.

[33] Stephen MacNeil, Celine Latulipe, Bruce Long, and Aman Yadav. 2016. Exploring Lightweight Teams in a Distributed Learning Environment. In *Proceedings of the 47th ACM Technical Symposium on Computing Science Education*. ACM, 193–198.

[34] Kalpani Manathunga and Davinia Hernández-Leo. 2015. Has research on collaborative learning technologies addressed massiveness? A literature review. *Journal of Educational Technology & Society* 18, 4 (2015), 357–370.

[35] James R Martin and Peter R White. 2003. *The language of evaluation*. Palgrave Macmillan.

[36] Larry K Michaelsen, Arletta Bauman Knight, and L Dee Fink. 2002. *Team-based learning: A transformative use of small groups*. Greenwood publishing group.

[37] Diana C Mutz. 2006. *Hearing the other side: Deliberative versus participatory democracy*. Cambridge University Press.

[38] Barbara Oakley, Richard M Felder, Rebecca Brent, and Imad Elhajj. 2004. Turning student groups into effective teams. *Journal of student centered learning* 2, 1 (2004), 9–34.

[39] Christian J Resick, Toshio Murase, Kenneth R Randall, and Leslie A DeChurch. 2014. Information elaboration and team performance: Examining the psychological origins and environmental contingencies. *Organizational Behavior and Human Decision Processes* 124, 2 (2014), 165–176.

[40] Keith Richards. 2006. *Language and professional identity: Aspects of collaborative interaction*. Palgrave Macmillan.

[41] Niloufar Salehi, Andrew McCabe, Melissa Valentine, and Michael Bernstein. 2017. Huddler: Convening Stable and Familiar Crowd Teams Despite Unpredictable Availability. In *Proceedings of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing (CSCW '17)*. ACM, New York, NY, USA, 1700–1713.

[42] Cass R Sunstein. 2006. *Infotopia: How many minds produce knowledge*. Oxford University Press.

[43] Yla R Tausczik and James W Pennebaker. 2013. Improving teamwork using real-time language feedback. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, 459–468.

[44] Stephanie D Teasley, Frank Fischer, Armin Weinberger, Karsten Stegmann, Pierre Dillenbourg, Manu Kapur, and Michelene Chi. 2008. Cognitive convergence in collaborative learning. In *International conference for the learning sciences*. 360–367.

[45] W Ben Towne and James D Herbsleb. 2012. Design considerations for online deliberation systems. *Journal of Information Technology & Politics* 9, 1 (2012), 97–115.

[46] Daan Van Knippenberg, Carsten KW De Dreu, and Astrid C Homan. 2004. Work group diversity and group performance: an integrative model and research agenda. *Journal of applied psychology* 89, 6 (2004), 1008.

[47] Miaomiao Wen, Diyi Yang, and Carolyn Penstein Rosé. 2015. Virtual Teams in Massive Open Online Courses. *Proceedings of Artificial Intelligence in Education* (2015).

[48] Betsy Anne Williams. 2015. Peers in MOOCs: Lessons Based on the Education Production Function, Collective Action, and an Experiment. In *Proceedings of the Second (2015) ACM Conference on Learning@ Scale*. ACM, 287–292.

[49] Anita Williams Woolley, Christopher F Chabris, Alex Pentland, Nada Hashmi, and Thomas W Malone. 2010. Evidence for a collective intelligence factor in the performance of human groups. *science* 330, 6004 (2010), 686–688.

[50] Zhilin Zheng, Tim Vogelsang, and Niels Pinkwart. 2014. The Impact of Small Learning Group Composition on Student Engagement and Success in a MOOC. *Proceedings of Educational Data Mining* 7 (2014).

[51] Haiyi Zhu, Steven P Dow, Robert E Kraut, and Aniket Kittur. 2014. Reviewing versus doing: Learning and performance in crowd assessment. In *Proceedings of the 17th ACM conference on Computer supported cooperative work & social computing*. ACM, 1445–1455.

Received June 2017; revised August 2017; accepted November 2017.