Does the transition to online meetings exacerbate or alleviate gender and racial inequities in student participation in project teams? Theory-guided machine-learning analysis of online collaborative work during the pandemic

Xu Dong[a], Victoria R. Nelson[b], Hanzhe Zhang[c], Young Anna Argyris[d*], Sinem Mollaoglu[e], Kenneth Frank[f], Arnav Jain[g], Xiao Qiao[h], Joseph Thekinen[i] and Haochen Liu[j]

[a]Department of Economics, Michigan State University, East Lansing, USA; [b]Department of Advertising + Public Relations, Michigan State University, East Lansing, USA; [c]Department of Economics, Michigan State University, East Lansing, USA; [d]Department of Media and Information, Michigan State University, East Lansing, USA; [e]Department of Construction Management, Michigan State University, East Lansing, USA; [f]Department of Counseling, Educational Psychology, and Special Education, Michigan State University, East Lansing, USA; [g]Department of Construction Management, Michigan State University, East Lansing, USA; [h]Department of Data Science, City University of Hong Kong, Hong Kong, People's Republic of China; [i]Department of Mechanical and Manufacturing Engineering, University of Calgary, Calgary, Canada; [j]Senior Data Analyst, Fidelity Investments, Massachusetts, USA

Contact: Young Anna Argyris, argyris@msu.edu, 404 Wilson Rd. Communication Arts and Sciences Building Michigan State University

Does the transition to online meetings exacerbate or alleviate gender and racial inequities in student participation in project teams? Theory-guided machine-learning analysis of online collaborative work during the pandemic

Online meeting modalities are believed to provide equal opportunities for participation to minorities. To examine whether they alleviated inequities, we observed students' participation in 17 project teams over 14 months. We analyzed each student's participation time and roles using machine-learning algorithms for voice recognition and role classification. These objective measures were complemented by self-reported responses of perceived contributions from 34 surveys. Our results indicate that females and racial minorities were affected heterogeneously. Females participated equally initially, but their participation decreased as the new modality settled. Contrarily, racial minorities were negatively affected, but these negative impacts decreased over time. Even so, females were perceived to contribute to teamwork as much as, while racial minorities were less than, their majority counterparts. These results align with the social affordance theory in that online modalities offer technical affordances for increasing equity, but, as the expectation states theory postulates, perceptions are yet to catch up.

Keywords: Project teams, online meetings, virtual meetings, video conference, natural language processing, student participation, females, minorities, COVID

Introduction

Background and Significance

While the COVID-19 pandemic imposed an abrupt transition to online modalities for student project teams, extant research has yet to explicate how collaborative work has changed group dynamics for team members, including those with diverse backgrounds. This dearth of research is primarily attributed to the absence of intelligent tools for unobtrusively collecting participation data (e.g., speech). Due to this challenge, many extant studies have relied solely on self-reported measures (Alharbi et al., 2021; Kaptelinin et al., 2021; Karl et al., 2021; Bayhan et al., 2022; Standaert and Thunus, 2022), which interferes with the fair observations of individuals from historically marginalized groups when used unaccompanied by objective measures. Furthermore, extant studies have examined a small number of teams in only one discipline, thus hindering the generalization of findings to diverse collaborations in various areas. Similarly, extant studies have observed group dynamics in a short period, often ranging from weeks to a month or two, which is insufficient to capture the changes in group dynamics over time. Consequently, these barriers have impeded obtaining generalizable scientific knowledge in this opportune time (i.e., the global pandemic and the accompanied transition to online) to expand theories on the effects of modality on team dynamics.

We have employed several novel approaches to alleviate these methodological issues and thus to make contributions to advancing theories relevant to human-computer interaction (HCI). First, we conducted a longitudinal study of student project work over the 14 months from March 2020 to April 2021, covering the onset, peaks, and troughs of changes due to COVID-19. Second, we collected data from multiple teams across disciplines—i.e., 17 undergraduate and graduate student project teams in social sciences and STEM majors. Third, we measured both objective participation behaviors and subjective/perceived contributions. To do so, we analyzed video/audio recordings of meetings using a machine-learning pipeline that consists of (i) speech diarization, speaker recognition, and transcription and (ii) classification of roles team members played in group work. To compare the objective participation behaviors to peers' perceived contributions, we conducted two surveys for every project team each of the four semesters—which amounts to 34 surveys in total. These expansive methodologies enabled us to examine whether the abrupt changes to online modalities have altered students' objective participation and subjective perceptions about their contributions to teamwork.

In so doing, we contribute to knowledge advancement in the field of HCI. Although a few extant studies on modality changes and teamwork have advocated the social affordances of video-conferencing tools designed to facilitate collaboration among team members (potentially those from underprivileged groups), it is unclear whether these technology affordances can overcome the existing inequity in teamwork. To answer this question, we employ a social psychological theory of expectation states that illuminates how one's expectations about group members affect their perceptions of group members. In other words, we use the expectation states theory to show how implicit beliefs surrounding immutable characteristics like gender and race impede the realization of the social affordances of synchronous video conferencing tools. As a result, this study demonstrates how technology interacts with social dynamics, thereby setting its boundary conditions of the technology affordance theory.

Eventually, this research illustrates how future researchers can utilize machine learning methods to automate the diagnoses of problems in project teams (Mikolov, 2013), which will be translated into groundwork to support diverse groups of individuals to work in complex social systems (Cross et al., 2010; Bayhan et al., 2022; Karl et al., 2022) and to create a more inclusive environment to reduce inequities (Wu et al., 2022; Bagmar et al., 2022, Do et al., 2022, Kim et al., 2021).

Literature Review and Gap in the Literature

Pre-pandemic Comparison Between Meeting Modalities and Inequities in Participation

Prior to the transformation of work and school to online modalities due to the COVID-19 pandemic, most HCI literature focused on online meetings in comparison to offline meetings and what the benefits and costs were between the two modalities. Settings for these comparisons include collaboration via online and technology-enabled meetings in education (Kinnula et al., 2018; Willermark and Pareto, 2020), healthcare (Islind et al., 2019; Constantinides, 2011), and co-authored writing (Larsen-Ledet and Korsgarrd, 2019). In educational collaboration, positive affordances of online meetings were identified: Participants communicate, share resources, engage in productive collaborative learning processes, monitor and regulate collaborative learning, and find and build groups and communities (Jeong and Hmelo-Silver, 2016).

However, other studies found that online meetings maintain the gender inequality found during in-person meetings. Dhawan et al. (2021) discussed gender inequity in in-person meetings, and hypothesized that many of those gender inequities were likely to transition into digital meetings, and suggested some video-conferencing behavioral etiquette to reduce this inequity. Other studies have found similar inequities both in-person and online for both racial and LGBTQ-identified minorities (Houtti et al., 2022). Likewise, during online meetings, there are significant mismatches in native English speakers' attributions of non-native speakers' behaviors, but no considerable mismatch exists in non-native speakers' attributions of native speakers' behaviors (He et al., 2017). Similarly, while team diversity is known to facilitate creativity (Ye and Robert Jr, 2017), intercultural conflict is found in global virtual teams and is shown to affect communication and project results negatively (He et al., 2017). Also, language and cultural barriers negatively affect minorities in terms of their ability to participate in communication and group tasks (He et al., 2017a, 2017b).

Pandemic Comparison between Meeting Modalities and Inequities in Participation

During the pandemic, a focus of many different disciplines' scholarship focused on how the transition to an online learning paradigm affects students' participation in learning processes and, thus, outcomes (Shu et al., 2023; Chen et al., 2021; Chen et al., 2021; Ravi et al., 2021; Lacy et al., 2022; Manshaei et al., 2022; Gui et al., 2022). Mixed results have been reported. Some studies have shown sustained disadvantages of online meetings in higher education compared to offline, co-located meetings (Kisworo et al. 2022). For instance, online meetings did not effectively support hands-on activities (Labrie et al., 2022) and the experience of "we-ness" (Kaptelinin et al., 2021) because online meetings made it harder to engage in private chats (Guo et al., 2022). In addition, the lack of structure in private chats is not conducive to reaching a consensus (Kim et al., 2021). A related study found that using text-based applications (e.g., Slack) for distributed teams decreases the ability to perceive, understand, and regulate emotions between team members (Benke et al., 2021).

Other studies (such as Houtti et al. 2022), contrarily, have maintained that some video-conferencing tools (e.g., the raise-hand feature) mitigate bias by giving members control over the meetings, thereby helping reduce adverse psychological outcomes caused by bias for employees. Finally, Liu et al (2021) investigated self-regulated learning during the pandemic, not necessarily teamwork but their results indicate that males adopted more behavioral strategies than females to deal with their disorientation during online learning, but that females were able to adapt better overall to the change.

Summary of Literature and Identified Gap

Many pre-pandemic studies found that online meetings had more disadvantages than offline, co-located conferences and that these disadvantages are often exacerbated for women and racial minorities. However, during the pandemic, extant studies have shown mixed results for these same groups. Some reported sustained disadvantages of online meetings, while others have advocated the benefits of synchronous video conferencing tools, especially in mitigating implicit biases. In general, very few studies have examined how these changing modalities interact with gender and race in group work. The above literature summary shows a need to investigate how the pandemic and its attendant changes in modalities have affected inequities in team participation.

Theoretical Framework and Hypothesis Development

In our investigations of the modality changes and their attendant inequity in group work, we resort to two strands of theories—(1) one that advocates the advantages of synchronous video conferencing tools on teamwork and (2) the other that pinpoints the disadvantages of such tools, especially for minoritized populations. The former consists of social affordance and social presence theories illuminating how information/communication technology facilitates interactions among individuals and members of a team. The latter comprises expectation states theory and its sub-theory, double standards theory, which sheds light on how implicit bias interacts with technology affordances.

Both strands of theories focus on the social aspects of group dynamics and have been used to study communication and group behaviors. We build upon these two strands of theories that advocate contrasting predictions of how synchronous video conferencing tools facilitate or hinder group work to solve conflicting and mixed results in the prior work.

Social affordances and presence in online meetings

Social affordances of technology refer to the capacity a technology or a medium provides individuals to create and maintain social interactions. Social affordances of synchronous videoconferencing technology platforms such as zoom identified in previous research include temporality, interactivity, multimediality, and portability (Yeshua-Katz et al., 2023). The social affordance theory, hence, provides theoretical support to the claims that the transition to online modalities during the pandemic facilitates teamwork, including for those from diverse backgrounds.

Similar theoretical support can be found in the social presence theory, albeit from a different angle. Social presence theory posits that different media offer distinct perceptions that other people are physically present during computer-mediated communication because of available social cues like eye contact, facial expressions, voice inflection, physical distance, and posture (Zelkowitz, 2010). These verbal and visual cues help to create and maintain productive social interactions in a group setting. Applying this theory to online meetings on video conferencing platforms, one can expect that all the individuals are given the same physical spaces (as in a window space assigned to each individual in Zoom meetings), and the ability to adjust the volume of their voices. This equal physical space and control over voice are contrasted to smaller physical space and lower volumes that some women (in case of physical space) and some minorities (in terms of volume) may take in-person meetings. In this way, a platform like

Zoom offers users the same opportunity to augment and create a social presence, which in turn has the potential to create an environment for more equal participation between group members.

Based on these social affordance and presence theories, one can expect that the modality change to online meetings during the pandemic has given minorities a chance to participate more and take up more active leadership roles than offline, in-person meetings before the pandemic. Simultaneously, empirical findings have not yet shown whether the potentially increased participation by minorities due to the facilitated social affordances and presence is sufficient to compensate for the pre-existing gap between them and majorities. As noted in the literature review section, Houtti et al. (2022) have maintained that some video-conferencing tools mitigate bias by giving members control over the meetings, as the social affordance theory has suggested. However, Houtti et al.'s (2022) study involved semi-structured interviews with only 22 individuals and hence cannot be generalized. Moreover, no empirical data shows whether the positive effects of social affordances/presence on minority participation are sustained throughout the pandemic or change as online meeting becomes a norm for project teams in a later part of the pandemic. Given this lack of empirical support, despite a plausible claim built on social affordances/presence theories supporting the positive effects of video conferencing tools on the equity in group work, we propose null hypotheses comparing minority members (women and racial minorities) and majority counterparts (men and non-Hispanic whites):

H1.1 There will be no difference in objective participation levels between female and male students. There will be no difference in objective participation during the transition to online meetings.

H1.2 There will be no difference in students' roles between female and male students. There will be no difference in roles during the transition to online meetings.

H2.1 There will be no difference in objective participation between majority and minority racial groups. There will be no difference in objective participation during the transition to online meetings.

H2.2 There will be no difference in students' roles (giving or asking for information) between majority and minority racial groups. There will be no difference in roles during the transition to online meetings.

Expectation states theory and double standards theory

Another network of theories states that group dynamics are shaped by social forces outside the communication modality and technology (moving from in-person to online). This outside set of social forces can be explained by two theories found in the social psychology literature: expectation states theory and its sub-theory, double standard theory.

Expectation states theory offers a plausible explanation of why this equal participation might not occur despite the social affordances provided and the presence created by synchronous video conferencing tools. Expectation states theory would posit that deeply and implicitly held expectations based upon beliefs about gender roles or stereotyping based upon race are deep-seated enough that this social psychological environment might overcome the advantages of synchronous video conferencing.

According to the expectation states theory, actors draw information from their social and cultural environment and then organize that information into expectations that dictate their interaction with others (Correll & Ridgeway, 2003). In expectation states research, the pressure for a team to reach a collective goal causes team members to forecast each member's future levels of contribution compared to other team members to decide how to complete a given task.

Expectation states theory historically has primarily been applied to gender and has invoked both positive as well as negative stereotypes or expectations.

A case of positive expectations occurs when team members think that one of the team members will perform a task better than others. In this case, the team will defer to this member who is perceived to be superior and will give that person more chances to participate and influence the group. The group anticipation of who will be the most competent is often driven by unconscious or implicit beliefs, which are split-second judgments about future contributions. As such, this positive stereotype explains why some majority, or high-status, group members seem to possess advantages over low-status group members, such as increased speaking time, automatic attention and validation to their ideas, and a perception that they are more influential than other lower-status group members.

In sum, the expectation states theory suggests a possibility that women are less favorably perceived than men in terms of the extent of contributions to group work. Simultaneously, the contributions of women who play the expected roles of supporters may still be perceived positively as those of men since the roles that those women play meet the other member's expectations. Again, there is no empirical study that examines how other members perceive women members compared to men counterparts during the transition to online meetings.

A case of negative stereotypes is often attendant with racial minorities. Racial minorities are often not expected to offer the same level of contributions to a group as majority members, so even when they do make equal contributions, others expect that they will not, leading to negative perceptions from teammates. The unique barriers facing students who identify as part of a racial minority can be further explained by a sub-theory of expectation states theory, double standard theory.

The double standard theory posits that lower-status members are given less of a chance to participate from the onset of group work and that when lower-status members are participating, they are evaluated more harshly than high-status individuals doing the same work. This results in reduced recognition for lower-status members by peers lead to a reduction in chances to become a high-status member. Correll and Ridgeway explained that because of these implicit beliefs which disadvantage racial minority members of the group that racial minority team members must contribute more to the group than majority race team members to be perceived as equally contributing to the group (2003).

As in all of the cases above, there is lack of empirical support to this double standard theory's utilities in group work during the transition from online to offline modalities. Given the lack of empirical support, instead of hypothesizing a directionality of an effect as in positive or negative, we propose a null hypothesis:

H3.1 There will be no difference in other team members' perceptions of female contributions compared to male counterparts'.

H3.2 There will be no difference in other team members' perceptions of racial minority members' contributions compared to majority counterparts'.

Materials and Methods

Data Collection

We developed a data collection protocol approved by the Institutional Review Board for Human Subject Research and collected data from students in courses adopting project team-based assignments in a large public university during four semesters: Spring 2020, Summer 2020, Fall 2020, and Spring 2021 (over the 14 months from March 2020 to April 2021). The teaching mode of these semesters was entirely or partially online, as shown in **Table 1**. Our team members recruited participants through voluntary and anonymous participation. Individual and team-participation based gift cards were utilized to improve the response rate. The sample covers 101 students from 17 teams across seven courses. **Table 2** presents summary statistics for different teams, including the semester, course type, team size, the number of videos and audios recorded, the average length of these recordings, and the span of the project. Course type captures whether the group project is a part of a course in economics or human resources and labor relations (HRLR) ("Social"), a course in computer science or engineering ("Technical"), or a course in civil engineering that combines Social and Technical dimensions ("Hybrid").

--- Insert Tables 1 and 2 Here ---

Machine Learning Analytical Pipeline

Prior researchers in HCI have called for machine-learning based speech recognition and speech diarization as a much-needed approach to investigating team dynamics for over 30 years (e.g., Bai et al., 2021; Zhao et al., 2015, Liao et al., 2022; Makhoul et al., 1992; Murviet, 1989). This call is to improve the accuracy and fluidity of diarization-created transcripts. A machine-learning automates laborious processes for manual recording, transcribing, and storing large amounts of unstructured datasets, such as video and audio tapes. That is, an automatic and analytical pipeline streamlines data flow to improve the speed and quality of data analyses, which is essential for facilitating a longitudinal study that involves multiple teams. As a result, a pipeline can aid in the early detection of team communication problems before the problems evolve into more significant issues detrimental to team functions (Arum and Roska, 2011). Indeed, Jagganath et al. (2018) validated this approach in a trauma-care unit in the healthcare setting. They used speech transcription and diarization to investigate the communication patterns between healthcare providers when deciding to resuscitate a patient. In addition, a pipeline allows for unobtrusive data collection, overcoming the well-known limitations of perceived measures for individual contributions to a team. Objective measures for participation eliminate subjectivity, implicit bias, and stereotypes among researchers and participants. Moreover, objectively measured participation can then be compared to subjective contributions collected through self-reported surveys, facilitating the discovery of implicit biases and stereotypes in teamwork.

Despite historically pressing needs and empirical validation in the healthcare setting, few studies have implemented and used machine-learning-based speech recognition and speech diarization to examine teamwork involving gender and racial minorities.

Speech Diarization, Speaker Recognition, and Transcription

To respond to this demand, we employed a machine learning pipeline to transform upstream original video and audio files to downstream intuitive tabulated data as described in the next section. Specifically, our machine-learning pipeline components include speech diarization, speaker recognition, transcription, and classification of participation (giving information, asking for information, and others).

The speech diarization component splits audio signals based on speaker identity. It first partitions the audio into small segments, each of which ideally contains a variable-length utterance from a unique speaker. Then, by comparing the audio signals in each segment, the module determines the number of speakers in a conversation and judges which speaker each segment belongs to. To perform this step, we applied a deep-learning-based speaker diarization model released in the Kaldi speech recognition toolkit (Povey et al., 2011). The model first encodes utterances to fixed-dimensional embeddings called "x-vectors" (Snyder et al., 2018), then performs agglomerative hierarchical clustering on x-vectors to produce an initial diarization output. Finally, a variational Bayes Hidden Markov model is applied over x-vectors to improve the diarization results. In our implementation, we use the pre-trained model from the Kaldi project (2021) and run it in ONNXRuntime (ONNXRuntime, 2021).

Following the speech diarization, the speech recognition component transcribes speakers' utterances in audio into text, which is next passed to the transcription component. The transcription component consists of three steps: (i) acoustic feature extraction, where an acoustic model is applied to convert raw audio signals into acoustic features; (ii) word selection, which chooses candidate words according to the acoustic features from the system's dictionary; and (iii) sentence-level matching, where a language model is applied to determine the final words based on their contexts. In our implementation, we use the pre-trained ASpIRE chain model (Povey, 2021) from the Kaldi project.

Classification of participants' roles

Following the transcription component, the classification component of the pipeline groups the transcribed text from speakers into three categories: giving information (G), asking for information (A), and others (O). This classification aims to determine whether the primary purpose of a transcribed text is to give information to others, ask for information from others, or neither (called "others"), to help us understand members' roles in project teams. This categorization has been used in prior studies that examined how members exchange information and collaborate to obtain a shared goal (Liang et al., 2017). GAO is an instrumental indicator of the effectiveness of knowledge-transfer processes, as group members must give information to others, not only ask for information from others (Argyris and Ransbotham, 2016). If group members only ask for information without giving information, knowledge transfer within the group is nonexistent, and accordingly, the group work cannot be sustained (Argyris and

Ransbotham, 2016). As such, we implemented GAO framework in our classification.

Specifically, we trained a deep-text classifier. Specifically, a 1-layer recurrent neural network (RNN) with gated recurrent units (GRU) is used as the text classification model. The size of word embeddings is set as 50, and the hidden size of the RNN is 128. The model is trained on 932 training instances manually labeled by human coders with a batch size of 64 for 20 epochs. As shown in **Table 3**, our classifier had an acceptable accuracy of 84.12, F1-macro of 80.70 and F1-micro for 84.12.

--- Insert Table 3 Here ---

Measure for the objective participation

We measured members' objective participation in team meetings with the percentage of speaking time generated by our speech-diarization algorithm as described in the section above. On average, across the pre-change and new-modality periods, each student speaks 18.71% of the total meeting duration. The percentage of speaking duration is unaffected by the length of the meeting, so it is more comparable across different teams than the raw speaking time.

Self-Reported Surveys

Peers' perception of a member's contribution to teamwork is measured based on self-reported survey data. For every project team in each semester, we conducted two surveys—(i) an entry survey to collect the students' demographic factors (e.g., gender, race, background) and (ii) a final survey to evaluate their perceived contributions to project teams. As there were 17 teams from which we collected data, there were a total of 34 online surveys distributed via e-mail over 14 months.

Measure for the perceived participation

Students were asked to evaluate their own and their teammates' contributions to the overall project. Perceived contribution has three sub-dimensions: proficiency (Kaufman et al., 1999; Ohland et al., 2012), adaptivity (Raelin et al., 2011; Mentzer et al., 2017), and proactivity (Raelin et al., 2011; Mentzer et al., 2017) as shown in **Table 4**. A measure for each category employs a 4-point Likert scale; a rating of 1 indicates a response of "strongly disagree," whereas a rating of 4 indicates "strongly agree." To eliminate the potential bias of self-evaluation, we use the average evaluation scores from the other team members as an individual's final score.

--- Insert Table 4 Here ---

The reliability of the 12-item scale was validated as follows. Cronbach's Alpha for the 12 items of 0.9757, indicating that the items have high internal consistency. A principal component analysis suggested that the scale was unidimensional. The results from our principal component analysis show that the first component has an eigenvalue of 9.72 which is much higher than the second eigenvalue (0.79). The first component explains 81% of the total variance. These analyses indicate that the scale is reliable and unidimensional.

Results

Descriptive Statistics and Model Specifications

Before testing the hypotheses, we visually examined objective participation behaviors across genders (*females* and *males*) and races (*minorities* vs. *majority*) in two periods: (1) *pre-change* (i.e., *Time 1* from Spring 2020 to Summer 2020 semesters) and (2) *new modality* (i.e., *Time 2* from Fall 2020 to Spring 2021). Recall that we operationalized participation type as speaking in general (*pdur*), giving (*givepdur),* and asking for information (*askpdur*), and measured them using the machine-learning pipeline. **Figures 1** and **2** show three types of participation by gender and race in two periods. As shown in **Figure 1**, gender appears to interact with the period such that the gap between females and males reversed from *pre-change* (*Time 1*) to the *new modality* (*Time 2*). Similarly, **Figure 2** shows the gap between minority and majority students decreased drastically in *Time 2* compared to *Time 1*, suggesting a potential interaction between minority statuses and modality change.

--- Insert Figures 1 and 2 Here ---

Given these likely interactions, we built two linear regression models—i.e., one for testing the main effects only (*Model 1*) and the second for accounting for interaction effects (*Model 2*). We also chose the ordinary least squares technique (OLS) for estimation because OLS regressions allow for controlling factors that may affect outcome variables, such as team size in our study (Cohen et al., 2013). Our two models, therefore, demonstrate how the modality change interacted with students' gender and minority statuses over time, exerting heterogeneous influences on their objective participation (*pdur* in **H1**; *givepdur* and *askpdur* in **H2**) and *perceived contributions* (**H3**).

Our first regression model (i.e., *Model 1*) focuses on the main effects as follows:

$$y_i = \beta_0 + \beta_1 * \textit{new-modality} + \beta_2 * female + \beta_3 * \textit{minority} + \beta_4 * size + \epsilon_i \qquad (1)$$

Our second regression model (i.e., *Model 2*) includes the interaction terms between the *new modality* and *gender* and *minority* statuses.

$$y_i = \beta_0 + \beta_1 * \textit{new-modality} + \beta_2 * female + \beta_3 * \textit{new-modality} * \textit{female} + \beta_4 * \textit{minority} + \beta_5 * \textit{new-modality} * \textit{minority} + \beta_6 * size + \epsilon_i \qquad (2)$$

Detailed descriptions and summary statistics of the data can be found in **Table 5**. The primary independent variable of interest is the change in the meeting modality (henceforth, "*new* *modality*”). The first interaction term, *new-modality**female*, indicates the change in gender gap (defined as female participation minus male participation) in the *new-modality* period (*Time 2*) compared with the *pre-change* period (*Time 2*). Likewise, the second interaction term, *new-modality* * *minority*, indicates the change of racial gap (defined as *minority* students' participation minus *majority* students' participation) in the *new modality* period (*Time 1*) compared with the *pre-change* period (*Time 2*).

--- Insert Table 5 Here ---

Hypothesis Testing Results

Hypothesis 1 states that there is no difference between *male* and *female* students regarding their objective speech time (*pdur*) and roles (*givepdur* and *askpdur*), obtained by our machine-learning analytical pipeline. **Table 6** below presents the results of two regression models per each outcome variable.

--- Insert Table 6 Here ---

We also conducted the regressions that controlled for course materials (technical and social) and student level (undergraduate or graduate student), but their estimated coefficients were significantly less than their reported standard errors. Adding them does not change our substantive conclusions, so we did not include them in our reported analyses for the parsimony of our models. For comparison, the regression results that include these three control variables are shown in **Appendix**.

The main effects of gender on *pdur* are shown in column 1 of **Table 6**. *Female* status does not have a significant influence on *pdur* (β = - 2.442, Standard Error [SE] = 2.040, p = 0.232, *n.s.*). This non-significant result means that the gap between *males* and *females* was not significant; in other words, female students participated as much as male students did. Likewise, the interaction effects between the *new modality* and *gender* (*Model 2*) were not significant (β = - 6.396, SE = 4.610, p = 0.166, *n.s.*), suggesting that the change of gender gap from *Time 1* to *Time 2* was not statistically significant. Although descriptively the slopes diverged, this could have occurred by chance alone if, in fact, the lines were parallel.

Figure 3 further illustrates the gender gap regarding *pdur* between the pre-change and new-modality periods. **Figure 3(a)** shows that the gender gap measured in *pdur* changed from 2.39 percentage points in *Time 1* (23.73 (female) vs. 21.34 (male)) to – 4 percentage points in *Time 2* (15.97 (female) vs. 19.97 (male)). These results, along with some signs of interactions between genders in **Figures 1 and 2**, suggest a possibility that there was interaction between *gender* and *new modality*, despite the non-significant main and the interaction effects.

--- Insert Figure 3 Here ---

Hence, we compared females and males using our regression *Model 1* in *Time 2* only because testing with *Model 1* allows us to control for the same control variables in all hypothesis tests, an option not available in a simple *t*-test. Our result showed a marginally significant difference at the .10 level between *females* and *males* in Time 2 (β = - 4.135^{+}, SE = 2.321, p = 0.076). Accordingly, **H1.1** was partially rejected, which means that, although there was no gender gap of *pdur* in the *pre-change* period, the gender gap may have changed as the modality changed to online, such that *female* students' initial participation through speaking decreased compared to their *male* counterparts did in the *new-modality* period.

Hypothesis 1.2 states no differences in roles that *females* and *males* played, which were answered in columns 3 (*givepdur*) and columns 5 (*askpdur*) in **Table 6**, respectively. *Model 1*s shows that *female* status had no influence on either *givepdur* (β = -1.859, SE =1.754, p = 0.290, *n.s.*) or *askpdur* (β = - 0.350, SE = 0.474, p = 0.461, *n.s*). These non-significant results mean that the gender gap in terms of giving and asking for information was not significant: in other words, female students both gave and asked for information as much as their male counterparts did. Likewise, as shown in model 2s, the interaction effects between *new modality* and gender on *givepdur* (β = - 5.248, SE = 3.858, *p* = 0.174, *n.s*) and *askpdur* (β = - 1.423, SE = .924, *p* = 0.122, *n.s*) were not significant.

Figures 3(b) and **3(c)** show a possibility that, as the modality changed, female students' role as information-givers declined more considerably than their roles as information-askers, as indicated by the steeper slope in **Figure 3(b)** than that in **Figure 3(c)**. Given these signs of interaction effects in Figures **3(b)** and **3(c)**, despite non-significant main and interaction effects, we compared females' and males' roles using our regression *Model 1*, as we did for **Hypothesis 1.1**. Our results once again showed non-significant differences in terms of roles that *females* and *males* played in *Time 2* (*givepdur*: β = - 3.298, SE = 2.018, *p* = 0.103, *n.s.*; *askpdur*: β = - 0.710, Standard Error [SE] = 0.585, *p* = 0.226, *n.s.*). Accordingly, **H1.2,** which states no difference in roles played between *females* and *males*, was supported.

Hypotheses 2 states there is no difference in *pdur*, *givepdur*, or *askpdur* between *minorities* and *majorities*. Column 1 in **Table 6** indicates that the racial gap in terms of *pdur* was marginally significant at the significance level of .10 ($\beta = 3.591^{+}$, SE = 2.014, *p* = 0.075). Likewise, the interaction effects between the *new modality* and race on *pdur* was significant ($\beta = 9.070$, SE = 4.486, *p* = 0.044*), which shows that the racial gap changed significantly as the modality changed to online such that minority students' participation increased more than majority students' participation increase in the new modality period. **Figure 4(a)** provides graphic support for this finding. Therefore, we proceeded with a comparison test using our regression *Model 1*, as we did for Hypotheses **1**. Our results confirmed no significant racial difference in *pdur* in *Time 2* (β = - 1.637, SE = 2.292, *p* = 0.476, *n.s.*). Accordingly, our **H2.1** was rejected. There was a significant gap in *pdur* in the *pre-change* period, but the racial gap was almost closed as the meeting modality changed to online such that minority students' *pdur* increased while majority students' *pdur* did not increase at the same pace.

--- Insert Figure 4 Here ---

Next, we tested **H2.2** regarding the differences in *givepdur* and *askpdur* between minority and majority students. As shown in *Model 1*s in columns 3 and 5, *minority* had no main effects on *givepdur* (β = - 2.598, SE = 1.695, *p* = 0.126, *n.s.*), but has a marginally significant negative effect on *askpdur* at the level of .10 (β = - 0.850^{+}, SE =0.511, *p* = 0.097). This result means that minority students gave as much information as majority students but asked for information less than majority students did. The interaction effects between *new modality* and *minority* (*Model 2*s in columns 4 and 6) were marginally significant on *givepdur* (β = 6.962+, SE = 3.766, *p* = 0.065) and on *askpdur* (β = 1.494+, SE = .859, *p* = 0.083). These results mean that the racial gap in terms of *givepdur* and *askpdur* changed significantly as the new modality set in, such that the gap in giving (albeit its statistically non-significant main effect) and asking for information closed substantially.

Figures 4(b) and **4(c)** further illustrate the racial gap change between *pre-change* and *new-modality* period. As such, we compared the racial differences regarding their roles as giving and asking for information in *Time 2*, using our regression *Model 1*. Our results confirmed that there was no racial difference in terms of roles that minority and majority students played in *Time 2* (*givepdur*: β = - 1.195, SE = 1.942, *p* = 0.539, *n.s.*; *askpdur*: β = - 0.505, SE = 0.640, *p* = 0.43, *n.s.*). Accordingly, **H2.2,** stating no difference in *majority* and *minority* students' roles, was partially rejected such that the racial gap in giving and asking for information decreased substantially as the modality changed.

Finally, we tested our **H3**, which concerns the differences in *perceived contributions* between *female* and *male* students. As explained in **Table 4, *Perceived Contributions to Teams***, in the section, **Self-Reported Surveys**, the outcome variable in Equation (2) is the *perceived contribution* score of an individual team member obtained from our surveys. As noted in that section, this measure reached an acceptable level of reliability. **Table 7** shows the results of our two regression models.

--- Insert Table 7 Here ---

Column 1 in **Table 7** shows that the main effects of *female* on *perceived contributions* were not significant ($\beta = 0.0415$, SE = 0.085, $p = 0.626$, *n.s.*). This result means that female students were perceived to make equal contributions to those of their male counterparts. The interaction effects between the *new modality* and *female* (Column 2 in **Table 7**) were also not significant ($\beta = .137$, SE = .179, $p = 0.446$, *n.s.*), which means that the gender gap did not change significantly in *Time 2* compared with *Time 1* but rather remained constant. As visibly apparent in **Figure 5**, the gender differences in *perceived contributions* were nominal. To confirm these non-significant main and interaction effects, we compared the difference in *perceived contribution* between *females* and *males*, using our regression *Model 1*. Our test result confirmed no difference between the two genders ($\beta = 0.058$, SE = 0.114, $p = 0.613$, *n.s.*). Therefore, **H3.1,** stating no gender differences in *perceived contribution*, was supported.

--- Insert Figure 5 Here ---

Finally, we tested **H3.2**, which states no difference in *perceived contributions* between *minority* and *majority*. As shown in *Model 1*, the main effects of *minority* had a significantly negative influence on *perceived contributions* (β = -0.292, SE = 0.511, p = 0.015*). As shown in *Model 2*, the interaction effects between the *new modality* and *minority* were not significant (β = 0.000421, SE = 0.239, p = 0.999, *n.s.*). These significant main effects and non-significant interaction effects indicate that the racial gap in terms of *perceived contributions* persisted regardless of the modality change. **Figure 6** further shows the racial gap between the *pre-change* and *new-modality* periods: The racial gap in *perceived contributions* appears consistent in both periods. As such, we compared the difference between the *majority* and *minority* regarding *perceived contribution* using our regression *Model 1* in *Time 2*. The results confirmed a marginally significant racial difference at the .10 level in *Time 2* (β = - 0.317^{+}, SE = 0.167, p = 0.063). Hence, our hypothesis **3.2**, which states no racial difference in *perceived contribution*, was rejected such that minorities were perceived as making fewer contributions than their majority counterparts were throughout time.

--- Insert Figure 6 Here ---

Robustness check for all research questions

Even though we attempt to control the factors that could affect *pdur* across teams (e.g. team size, class type), team size may still pose a systematic effect on participation. To address this concern, we construct two normalized versions of *pdur* and repeat our analyses. The measure *pdur1* is normalized by the entire sample, as shown:

$$\mathrm{pdur}_1 = \frac{\mathrm{pdur} - \mathrm{mean(pdur)}}{\mathrm{sd(pdur)}} \tag{3}$$

The measure *pdur2* considers that group size is, by construction, inversely related to the average speaking time in percentage terms. We subtract the group mean from *pdur* (100/team size) and divide the normalized standard deviation from the team mean:

$$\mathrm{pdur}_2 = \frac{(\mathrm{pdur} - 100 / \mathrm{team\ size})}{\mathrm{sd}\ (\mathrm{pdur\ within\ team\ within\ meeting})} \tag{4}$$

We did not observe any noticeable differences in regression results compared with the main results presented above, illustrating that our findings are robust to alternative participation measures. The detailed robustness results can be found in **Appendix**.

Discussion

Summary of the Findings

From a unique dataset collected at a large public university from March 2020 to April 2021, we examine how the COVID-19 pandemic and its imposed transition to online meetings have affected students' participation in project teams. In this analysis, we divide the dataset into two periods: the pre-change period in which the initial transition to online meetings occurred (Spring 2020 and Summer 2020), and the new-modality period in which students had acclimated to online meetings during the Fall 2020 and Spring 2021 semesters. Our results indicate that female and minority students were affected by the transition to online meetings due to the COVID-19 pandemic. However, these groups were not affected in the same way. For female students, no significant differences existed between female and male students in terms of the roles they played or the way they were perceived. However, there is still some sign of gender inequality because female speaking time reduced over time. However, critically, they were still perceived by their peers to be making the same amount of contributions as their male counterparts. So, the change of modality did not seem to affect female versus male student participation beyond this decrease in speaking time. In the case of racial minority students, there was time and role gaps between them and their white peers during the pre-change period. Initially, minority students participated less compared to white peers. However, after new modality settled in, minority students closed the gap in objective participation with their white peers, but, critically, were perceived by their peers to be making less of a contribution. So, the modality change, overall, seemed to have a mild equalizing effect in objective participation for disadvantaged groups, but there was still evidence of both gender and racial inequity.

Theoretical Contributions

Overall, the results from this study seem to support the social affordances of technology and social presence theories because a lack of significant objective participation gaps for both female and racial minority students greater equity after the change to a synchronous online modality. These theories posit that the advancement of videoconferencing technology and the affordances that are gained by synchronous online meetings due to the use of a platform like Zoom allow for more equity within groupwork. This supports the continued use of offline, synchronous video meetings as a viable alternative to offline, collocated meetings for diverse groups.

Simultaneously, our findings indicate the discrepancy between the objective participation behaviors and the perceived contributions for racial minority students. This gap between actual and perceived contribution gives credence to the double standard theory and expectation states theory. The results indicate the lack of inequity in actual contributions in online meetings was not able to overcome barriers faced due to pre-existing social bias for perceived contributions. The misalignment of objective participation and subjective evaluation (in the form of perceived contributions) can be explained by expectation states theory. Correll and Ridgeway (2006) ascertain that when lower status members, those from racial minorities, do participate, their performances are evaluated by a stricter standard, and their participation is perceived to be less than their majority peers. As such, both of our results—i.e., non-significant main effects of gender and race yet the disparity between objective and subjective measures of participation for racial minority students—help set the boundary conditions for the social affordances and presence theories. That is, technology affordances to facilitate social interactions are overcome by the existing stereotype against racial minorities.

Practical Implications

The discrepancy between objective and subjective measures of participation shows how difficult it is for competent performances by lower status members to be noticed as such, which further reduces their ability to achieve high status in the group. This raises the concern about equality during evaluation and promotes objective measures as an alternative way to conduct evaluation of participation level in teamwork. We also demonstrate the promise of the machine-learning pipeline to generate speaker diarization for analyzing team members' participation. The generated G/A/O information classification enables early detection of member roles and thus identifications of potential communication problems and inequity in participation. The machine-learning pipeline also has allowed for unobtrusive data collection on participation levels and allowed this study to examine questions of race and gender with data less tinged by implicit bias thus improving the study's ability to diagnose problems and perhaps increase equity in diverse teams. Our multi-method approach combining these machine-generated results with survey results can construct progress loops for teams and detect issues in team coordination and communication.

Limitations and Suggestions for Future Research

As with any research, there are some potential limitations to this study which can be addressed in future research. Future researchers may consider recruiting more underrepresented minorities to be able to have more analyses about individual racial minority populations that are sufficiently powered since we are aware of the important cultural differences and perceptions of different racial minority populations. Specifically, future studies should include more demographic information pertaining to participants' home life, level of income, and technology access, as these data points could illuminate whether racialized gaps in income and technology access affected participation level. If true, these findings might indicate that future solutions should continue to be sensitive to the diversity of historically marginalized populations and should integrate the idea that a multitude of interventions might be necessary to increase equity amongst culturally distinct subpopulations. Given our results showing some inequity in participation and definite inequity in perceived contributions, future research may examine and develop a set of interventions, which include (i) team meetings to set milestones at the beginning of the project, and (ii) individual-level nudges to improve individual—and ultimately team—performance, among others. These milestones and nudges can depend on demographics and other performance-related characteristics.

Conclusion

In this study, we evidenced how changing modalities from March 2020 until April 2021 affected both female and minority student participation in project teams using the observations obtained from a machine-learning pipeline and self-reported perceptions. This paper contributes to the expansion of social affordance and presence theories by setting the boundary conditions for equalizing effects of online modalities. The results show that researchers must avoid the uniform understanding of inequities among various subpopulations of historically marginalized groups and instead consider the intersections between these subpopulations and other socioeconomic factors as postulated by expectation states and double standard theories.

References

Alharbi, R. A., Altayyari, F. T., Alamri, F. S., & Alharthi, S. A. (2021, October). Pandemic-Driven Technology During COVID-19: Experiences of Older Adults. In *Companion Publication of the 2021 Conference on Computer Supported Cooperative Work and Social Computing* (pp. 5-9).

Argyris, Y. A., & Ransbotham, S. (2016). Knowledge Entrepreneurship: Institutionalising Wiki-based Knowledge-management Processes in Competitive and Hierarchical Organisations. Journal of Information Technology, 31(2), 226–239.

Arum, R., & Roksa, J. (2011). *Academically adrift: Limited learning on college campuses.* University of Chicago Press.

Ayobi, A., Eardley, R., Soubutts, E., Gooberman-Hill, R., Craddock, I., & O'Kane, A. A. (2022). Digital Mental Health and Social Connectedness: Experiences of Women from Refugee Backgrounds. *Proceedings of the ACM on Human-Computer Interaction*, 6(CSCW2), 1-27.

Bagmar, A., Hogan, K., Shalaby, D., & Purtilo, J. (2022). Analyzing the Effectiveness of an Extensible Virtual Moderator. *Proceedings of the ACM on Human-Computer Interaction*, 6(GROUP), 1-16.

Bayhan, H. G., Mollaoglu, S., Zhang, H., & Frank, K. A. (2022). Project Team Collaborations during Time of Disruptions: Transaction Costs, Knowledge Flows, and Social Network Theory Perspective. In *Construction Research Congress (CRC) on Project Management and Delivery, Contracts, and Design and Materials* (pp. 1012-1023).

Benke, I., Knierim, M. T., & Maedche, A. (2020). Chatbot-based emotion management for distributed teams: A participatory design study. *Proceedings of the ACM on Human-Computer Interaction*, 4(CSCW2), 1-30.

Bird, S. (2006). NLTK: the natural language toolkit. In Proceedings of the COLING/ACL 2006 Interactive Presentation Sessions. 69–72.

Chen, X., Chen, S., Wang, X., & Huang, Y. (2021). " I was afraid, but now I enjoy being a streamer!" Understanding the Challenges and Prospects of Using Live Streaming for Online Education. *Proceedings of the ACM on Human-Computer Interaction*, 4(CSCW3), 1-32.

Chen, Z., Cao, H., Deng, Y., Gao, X., Piao, J., Xu, F., ... & Li, Y. (2021, May). Learning from home: A mixed-methods analysis of live streaming based remote education experience in Chinese colleges during the COVID-19 pandemic. In *Proceedings of the 2021 CHI Conference on human factors in computing systems* (pp. 1-16).

Cohen, J., Cohen, P., West, S. G., & Aiken, L. S. (2013). Applied multiple regression/correlation analysis for the behavioral sciences. Routledge.

Copur-Gencturk, Y., Cimpian, J. R., Lubienski, S. T., & Thacker, I. (2020). Teachers' Bias Against the Mathematical Ability of Female, Black, and Hispanic Students. Educational Researcher, 49(1), 30–43.

Correll, S. J., & Ridgeway, C. L. (2006). Expectation states theory. Handbook of social psychology, 29-51.

Cross, R., Gray, P., Cunningham, S., Showers, M., & Thomas, R. J. (2010). How to make employee networks really work. *MIT Sloan management review*, 52(1), 83-90.

Daphna Yeshua-Katz, Stav Shapira, Limor Aharonson-Daniel, A. Mark Clarfield & Orly Sarid (2023) Matching Digital Intervention Affordances with Tasks: The Case of a Zoom and WhatsApp Mental Health Intervention for Seniors during the COVID-19 Pandemic, Health Communication, 38:3, 499-511.

Deming, D. J. (2017). The growing importance of social skills in the labor market. The Quarterly Journal of Economics, 132(4), 1593-1640.

Dhawan, N., Carnes, M., Byars-Winston, A., & Duma, N. (2021). Videoconferencing Etiquette: Promoting Gender Equity During Virtual Meetings. *Journal of Women's Health*, 30(4), 460–465.

Do, H. J., Kong, H. K., Lee, J., & Bailey, B. P. (2022). How Should the Agent Communicate to the Group? Communication Strategies of a Conversational Agent in Group Chat Discussions. *Proceedings of the ACM on Human-Computer Interaction*, 6(CSCW2), 1-23.

Garcia, A. J., Mollaoglu-Korkmaz, S., & Miller, V. D. (2014). Progress loops in interorganizational project teams: An IPD case. In *Construction Research Congress 2014: Construction in a Global Network* (pp. 2011-2020).

Garicano, L. (2000). Hierarchies and the Organization of Knowledge in Production. *Journal of political economy*, 108(5), 874-904.

Garicano, L., & Rossi-Hansberg, E. (2015). Knowledge-based hierarchies: Using organizations to understand the economy. *Annual Review of Economics*, 7(1), 1-30.

Gui, X., Li, Y., & Wu, Y. (2021). Teacher-Guardian Collaboration for Emergency Remote Learning in the COVID-19 Crisis. *Proceedings of the ACM on Human-Computer Interaction*, 5(CSCW2), 1-26.

Guo, J., & Fussell, S. R. (2022). " It's Great to Exercise Together on Zoom!": Understanding the Practices and Challenges of Live Stream Group Fitness Classes. *Proceedings of the ACM on Human-Computer Interaction*, 6(CSCW1), 1-28.

Grønbæk, J. E., Rasmussen, M. K., Halskov, K., & Petersen, M. G. (2020, April). Kirigami Table: Designing for proxemic transitions with a shape-changing tabletop. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems* (pp. 1-15).

He, H. A., Yamashita, N., Hautasaari, A., Cao, X., & Huang, E. M. (2017, February). Why did they do that? Exploring attribution mismatches between native and non-native speakers using videoconferencing. In *Proceedings of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing* (pp. 297-309).

He, H. A., Yamashita, N., Wacharamanotham, C., Horn, A. B., Schmid, J., & Huang, E. M. (2017). Two Sides to Every story: Mitigating intercultural conflict through automated feedback and shared self-reflections in global virtual teams. *Proceedings of the ACM on Human-Computer Interaction*, 1(CSCW), 1-21.

Houtti, M., Zhou, M., Terveen, L., & Chancellor, S. (2022). "All of the White People Went First": How Video Conferencing Consolidates Control and Exacerbates Workplace Bias. *arXiv preprint arXiv:2212.00849*.

Hutto, C., & Gilbert, E. (2014, May). Vader: A parsimonious rule-based model for sentiment analysis of social media text. In *Proceedings of the international AAAI conference on web and social media* (Vol. 8, No. 1, pp. 216-225).

Hy Murveit. (1989). Real-time speech recognition systems. In Proceedings of the workshop on Speech and Natural Language (HLT '89). Association for Computational Linguistics, USA, 458.

Islind, A. S., Snis, U. L., Lindroth, T., Lundin, J., Cerna, K., & Steineck, G. (2019). The virtual clinic: two-sided affordances in consultation practice. *Computer Supported Cooperative Work (CSCW)*, *28*(3), 435-468.

Jeong, H., & Hmelo-Silver, C. E. (2016). Seven affordances of computer-supported collaborative learning: How to support collaborative learning? How can technologies help? *Educational Psychologist*, 51(2), 247-265.

John Makhoul and Richard Schwartz. (1992). Robust continuous speech recognition. In Proceedings of the workshop on Speech and Natural Language (HLT '91). Association for Computational Linguistics, USA, 464.

Junwei Liao, Sefik Eskimez, Liyang Lu, Yu Shi, Ming Gong, Linjun Shou, Hong Qu, and Michael Zeng. (2023). Improving Readability for Automatic Speech Recognition Transcription. ACM Trans. Asian Low-Resour. Lang. Inf. Process. 22, 5, Article 142 (May 2023), 23 pages.

Kaptelinin, V., Danielsson, K., Kaiser, N., Kuenen, C., & Nordin, M. (2021, October). Understanding the Interpersonal Space of Online Meetings: An Exploratory Study of “We-ness”. In *Companion Publication of the 2021 Conference on Computer Supported Cooperative Work and Social Computing* (pp. 79-83).

Karl, K. A., Peluchette, J. V., & Aghakhani, N. (2022). Virtual work meetings during the COVID-19 pandemic: The good, bad, and ugly. *Small Group Research*, 53(3), 343-365.

Kaufman, D.B., Felder, R.M. and Fuller, H. (1999). Peer ratings in cooperative learning teams. Annual American Society for Engineering Education Meeting Proceedings of the 1999 Annual ASEE Meeting.

Kim, S., Eun, J., Seering, J., & Lee, J. (2021). Moderator chatbot for deliberative discussion: Effects of discussion structure and discussant facilitation. *Proceedings of the ACM on Human-Computer Interaction*, 5(CSCW1), 1-26.

King, M. M., & Frederickson, M. E. (2021). The pandemic penalty: the gendered effects of COVID-19 on scientific productivity. *Socius*, 7, 23780231211006977.

Kinnula, M., Iivari, N., Isomursu, M., & Laari-Salmela, S. (2018). 'Worksome but Rewarding'– Stakeholder Perceptions on Value in Collaborative Design Work. *Computer Supported Cooperative Work (CSCW)*, *27*(3), 463-494.

Kirshenbaum, N., Davidson, K., Harden, J., North, C., Kobayashi, D., Theriot, R., ... & Leigh, J. (2021). Traces of Time through Space: Advantages of Creating Complex Canvases in Collaborative Meetings. *Proceedings of the ACM on Human-Computer Interaction*, 5(ISS), 1-20.

Kisworo, A. Y., Wardhanie, A. P., & Erdiana, S. P. (2021, September). One Year Reflection of Online Learning: The Students' Perceptions of the Effectiveness of Online Learning amidst the COVID-19 Pandemic in a University. In *Proceedings of the 5th International Conference on Learning Innovation and Quality Education* (pp. 1-7).

Labrie, A., Mok, T., Tang, A., Lui, M., Oehlberg, L., & Poretski, L. (2022). Toward Video-Conferencing Tools for Hands-On Activities in Online Teaching. *Proceedings of the ACM on Human-Computer Interaction*, 6(GROUP), 1-22.

Lacy, A., Polsley, S., Ray, S., & Hammond, T. (2022). A seat at the virtual table: Emergent inclusion in remote meetings. *Proceedings of the ACM on Human-Computer Interaction*, 6(CSCW2), 1-20.

Larsen-Ledet, I., & Korsgaard, H. (2019). Territorial functioning in collaborative writing. *Computer Supported Cooperative Work (CSCW)*, *28*(3), 391-433.

Liang, Y. (2017, February). Knowledge sharing in online discussion threads: What predicts the ratings? In *Proceedings of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing* (pp. 146-154).

Liu, X., He, W., Zhao, L., & Hong, J. C. (2021). Gender differences in self-regulated online learning during the COVID-19 lockdown. *Frontiers in Psychology*, 12.

MacArthur, C., Wong, C., & Hancock, M. (2019). Makers and quilters: Investigating opportunities for improving gender-imbalanced maker groups. *Proceedings of the ACM on Human-Computer Interaction*, 3(CSCW), 1-24.

Manshaei, R., Mayat, U., Imtiaz, S. A., Andric, V., Aliar, K., Abu Hantash, N., ... & Mazalek, A. (2022). Tangible Chromatin: Tangible and Multi-surface Interactions for Exploring Datasets from High-Content Microscopy Experiments. *Proceedings of the ACM on Human-Computer Interaction*, 6(ISS), 51-72.

McInnis, B., Xu, X., & Dow, S. P. (2018). How features of a civic design competition influence the collective understanding of a problem. *Proceedings of the ACM on Human-Computer Interaction*, 2(CSCW), 1-25.

Mentzer, N., Laux, D., Zissimopoulos, A., & Richards, K. A. R. (2016). Peer evaluation of team member effectiveness as a formative educational intervention. *Journal of Technology Education*, 28, 53-82.

Mikolov, T., Sutskever, I., Chen, K., Corrado, G. S., & Dean, J. (2013). Distributed representations of words and phrases and their compositionality. *Advances in neural information processing systems*, 26.

Morris, M. R., Morris, D., & Winograd, T. (2004, November). Individual audio channels with single display groupware: effects on communication and task strategy. In *Proceedings of the 2004 ACM conference on Computer supported cooperative work* (pp. 242-251).

Nepal, S., Wang, W., Vojdanovski, V., Huckins, J. F., DaSilva, A., Meyer, M., & Campbell, A. (2022, April). COVID Student Study: A Year in the Life of College Students during the COVID-19 Pandemic Through the Lens of Mobile Phone Sensing. In *CHI Conference on Human Factors in Computing Systems* (pp. 1-19).

ONNXRuntime. (201) Optimize and accelerate machine learning inferencing and training. https://github.com/Microsoft/onnxruntime

Ohland, M. W., Loughry, M. L., Woehr, D. J., Bullard, L. G., Felder, R. M., Finelli, C., Layton, R., Pomeranz, H., & Schmucker, D. G. (2012). 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.

Pan Zhou, Hui Jiang, Li-Rong Dai, Yu Hu, and Qing-Feng Liu. (2015). State-clustering based multiple deep neural networks modeling approach for speech recognition. IEEE/ACM Trans. Audio, Speech and Lang. Proc. 23, 4 (April 2015), 631–642.

Povey, D. (2021). ASpIRE Chain Model. http://kaldi-asr.org/models/m1

Povey, D., Ghoshal, A., Boulianne, G., Burget, L., Glembek, O., Goel, N., Hannemann, M., Motlicek, P., Qian, Y., Schwarz, P., et al. 2011. The Kaldi speech recognition toolkit. In *IEEE 2011 workshop on automatic speech recognition and understanding*. IEEE Signal Processing Society.

Raelin, J. A., Bailey, M., Hamann, J., Pendleton, L., Raelin, J., Reisberg, R., & Whitman, D. (2011). The effect of cooperative education on change in self-efficacy among undergraduate students: Introducing work self-efficacy. *Journal of Cooperative Education and Internships*, 45(2), 17-35.

Ravi, P., Ismail, A., & Kumar, N. (2021). The pandemic shift to remote learning under resource constraints. *Proceedings of the ACM on Human-Computer Interaction*, 5(CSCW2), 1-28.

Riedl, C., Kim, Y. J., Gupta, P., Malone, T. W., & Woolley, A. W. (2021). Quantifying collective intelligence in human groups. *Proceedings of the National Academy of Sciences*, 118(21), e2005737118.

Shi, W. W., Jagannadharao, A., Lee, J., & Bailey, B. P. (2021). Challenges and Opportunities for Data-Centric Peer Evaluation Tools for Teamwork. Proceedings of the ACM on Human-Computer Interaction, 5(CSCW2), 1-20.

Shrivastava, A., & Pandey, A. (2022). Understanding intra-group response biases in culturally diverse teams. *International Journal of Intercultural Relations*, *88*, 189-199.

Snyder, D., Garcia-Romero, D., Sell, G., Povey, D., & Khudanpur, S. (2018, April). X-vectors: Robust DNN embeddings for speaker recognition. In *2018 IEEE international conference on acoustics, speech, and signal processing* (ICASSP) (pp. 5329-5333). IEEE.

Standaert, W., & Thunus, S. (2022). Virtual Meetings during the Pandemic: Boon or Bane for Gender Inequality.

Staniscuaski, F., Kmetzsch, L., Soletti, R. C., Reichert, F., Zandonà, E., Ludwig, Z. M., ... & de Oliveira, L. (2021). Gender, race, and parenthood impact academic productivity during the COVID-19 pandemic: from survey to action. *Frontiers in Psychology*, 12, 663252.

Swathi Jagannath, Aleksandra Sarcevic, and Ivan Marsic. (2018). An Analysis of Speech as a Modality for Activity Recognition during Complex Medical Teamwork. In Proceedings of the 12th EAI International Conference on Pervasive Computing Technologies for

Healthcare (PervasiveHealth '18). Association for Computing Machinery, New York, NY, USA, 88–97.

Tomprou, M., Kim, Y. J., Chikersal, P., Woolley, A. W., & Dabbish, L. A. (2021). Speaking out of turn: How video conferencing reduces vocal synchrony and collective intelligence. *PLoS One*, 16(3), e0247655.

Tripathi, P., & Burleson, W. (2012, February). Predicting creativity in the wild: Experience sample and sociometric modeling of teams. In *Proceedings of the ACM 2012 conference on computer supported cooperative work* (pp. 1203-1212).

Valente, Thomas. 2012. Network Interventions. *Science* 337, 6090 (2012), 49–53.

Willermark, S., & Pareto, L. (2020). Unpacking the role of boundaries in computer-supported collaborative teaching. *Computer Supported Cooperative Work (CSCW)*, 29(6), 743-767.

Wu, F., and Teets, T. S. (2021). Effects of the Covid-19 pandemic on student engagement in a general chemistry course. *Journal of Chemical Education*, 98, 12: 3633-3642.

Wu, Y., Sun, Y., and Sundar, S. S. (2022, November). What Do You Get from Turning on Your Video? Effects of Videoconferencing Affordances on Remote Class Experience During COVID-19. In *Proceedings of the ACM on Human-Computer Interaction* 6, no. CSCW2.

Xue, F., Merrill, M., Housefield, J., & McNeil, T. (2021, October). Beyond isolation: Benefits and challenges as perceived by students throughout online collaboration during the covid-19 pandemic. In *Companion Publication of the 2021 Conference on Computer Supported Cooperative Work and Social Computing* (pp. 195-198).

Ye Bai, Jiangyan Yi, Jianhua Tao, Zhengkun Tian, Zhengqi Wen, and Shuai Zhang. (2021). Fast End-to-End Speech Recognition Via Non-Autoregressive Models and Cross-Modal

Knowledge Transferring From BERT. IEEE/ACM Trans. Audio, Speech and Lang. Proc. 29 (2021), 1897–1911.

Ye, T., & Robert Jr, L. P. (2017, February). Does collectivism inhibit individual creativity? The effects of collectivism and perceived diversity on individual creativity and satisfaction in virtual ideation teams. In Proceedings of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing (pp. 2344-2358).

Zelkowitz, M. (2010). Improving the Web. Elsevier.

Appendix

Appendix 1: Robustness check using alternative measures of participation

The measure *pdur1* is normalized by the entire sample, which transforms the distribution of *pdur* to resemble a standard normal distribution:

$$\boldsymbol{pdur_1} = \frac{\boldsymbol{pdur - mean(pdur)}}{\boldsymbol{sd(pdur)}} \qquad \textbf{(5)}$$

The measure *pdur2* considers that group size has an inevitable influence on the value of speak time percent. It thus subtracts the group mean from *pdur* (100/team size) and divides the normalized standard deviation from the team mean:

$$\mathrm{pdur}_2 = \frac{(\mathrm{pdur} - 100\ /\mathrm{team\ size})}{\mathrm{sd\ (pdur\ within\ team\ within\ meeting)}}. \qquad (6)$$

Table A1: Examine the gender and racial participation gap using alternative measures of participation

	Model 1 **pdur1**	Model 2 **pdur1**	Model 1 **pdur2**	Model 2 **pdur2**
New-modality=1	-0.134 (0.177)	-0.266 (0.247)	-0.104 (0.152)	-0.159 (0.238)
Female=1	-0.122 (0.102)	0.120 (0.202)	-0.0755 (0.0909)	0.128 (0.183)
Minority=1	-0.179^{+} (0.100)	-0.522** (0.190)	-0.200* (0.0962)	-0.424* (0.181)
Size	-0.183*** (0.0538)	-0.197*** (0.0557)	-0.0345 (0.0480)	-0.0423 (0.0508)
New-modality=1 * female=1		-0.319 (0.230)		-0.270 (0.207)
New-modality=1 * minority=1		0.452* (0.224)		0.295 (0.206)
Constant	1.267**	1.471**	0.395	0.499

	(0.436)	(0.490)	(0.384)	(0.458)
Observations	433	433	433	433
Adjusted R^2	0.041	0.049	0.004	0.007

Note: Coefficients from OLS are shown in each cell, and robust standard errors are shown in the parentheses. In some cases, conventional standard errors were larger. + $p < 0.1$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

Appendix 2: Estimation results when having more control variables

Table A2 displays the regression results when adding course materials (technical and social) and student level (undergraduate or graduate student) variables. We excluded them from the main specification for two reasons: firstly, their estimated coefficients are significantly less than their standard errors; secondly, adding them does not make noticeable changes to other coefficients.

Table A5: Regression Results on Objective Participation Behaviors when adding more control variables

	Model 1 **Percent of speaking time duration** (*pdur*)	Model 2 **Percent of speaking time duration** (*pdur*)	Model 1 **Percent of time giving information** (*givepdur*)	Model 2 **Percent of time giving information** (*givepdur*)	Model 1 **Percent of time asking for information** (*askpdur*)	Model 2 **Percent of time asking for information** (*askpdur*)
New-modality=1	-0.436 (5.619)	-2.775 (6.139)	-0.392 (4.634)	-1.921 (5.006)	0.515 (1.074)	0.376 (1.029)
Female=1	-2.956 (2.505)	2.566 (4.207)	-2.408 (2.118)	2.161 (3.456)	-0.299 (0.526)	0.826 (0.776)
Minority=1	-3.767^{+} (2.077)	-9.966^{**} (3.757)	-2.702 (1.749)	-7.451^{*} (3.090)	-0.954^{+} (0.505)	-1.902^{**} (0.655)
Size	-2.662 (2.013)	-3.307^{+} (1.921)	-2.239 (1.670)	-2.716^{+} (1.593)	-0.256 (0.398)	-0.340 (0.364)

Technical	1.842 (4.534)	-1.514 (4.333)	1.217 (3.840)	-1.499 (3.680)	0.815 (0.906)	0.181 (0.924)
Social	3.119 (6.995)	0.852 (6.501)	2.070 (5.891)	0.314 (5.459)	1.215 (1.340)	0.852 (1.273)
Undergraduate	-1.272 (2.738)	-2.084 (2.725)	-1.357 (2.334)	-2.015 (2.344)	0.0392 (0.532)	-0.115 (0.512)
(new-modality=1) * (female=1)		-8.775^{+} (5.110)		-7.264^{+} (4.275)		-1.791^{+} (1.065)
(new-modality=1) * (minority=1)		9.003^{*} (4.341)		6.935^{+} (3.607)		1.407^{+} (0.844)
Constant	34.62 (21.66)	44.32^{*} (20.40)	29.29 (17.94)	36.56^{*} (16.78)	3.236 (4.220)	4.586 (3.839)
Observations	433	433	433	433	433	433
Adjusted R2	0.037	0.045	0.033	0.040	0.016	0.020

Note: Coefficients from OLS are shown in each cell, and robust standard errors are shown in the parentheses. In some cases, conventional standard errors were larger. + $p < 0.1$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

Table 1: The University's course modality policies

Pre-change period	Spring 2020	January 2020 to May 2020 On March 11, 2020, the university switched all classes to online meetings.
	Summer 2020	May 2020 to August 2020 All classes were online.
New-modality period	Fall 2020	August 2020 to December 2020 All classes were online (announced before the beginning of Fall semester).
	Spring 2021	January 2021 to May 2021 Most classes were online, with some in-person classes (not in our sample).

Table 2: Summary of project teams

Semester	Team ID	Course type	Team size	number of videos	Average length (minutes)	Video and audio recording date range
2020 spring	A	Social	6	3	25	03/27/2020-04/24/2020
2020 spring	C	Social	8	2	33	04/16/2020-04/27/2020
2020 spring	D	Social	8	4	21	03/27/2020-04/24/2020
2020 spring	F	Technical	9	3	39	04/09/2020-04/20/2020
2020 summer	H	Hybrid	5	5	46	07/14/2020-08/01/2020
2020 fall	I	Hybrid	4	13	43	10/02/2020-11/27/2020
2020 fall	J	Social	5	3	42	11/03/2020-11/19/2020
2020 fall	K	Social	5	7	53	10/30/2020-11/24/2020
2020 fall	L	Social	5	3	44	11/03/2020-11/17/2020
2020 fall	M	Social	5	5	26	10/21/2020-11/20/2020
2020 fall	N	Social	4	3	29	11/17/2020-11/29/2020
2020 fall	O	Social	4	3	12	11/18/2020-12/04/2020
2020 fall	P	Technical	6	7	17	11/06/2020-12/11/2020
2020 fall	Q	Technical	5	2	37	10/23/2020-11/17/2020
2020 fall	R	Technical	6	7	22	11/03/2020-12/02/2020
2021 spring	S	Social	5	8	22	03/04/2021-04/07/2021
2021 spring	T	Social	5	3	23	03/10/2021-03/29/2021
Total	**17**		**101**	**81**	**32**	**03/27/2020-04/07/2021**

Table 3: The Confusion Matrix of Recurrent Neural Network

		Predicted Labels		
		G	A	O
True Labels	G	119	9	13
	A	2	51	0
	O	10	3	26

Table 4: Perceived Contributions to Teams

Notation	Description
Proficiency (adopted from Kaufman et al. 1999, Ohland et al. 2012 and Mentzer et al. 2017)	
*l*21	Completed his/her tasks with the expected quality
*l*22	Completed his/her tasks to achieve the overall project goals
*l*23	Collaborated with team members from other disciplines to achieve the project goals
*l*24	Provided information to team members from other disciplines when needed
Adaptivity (adopted from Raelin et al. 2011 and Mentzer et al. 2017)	
*l*31	Adapted well to changes in the project that altered how he/she completed his/her task
*l*32	Coped with unforeseen demands placed on him/her
*l*33	Understood well how to deal with changes in the project to achieve the project goals
*l*34	Coped with changes to the project that altered how members from different disciplines collaborated on project goals
Proactivity (adopted from Raelin et al. 2011 and Mentzer et al. 2017)	
*l*41	Contributed new, improved ways to develop his/her tasks
*l*42	Initiated changes to the ways in which his/her tasks were done that helped accomplish project goals
*l*43	Created innovative solutions to improve the project quality
*l*44	Developed alternative solutions to achieve the project goals ahead of time

Table 5: Explanation of the Variables and Summary Statistics

Item	Definition	Mean	Standard Deviations (SDs)	Min	Max
pdur	Percent of speaking time duration	18.71	20.06	0	99
dur	Minutes of speaking time duration	5.91	8.73	0	88
givepdur	Percent of time giving information	15.45	17.03	0	99
askpdur	Percent of time asking for information	2.86	4.60	0	33
new-modality	1 if the new-modality period 0 if the pre-change period	0.73	0.45	0	1
female	1 if female, 0 if male	0.50	0.50	0	1
Minority[1]	1 if ethnicity is Asian, African American, or Hispanic; and 0 if White	0.44	0.50	0	1
size	Number of the team members	5.64	1.41	4	9

Notes: The unit of observation is individual member * attended meetings. As we had 101 individual members, who attended, on average, 4.29 meetings, our sample consists of 433 such pairs (i.e., data points).

[1.] To deal with a small sample size for Hispanic and African American students [4 Hispanic students and 5 African American students, out of 95 students total], we used only two categories, namely white and minority students, where minority students contain all non-white students. While acknowledging the racial subgroup differences, we opted for the binary grouping to obtain an adequate level of power for the quantitative analysis.

Table 6: Regression Results on Objective Participation Behaviors

	Model 1 **Percent of speaking time duration** (*pdur*)	Model 2 **Percent of speaking time duration** (*pdur)*	Model 1 **Percent of time giving information** (*givepdur*)	Model 2 **Percent of time giving information** (*givepdur*)	Model 1 **Percent of time asking for information** (*askpdur*)	Model 2 **Percent of time asking for information** (*askpdur*)
New-modality=1	-2.681 (3.553)	-5.326 (4.957)	-1.970 (2.882)	-3.834 (4.161)	-0.249 (0.790)	-0.499 (0.889)
Female=1	-2.442 (2.040)	2.398 (4.045)	-1.859 (1.754)	2.108 (3.358)	-0.350 (0.474)	0.729 (0.717)
Minority=1	-3.591^{+} (2.014)	-10.47^{**} (3.804)	-2.598 (1.695)	-7.880^{*} (3.130)	-0.850^{+} (0.511)	-1.987^{**} (0.678)
Size	-3.663^{***} (1.079)	-3.954^{***} (1.117)	-2.970^{**} (0.918)	-3.185^{**} (0.966)	-0.572^{*} (0.235)	-0.611^{**} (0.224)
(New-modality=1) * (female=1)		-6.396 (4.610)		-5.248 (3.858)		-1.432 (0.924)
(New-modality=1) * (minority=1)		9.070^{*} (4.486)		6.962^{+} (3.766)		1.494^{+} (0.859)
Constant	44.12^{***} (8.735)	48.21^{***} (9.837)	35.71^{***} (7.317)	38.70^{***} (8.412)	6.821^{***} (1.959)	7.316^{***} (1.889)
Observations (n)	433	433	433	433	433	433
Adjusted R^2	0.041	0.049	0.037	0.043	0.021	0.025

Note: Coefficients from OLS are shown in each cell, and robust standard errors are shown in the parentheses. In some cases, conventional standard errors were larger. + $p < 0.1$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 7: Changes in Students' Perceived Contributions to Team

	Model 1 **Perceived contribution**	Model 2 **Perceived contribution**
New-modality=1	0.266 (0.182)	0.209 (0.158)
Female=1	0.0415 (0.0847)	-0.0454 (0.151)
Minority=1	**-0.292*** **(0.118)**	**-0.294**$^{+}$ **(0.176)**
Size	0.0382 (0.0496)	0.0400 (0.0512)
(New-modality=1) * (female=1)		0.137 (0.179)
(New-modality=1) * (minority=1)		0.000421 (0.239)
Constant	3.387*** (0.433)	3.405*** (0.407)
Observations	95	95
Adjusted R^2	0.115	0.100

Note: Coefficients from OLS are shown in each cell, and robust standard errors are shown in the parentheses. In some cases, conventional standard errors were larger. + $p < 0.1$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

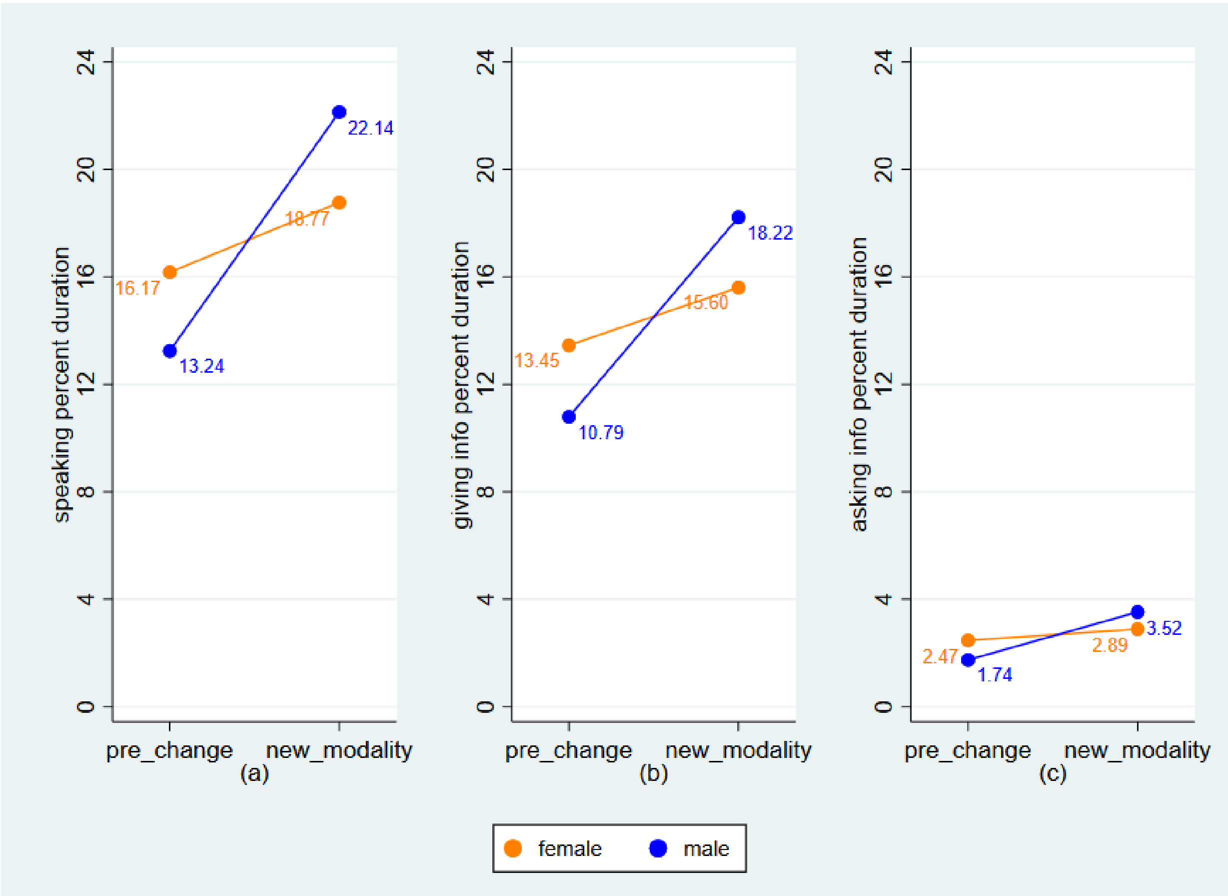

Figure 1: Participation of female and male students in pre-change and new-modality periods.

Notes: The values are the raw, unprocessed averages of speaking in general, giving, and asking for information.

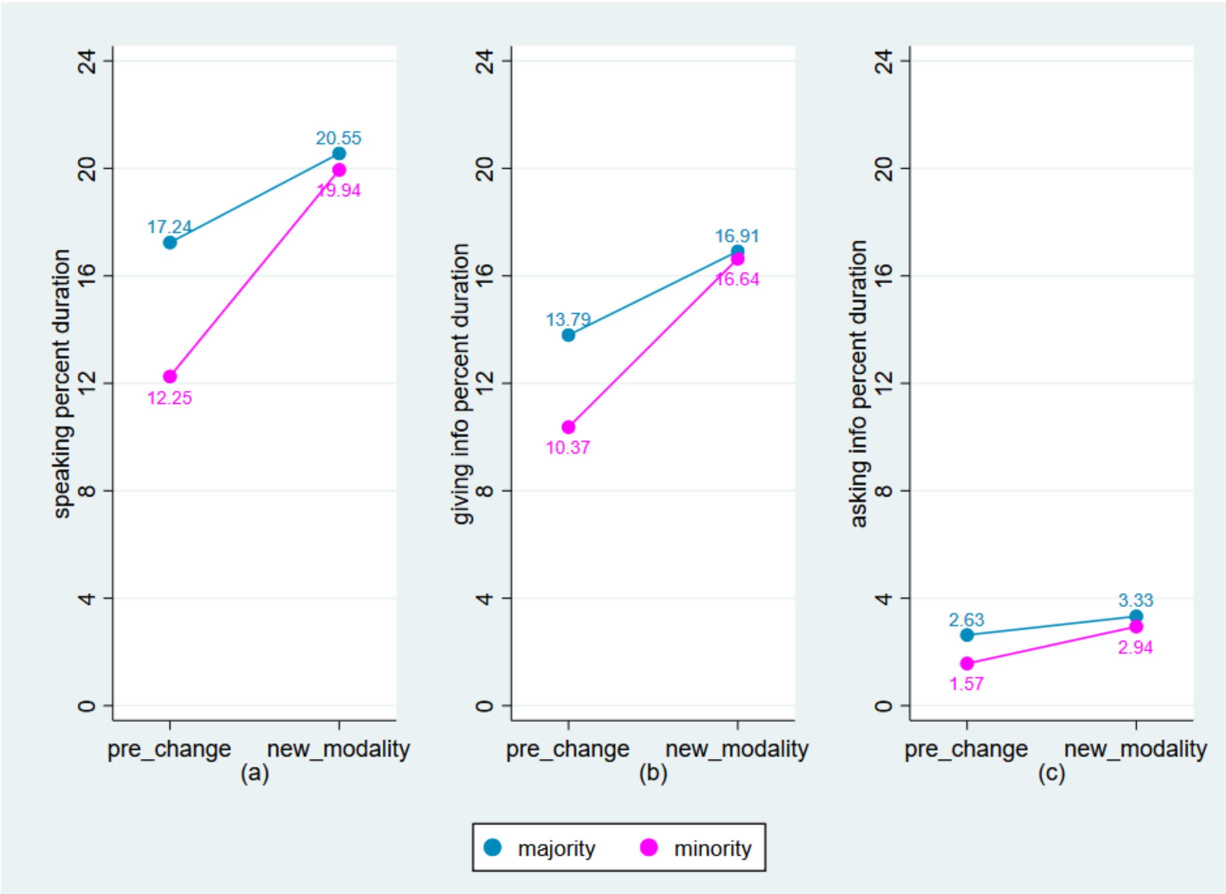

Figure 2: Participation of majority and minority students in pre-change and new-modality periods.

Notes: The values are the raw, unprocessed averages of speaking in general, giving, and asking for information.

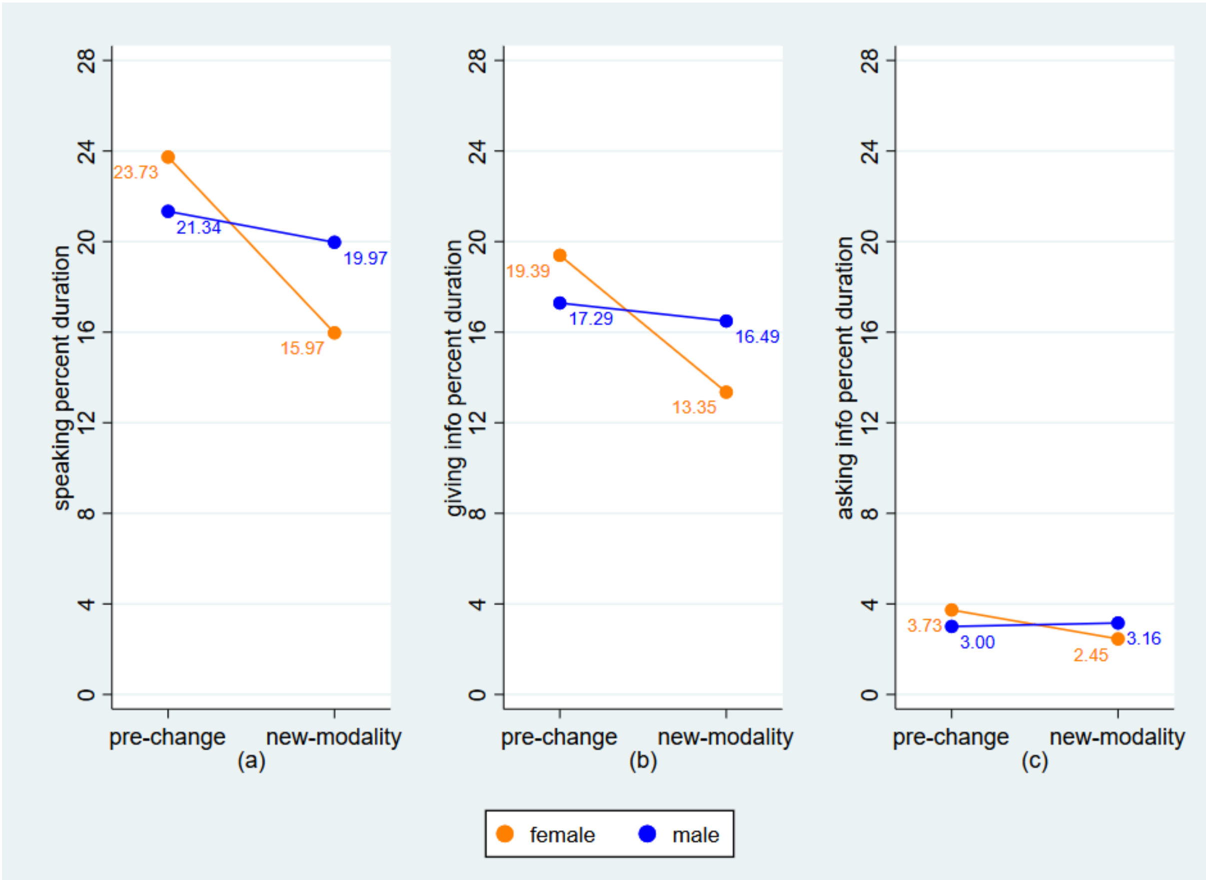

Figure 3: Gender gaps of participation in pre-change and new-modality periods.

Note: The three panels are percentages of speaking duration, giving information, and asking for information derived from the coefficients in **Table 6**, *not* raw time as in **Figures 1** and **2**.

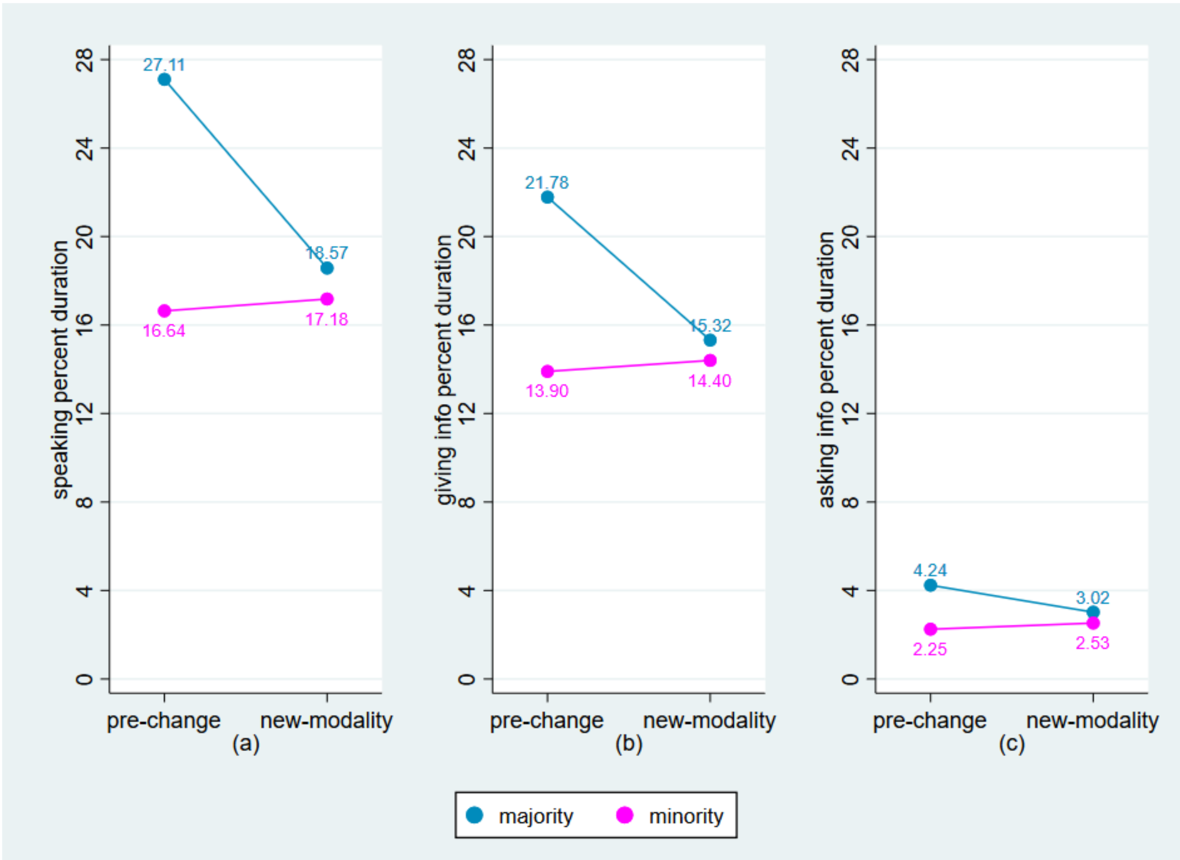

Figure 4: Racial gaps of participation in pre-change and new-modality periods.

Note: The three panels are percentages of speaking duration, giving information, and asking for information derived from the coefficients in **Table 6**, *not* raw time as in **Figures 1** and **2**.

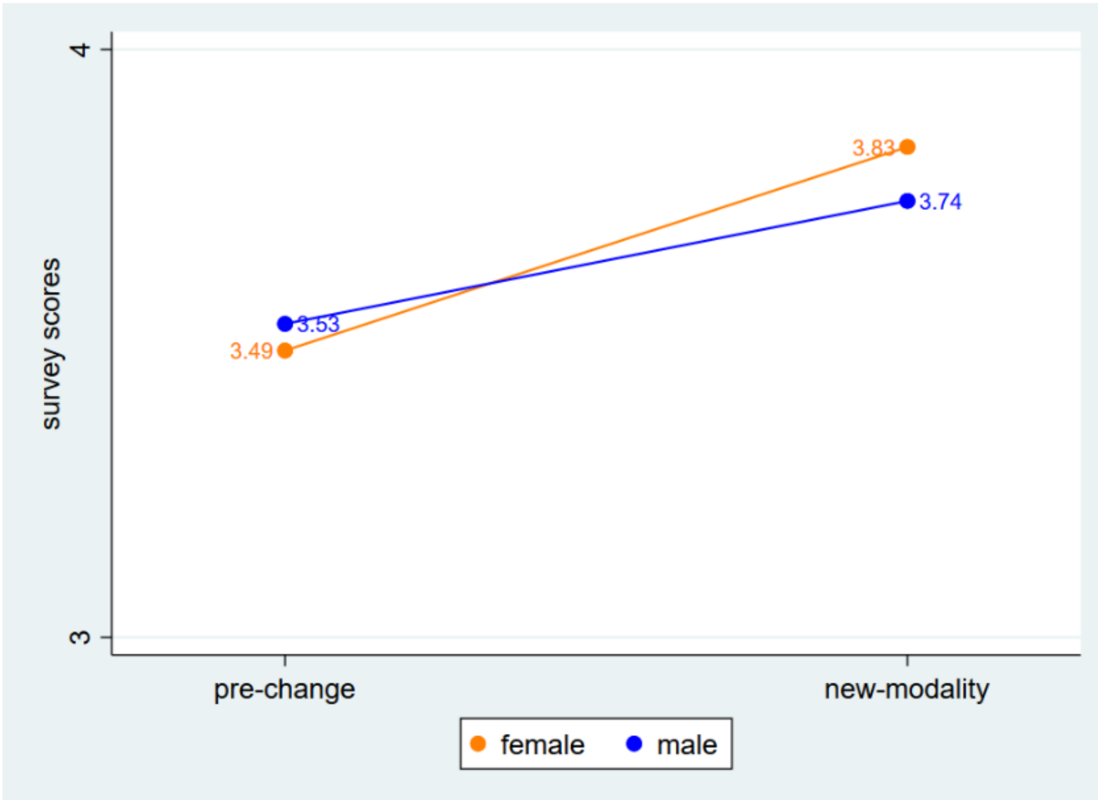

Figure 5: Perceived contribution of male and female in pre-change and new-modality periods.

Note: The three panels are percentages of perceived contribution derived from the coefficients in **Table 6**, *not* raw time as in **Figures 1** and **2**.

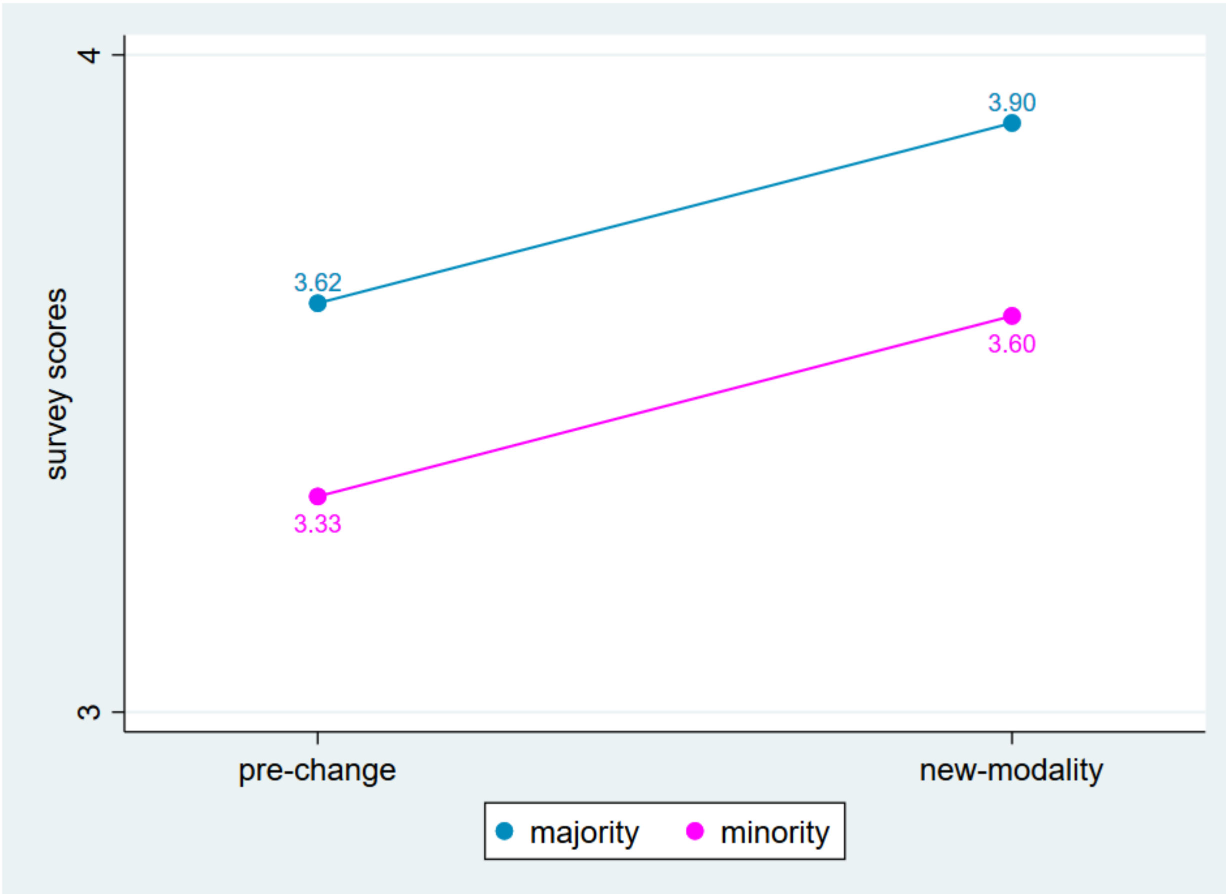

Figure 6: Perceived contribution of majority and minority students in pre-change and new-modality periods

Note: The three panels are percentages of perceived contribution derived from the coefficients in **Table 6**, *not* raw time as in **Figures 1** and **2**.

List of Figures