Feedback for Guiding Reflection on Teamwork Practices

Gilly Leshed, Jeffrey T. Hancock, Dan Cosley, Poppy L. McLeod, Geri Gay
Information Science and Communication
Cornell University
301 College Ave., Ithaca, NY 14850, USA
{gl87, jth34, drc44, plm29, gkg1}@cornell.edu

ABSTRACT
Effective communication in project teams is important, but not often taught. We explore how feedback might improve teamwork in a controlled experiment where groups interact through chat rooms. Collaborators who receive high feedback ratings use different language than poor collaborators (e.g. more words, fewer assents, and less affect-laden language). Further, feedback affects language use. This suggests that a system could use linguistic analysis to automatically provide and visualize feedback to teach teamwork. To this end, we present GroupMeter, a system that applies principles discovered in the experiment to provide feedback both from peers and from automated linguistic analysis.

Categories and Subject Descriptors
H.5.3 [**Information Interfaces and Presentation**]: Group and Organization Interfaces – *Computer-supported cooperative work.*

General Terms
Measurement, Design, Experimentation.

Keywords
CSCL, feedback, peer evaluation, linguistic analysis, teamwork.

1. INTRODUCTION
Many college courses now involve group work. Practicing teamwork in college prepares students to tackle problems upon entering the workforce, where teamwork is common. Further, the importance of social interaction for effective learning processes has long been recognized. Groups promote critical thinking skills, involve students in the learning process, improve classroom results, and model effective student problem solving techniques [9].

However, groups as learning tools also have pitfalls, such as *free riding*, where some members don't pull their weight [9]. These problems can sabotage group activity and learning goals; students hate free riders who can sometimes achieve a passing grade while doing little work and learning little. More generally, a prerequisite to successful learning in teams is the construction of effective collaboration practices [2]. However, little attention or guidance is given to learning teamwork processes in college courses.

Our goal is to find ways in which technology can be used to illuminate the group process so that students learn collaboration skills along with course content. In this paper, we present a feasibility study to investigate how automated linguistic analysis of conversation style might be used in a system that provides feedback about collaboration behavior. We asked groups to perform a decision-making task using a chat interface, and to provide peer feedback on each others' collaborative performance using a web interface. By collecting conversation data and explicit ratings of collaboration performance, we were able to address two key questions that a system that provides linguistic feedback must answer:

- Does conversation style predict collaboration ratings?
- Do collaboration ratings then change conversation style?

We found evidence for both questions, suggesting that building a system that provides linguistic feedback is a worthwhile thing to do. We end by presenting a prototype of the GroupMeter system which we will use to conduct further studies of how feedback from both peers and algorithms can improve group learning.

2. FEEDBACK TO GUIDE REFLECTION
Anyone who has worked in groups could produce a reasonable model of what effective collaboration entails, but that model would be generic and abstract. Most people would agree that effective collaboration involves making sure that all group members have the opportunity to contribute their expertise, but what can group members actually do to achieve this? Should they set aside an equal amount air time in meetings for each member? Should a leader or facilitator be given the responsibility to poll all members? How do the kind of task being performed and its progress in time affect the ways in which people should contribute?

Guiding students to consider how their groups should respond to these types of questions will develop the skills to adapt their collaborative techniques according to the demands of the situation. In a learning team, students can be taught lists or models of effective collaboration behaviors, but they must learn to apply these behaviors concretely in a given situation. They must learn to judge, for example, when a critical comment would be constructive, whether a joke would be distracting or a welcome tension reliever, or when keeping silent is more of a contribution than talking. Feedback about conversational behavior, may increase awareness of the team process and as such serve as a tool for training collaborative skills in learning groups.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.
GROUP'07, November 4–7, 2007, Sanibel Island, Florida, USA.
Copyright 2007 ACM 978-1-59593-845-9/07/0011...$5.00.

The SYMLOG framework (SYstematic Multiple Level Observation of Groups) [1] is commonly used to analyze group dynamics on three dimensions: dominance/submissiveness (also labeled *participation level*), individual/group orientation (*friendliness*), and task focus/socioemotional expressiveness (*task focus*). SYMLOG proposes that members of a group influence and are influenced by their behaviors on these dimensions. Using SYMLOG to provide team feedback in computer-mediated settings was shown to increase socioemotional behaviors, suggesting that feedback increases self-focus mediated by collaborative technology [7].

We used SYMLOG to collect peer feedback, asking group members to rate each other along these three dimensions. Peer evaluations are often used to provide feedback because unlike teachers, peers observe each other throughout the process and often have detailed knowledge of each others' contributions. Further, [3] found peer feedback to be superior in the quality of its results to expert-based feedback in learning environments. Peer feedback has been shown to motivate teams to alter their communication style, such as focusing more on the task at hand and using fewer words that were correlated with disliked collaborative moves [10].

However, peer feedback can be costly. Some students may be reluctant to give or receive feedback from other students, while attending to feedback during a task might be distracting. Automatic analysis of behavior is another approach for measuring collaborative practices. Latent Semantic Analysis (LSA) on team discourses was shown to be valuable in predicting team performance and could be used for providing real-time feedback [5]. Further, visualizing turn-taking patterns and participation levels based on audio input stimulates reflection [6], causing dominant contributors to become aware of their behavior [4].

Most existing research on automatic analysis of conversation focuses on high-level phenomena such as turn-taking and performance. In contrast, we propose using linguistic analysis tools to uncover more subtle features of conversational style that correspond to collaborative behaviors. We use Linguistic Inquiry and Word Count (LIWC) [8], a dictionary-based tool for analyzing language features. LIWC counts what percentage of words in a block of text occurs in various content categories such as emotion words, self-references, and assents. These LIWC indicators can be seen as measures of conversation style. For example, one could use LIWC to characterize people as self-directed versus other-directed by comparing how often they use self-references versus second- and third-person pronouns. Note that we envision LIWC-based analyses as a simple first step in linguistic analysis. Much more complex analyses of language (e.g., syntactic and pragmatic levels) can be implemented in the future.

3. FEASIBILITY STUDY

To guide reflection on group processes, a system that uses feedback on conversation style must at a minimum be able to identify linguistic features that correspond to effective teamwork and provide useful feedback that improves conversation style. Figure 1 presents an overview of an experimental feasibility study that focuses on these two questions:

RQ1. What elements of communication style predict peer evaluations on the SYMLOG dimensions?

RQ2. How does feedback affect subsequent collaboration practices, as indicated by communication style?

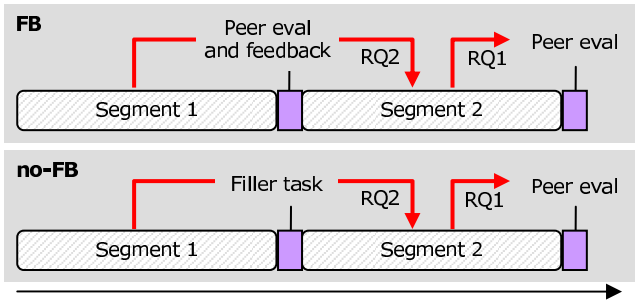

Figure 1. An overview of the experiment.

Participants. One-hundred and four undergraduate students (62 females, 42 males) in a large northeastern university volunteered to participate in the experiment for course credit and for a chance to win $40. Participants were randomly assigned to mixed-gender 4-person groups in one of two conditions: *Feedback (FB*, 13 groups*)* or *No Feedback (no-FB*, 13 groups*)*.

Procedure. Upon arrival, participants were seated at isolated computer workstations. Participants were informed that they would be working as a team on a decision making task and that their part is very important for the team's success. The instructions included a description of the three SYMLOG dimensions and behavioral propositions associated with effective teamwork on these dimensions.

Groups used an iChat chatroom to complete the group task. In the task, *Lost-on-the-Moon*, teams need to reach a decision with respect to the ranking of 15 items necessary for the survival as a team of astronauts on the moon. We chose not to reveal group members' actual names (to reduce potential gender biases); members were identified by a color: Blue, Red, Green, or Yellow.

An experimenter monitored the group's progress. When the group completed ranking seven items out of 15, they were prompted to pause the conversation and fill out an evaluation questionnaire. FB groups completed peer evaluations that consisted of three 7-point scales corresponding to the three SYMLOG dimensions along with open-ended responses to explain each rating they provided. no-FB groups evaluated the chat user interface as a filler task. Once all group members completed the questionnaire, they continued the group task. The experimenter tallied the ratings and sent FB groups a summary. The summary was an image with three bar graphs showing the group members' average ratings and reiterated the behavioral propositions from the initial instructions.

Upon completion of the task, all participants filled out post-session peer evaluations identical to those completed by the FB groups earlier. Participants were then debriefed and thanked for their participation. The whole session lasted about 50 minutes.

4. RESULTS

We divided each transcript into two segments, as shown in Figure 1. Segment 1 and segment 2 correspond to the conversation before and after the intervention, respectively. We used LIWC to analyze language use in each segment both at the individual and at the group level. For individuals, we extracted their contribution to the conversation and used that as input to LIWC, generating a list of linguistic features and the percentage of the text that corresponds to each feature. For groups, we submitted the entire transcript of the conversation to LIWC, less communication to and

from the experimenter. The group measurements are effectively an average of each individual's conversational style, weighted by the amount each individual contributed overall to the group.

4.1 Communication style predicts feedback

Our first research question was "What elements of communication style predict peer evaluations on the SYMLOG dimensions?" To address this question, we examined the relationship between individuals' LIWC results in segment 2 and peer evaluations provided at the end of the session. We created a number of hierarchical linear models using data from all participants, with individual nested within group, and group nested within feedback condition, to control for non-independence within groups. Each model used one LIWC indicator to predict the rating of one SYMLOG dimension—that is, does this indicator of language use correspond to higher or lower ratings? Because this is an exploratory study of the effect of language use, we examined 25 LIWC indicators. Table 1 presents a summary of the linguistic factors that were significant predictors of peer ratings.

Participants' word production was positively related to peer evaluations of group participation and task focus. Using achievement words (e.g., *best*, *goal*) and inclusive words (e.g., *also*, *plus*) was also positively related to peer evaluations of participation and task focus. Conversely, using affect-laden words, particularly related to positive emotion, was negatively related to participation and task focus evaluations. The use of assents (e.g., *yes, agree*) was also negatively related to all three evaluation dimensions.

The open-ended responses in which group members explained their ratings provide insight into why team members with various language styles were rated as better or worse collaborators. Those who received high ratings on participation were evaluated as "active", "leader", and "keeps us on track", while those with lower ratings were evaluated as "didn't talk much", "passive", and "only responds when necessary". Those with high ratings on task focus were evaluated as "organizes things" and "helped the group come to decisions", while those with low ratings were evaluated as "humorous" and "makes jokes". Finally, low evaluations were also followed by comments such as "seemed just to go with the flow and agree with what everybody else was saying" and "didn't seem to have many insights".

The open-ended responses help explain the positive relationship between peer evaluations and word count, achievement words, and inclusive words: 1) people may interpret verbosity as contribution, 2) being inclusive may be interpreted as suggestive and detailed, and 3) discussing achievement may be interpreted as being interested in the team's success. Alternatively, frequently using agreement terms may have been perceived as passivity (i.e. passively agreeing without active contribution). Using affect terms, especially positive emotion, may have been interpreted as straying off the task at hand.

4.2 Feedback affects communication style

Our second research question was "How does feedback affect subsequent collaboration practices, as indicated by communication style?" To answer this question, we examined whether the feedback intervention resulted in linguistic style change from segment 1 to segment 2 (RQ2, see Figure 1), comparing changes in LIWC results at the group level between the two conditions.

Table 1. Estimates of parameter coefficients in hierarchical linear models using linguistic measures as covariates and peer ratings on SYMLOG dimensions as predicted variables. Each value represents the coefficient parameter estimate on a single predictor model. (Significance: *p<0.1, **p<0.05, ***p<0.001).

		SYMLOG peer evaluation ratings		
		Participation	**Friendliness**	**Task focus**
Linguistic measures	Mean (SD)	5.29 (0.89)	5.80 (0.89)	5.49 (0.85)
Word count	137.6 (91.7)	0.005***	0.0002	0.002*
Achievement *(best, solve, win)*	1.32 (1.16)	0.151**	0.111*	0.148**
Inclusive *(also, and, plus)*	2.90 (1.71)	0.157***	0.061	0.086**
Assents *(ok, yes, agree)*	6.64 (5.91)	-0.063***	-0.028**	-0.036***
Affect *(funny, hate, good)*	7.12 (4.85)	-0.061***	-0.019	-0.036**
Positive emotion *(love, lucky, neat)*	6.54 (4.91)	-0.061***	-0.020	-0.036**

Figure 2 shows three LIWC indicators where the FB and no-FB conditions differed. While FB groups did not differ from no-FB groups in segment 1 on the use of total first person pronouns (*I*, *me*, *we*, *us*), cognitive process terms (*cause*, *know*, *think*, *should*), and assents (*ok*, *agree*, *yes*), in segment 2 FB groups use significantly more first person pronouns, more cognitive process terms, and less assents. As demonstrated in Figure 2, these differences between FB and no-FB groups resulted from the FB groups *not* changing their communicative behavior from segment 1 to segment 2, whereas the no-FB groups decreased their self pronoun use and cognitive process terms, and increased the use of assents.

We argue that the pattern of change observed in the no-FB condition reflects the natural communication course for this task without a feedback intervention. Toward the end of the session team members sought consensus, leading to fewer self pronouns and cognitive process terms and more assents (less "I think" and more "yeah, yeah, yeah"). It appears that the feedback intervention helped groups stay on track and keep the teamwork going with the same cognitive effort and involvement as in the first part of the conversation. This finding is also aligned with [7], suggesting that feedback increases self-focus compared to no feedback in a technology-mediated environment. Note that the no-FB groups talked just as much as the FB groups in segment 2 (FB: mean=506.9, SD=385.4; no-FB: mean=558.1, SD=269.1; t(24)=-.393, p=.698). This shows that conversation style can change while word count remains the same and supports the idea of using more sophisticated language analysis tools to understand group dynamics.

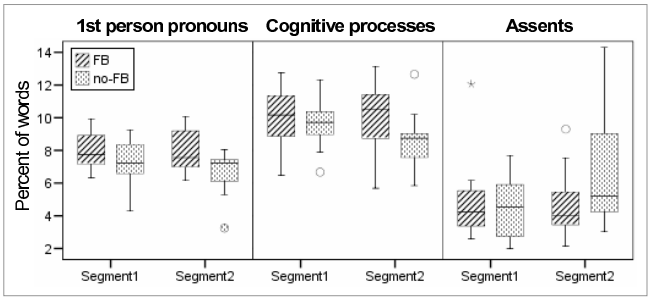

Figure 2. Use of first person pronouns, cognitive process terms, and assents in segments 1 and 2.

5. GROUPMETER

Our results are encouraging for the idea of developing systems that help people reflect on group processes. However, a number of questions remain:

- Can feedback affect task performance? In the study, it did not (FB: mean=38.5, SD=11.3; no-FB: mean=34.9, SD=8.3; t(24)=.907, p=.373). The short time frame (approx. 30 mins) may have been a factor; we think reflection on collaboration is more likely to have consequences in the long-run.
- Will giving and attending to feedback be distracting? Based on survey data from the study, it was somewhat distracting (mean=3.92, SD=1.84 on a 7-point scale, 1= not distracting at all, 7= very distracting). Designs will need to balance the goals of guided reflection and learning outcomes.
- What are effective ways to distill linguistic feedback into information people can use? The results of RQ1 suggest that practically we can replace expensive expert evaluations with LIWC indicators. However, presenting raw linguistic data ("you say yes a lot") will have different effects from presenting suggested behaviors ("contribute more ideas").
- How should feedback be visualized? Should it emphasize individual or group-level behaviors? Should the visualization include normative elements, or simply present information for interpretation? Does it matter whether people believe the feedback comes from peers or a program?

Inspired by the feedback concepts and questions laid out above and the results from this feasibility study, we developed GroupMeter, a system designed to help groups reflect on their collaboration practices and support research into how language, feedback, and practices interact. A server runs under Apache Tomcat, supporting chat sessions, collecting conversational data and peer ratings, performing linguistic analysis on collected conversations, aggregating peer ratings, and providing access to the collected information about ratings and language behavior. Figure 3 presents the GroupMeter interface. The user interface runs in a web browser and presents a synchronous chat client (main portion), augmented with features for collecting peer ratings (right) and presenting visualizations of group behavior based on the peer ratings and linguistic analyses (bottom).

We chose synchronous chat partly because it resembles instant messaging, a tool commonly used by college students, and partly because it facilitates the collection and analysis of linguistic data. We plan to leverage what we learn through this initial version of GroupMeter as an educational tool to support group reflection in an array of collaboration media used in team learning environments, including asynchronous interaction in blogs and wiki discussions; and using voice recognition to support group reflection in face-to-face meetings, audio- and video-conferencing.

6. CONCLUSIONS AND FUTURE WORK

In the experiment we demonstrated peer feedback effects on communicative style and the potential of automated linguistic analysis to measure teamwork behaviors. We then incorporated linguistic analysis into the GroupMeter system, suggesting that it can guide reflection on collaborative practices in addition to peer feedback procedures. Our next steps are to run user studies to evaluate the extent to which GroupMeter fulfills this promise, and to develop a theoretical framework that elucidates the linkages we discovered between collaboration ratings and linguistic style.

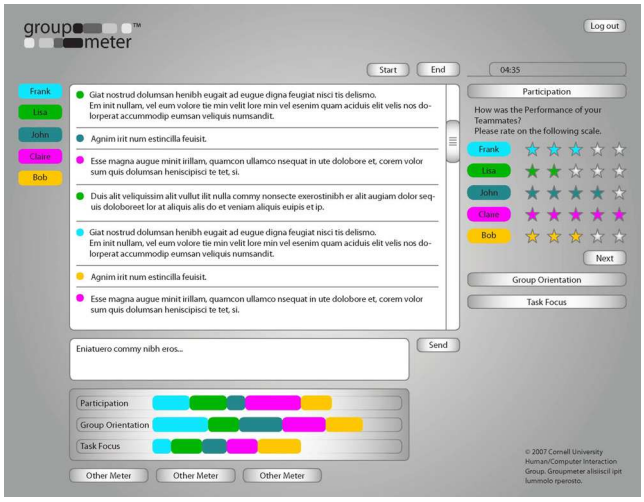

Figure 3. The GroupMeter user interface.

We plan to deploy GroupMeter in classroom environments, where students work on group projects throughout the semester. We will examine changes not only in collaborative practices, but also in the reflective process students undergo, project quality and learning outcomes. By this we hope to foster more educational, reflective, and enjoyable teamwork learning environments.

7. REFERENCES

[1] Bales, R.F., & Cohen, S.P. (1979). *SYMLOG: A system for the multiple level observation of groups*. NY: Free Press.

[2] Bransford, J. (2000). *How people learn: Brain, mind, experience, and school*. National Academy Press.

[3] Cho, K., Chung, R., King, W., & Schunn, C.D. (in press). Creating a peer-based computer-supported knowledge refinement process. *Communications of the ACM*.

[4] DiMicco, J.M., Pandolfo, A., & Bender, W. (2004). Influencing group participation with a shared display. *Proc. CSCW'04*. Chicago, Il. 614-623.

[5] Foltz, P.W., & Martin, M.J. (2004). Automated team discourse annotation and performance prediction using LSA. *Proc. HLT/NAACL*, Boston, MA.

[6] Karahalios, K.G., & Bergstrom, T. (2006). Visualizing audio in group table conversation. *Proc. IEEE TableTop 2006*.

[7] Losada M., Sánchez, P., Noble, E.E. (1990). Collaborative technology and group process feedback: Their impact on interactive sequences in meetings. *Proc. CSCW'90*.

[8] Pennebaker, J.W., Francis, M.E., & Booth, R.J. (2001). *Linguistic Inquiry and Word Count LIWC 2001*. Erlbaum.

[9] Roberts, T.S. (2005). *Computer-Supported Collaborative Learning in Higher Education*. Idea Group Publishing.

[10] Turner, B., & Schober, M.F (2007). Feedback on collaborative skills in remote studio design. *HICSS 2007*, 44.