Monitoring Email to Indicate Project Team Performance and Mutual Attraction

Sean A. Munson
Human Centered Design & Engineering
DUB group, University of Washington
smunson@uw.edu

Karina Kervin, Lionel P. Robert Jr.
School of Information
University of Michigan
{kkervin,lprobert}@umich.edu

ABSTRACT
Many managers and mentors for project teams desire more efficient and more effective ways of monitoring and predicting the quality of social relationships and the performance of teams under their purview. A previous study [13] found that one form of linguistic mimicry, linguistic style matching, and some lexical features indicated team performance and mutual attraction in short-term, laboratory tasks. In this paper, we evaluate whether these measures also work as indicators for performance, shared understanding, and team trust in longer-duration project teams, using only limited, unobtrusively obtained communication traces. In our four-month evaluation using student project team emails, we found no support for LSM or most of the previously identified measures as practical indicators in our field setting. We did find some support for using future-oriented words to indicate team performance over time.

Author Keywords
Email, groups, workgroups, performance, LIWC, team, process, language, mimicry, trust, mutual attraction, computer-mediated communication

ACM Classification Keywords
H.5.e. Group and Organization Interfaces: Computer-supported cooperative work

INTRODUCTION
Managers and project team leaders regularly seek more accurate or more efficient ways to keep track of the health and performance of the teams they manage [19, 20]. Team members also may benefit from better awareness of their team's performance [16]. Despite the importance of team performance and functioning in many workplaces, relatively little is known about effective and unobtrusive indicators of team performance and processes [4].

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. Copyrights for components of this work owned by others than the author(s) must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.
CSCW'14, February 15 - 19 2014, Baltimore, MD, USA
Copyright is held by the owner/author(s). Publication rights licensed to ACM. ACM 978-1-4503-2540-0/14/02...$15.00.
http://dx.doi.org/10.1145/2531602.2531628

In efforts to find unobtrusive indicators – with respect to additional work for team members and leaders, if not with respect to privacy – several researchers have mined face-to-face or electronic communication among different types of teams or workgroups. Measures extracted from this communication, such as affect, linguistic style, coordination routines, and communication volume (number of messages or word count), have been found to be valuable indicators in such diverse teams as the electronic communication of Wikipedians [9], face-to-face interactions among teams in simulated search and rescue missions [11], and student teams [7, 6]. While some of these indicators are easily automated, others have required extensive human coding of communication logs.

Linguistic mimicry between dyads or among small teams is one particularly promising indicator that can be easily automatically extracted from communication records. Linguistic mimicry is the extent to which people align in the cognitive complexity, formality, emotionality, and/or terms in their communication [5, 22]. In a laboratory setting, Gonzales et al. used one measure of linguistic mimicry, linguistic style matching (LSM) [21], as an indicator for group cohesion and, in computer-mediated settings, team performance [13].

In this paper, we present a study that replicates Gonzales et al.'s approach [13] in the field. We evaluate LSM as an indicator of mutual attraction, operationalized as trust and shared understanding, and performance in real-world project teams whose members communicate in-person, by email, and through other channels. We chose to use team emails as the source data, restricting ourselves to data that might be reasonably obtained in a corporate or educational environment, by mining the digital traces of communication among team members as they go about their work. We selected emails because groupware that integrates with activities and tools that people already use may have lower barriers to adoption and continued use [14]. Other than an initial setup, tracking team emails requires no further work from team members. Supporting this choice, [21] notes that emails are one form of communication to which the LSM metric should be applicable. As a secondary goal, we evaluated other linguistic indicators that Gonzales et al. [13] found indicated performance or mutual attraction. This included the use of first-person plural pronouns and word count to indicate group cohesion and future- and achievement-oriented language for performance.

This study, then, is an important step toward assessing the feasibility of using LSM and other linguistic measures as indcators of performance and mutual attraction among real-world

project teams. In our setting and with our collected data, we did not find that linguistic mimicry was an indicator of team trust, shared understanding, or team performance. Of the other indicators evaluated, we found only the proportion of future-oriented words to be indicative of performance and none to be indicators of team trust or shared understanding.

In the remainder of this paper, we present additional background on linguistic mimicry and the particular measure, Linguistic Style Matching (LSM), that [13] and we used, the details of our study design, our analysis, and details of our results. We then discuss the implications of these findings and conclude with ideas for future work.

RELATED WORK AND MOTIVATION

Previously, linguistic mimicry has been used as an indicator for team performance [13, 32], group cohesion [13], and trust [29]. Niederhoffer and Pennebaker propose one measure of similarity in text, Linguistic Style Matching (LSM), based on conversation participants' use of function ("content-free") words, such as adverbs, articles, auxiliary verbs, conjunctions, negations, prepositions, pronouns, and quantifiers [21]. Function words occur frequently and are context-independent, so similarity in the use of these words may be a better indicator of convergence in conversation style than similarity in use of words that are task-specific. For example, if project teams are working in different domains, their use of task-specific words may vary from project to project, but function words will be present across all communication. Similarly, function words will be present regardless of whether teams are communicating to socialize or to accomplish work [13], and so this measure can be used across messages in which participants are catching up about the weekend, planning a bonding happy hour, planning tasks for an upcoming deliverable, or sharing thoughts about the content of that deliverable.

In a short-duration lab experiment, Gonzales et al. found that groups' LSM scores were correlated with self-reported measures of group cohesion in both face-to-face (FTF) and computer mediated (CMC) contexts and are correlated with performance in the FTF, but not the CMC, context [13]. In the same study, Gonzales et al. also found a positive correlation between future-oriented words and performance, a negative correlation between achievement oriented words and performance, a positive correlation between word count and team cohesiveness, and a negative relationship between use of first-person plural pronouns and team cohesiveness. While this work is promising, we were curious if we could reproduce these results in a setting that is more authentic for many types of collaboration: longer-duration, real-world tasks in which participants communicate across diverse media, only some of which is unobtrusively observable.

We are not the first to study the application of the LSM measure to real-world settings or tasks. LSM, alongside other measurements, has been used to provide real-time feedback that has improved the performance of poorly functioning classroom teams [31]. In another context, higher LSM between hostage takers and hostage negotiators has been found to be correlated with successful outcomes [32].

In this study, we evaluated the potential of LSM as an indicator of team performance, team trust, and shared understanding in project teams whose members communicate in-person, by email, and through other channels. We further restricted ourselves to data that might be reasonably obtained in a corporate or educational environment, simply by mining the digital traces of communication among team members as they go about their work. Prior work has shown that mining online discussion patterns can indicate students' performance and retention [6]. Unlike [6], however, we study communication only among project group members and analyze the content of that communication, rather than the network of communication among an entire class of students. Our outcomes are also at the group, rather than individual, level.

We sought to replicate Gonzales et al.'s study of LSM [13] among teams working on longer-term projects with greater authenticity to the workplace. We also wanted to be able to distinguish between two forms of cohesiveness, or mutual attraction, among teams: shared understanding and team trust. Thus, we designed a study to test the following hypotheses about LSM's value as an indicator:

- **H1a.** Higher LSM will correlate to greater team trust.
- **H1b.** Higher LSM will correlate to greater team shared understanding.
- **H2.** Higher LSM will correlate to greater team performance on class assignments.

We also developed hypotheses based on the other linguistic indicators identified by Gonzales et al. [13], again dividing team cohesiveness into team trust and shared understanding:

- **H3a.** A lower proportion of first-person plural pronouns will correlate with greater team trust.
- **H3b.** A lower proportion of first-person plural pronouns will correlate with greater shared understanding.
- **H4a.** Greater word count overall will correlate with greater team trust.
- **H4b.** Greater word count overall will correlate with greater shared understanding.
- **H5.** A higher proportion of future-oriented words will correlate with higher performance.
- **H6.** A higher proportion of achievement-oriented words will correlate with lower performance.

If these hypotheses were to be supported – indicating that [13]'s measures work for longer-term communication routines in complex teams – these indicators could be useful for efforts to build dashboards that give managers, mentors, teams, and team members visibility of and feedback on their performance, processes, and communication styles (e.g., [15, 3]). Such dashboards have historically been more limited to providing feedback on specific, shorter-duration interactions, even when used among longer term project teams (e.g., [10, 18, 31]).

STUDY DESIGN

We chose to evaluate LSM's ability to indicate team performance, trust, and shared understanding among groups in an introductory masters-level course at the University of Michigan School of Information. In this course, students are assigned to groups of four to six students (in one team, three) to work on a semester-long project. Each project team is assigned to an external client who has an information-centric process. Each team is tasked with interviewing client stakeholders, identifying themes in the interview data, modeling the client's process, and providing recommendations to improve the client's process.

The selection of this site and context was advantageous for several reasons. First, each team worked on an authentic project, yet each project was bounded in time to one semester. Each project team earned several grades over the course of the semester, including a final report grade, which could be used as performance measures.

At the beginning of the Fall 2012 semester, we introduced the study to teams during a class session. Afterwards, they had the opportunity to review the informed consent materials and decide whether to participate. Each participating project team created a group mailing list and was instructed to add a unique email address, supplied by the research team, as a subscriber to this list. Email sent to that address would be accessed by our data analysis scripts. Students were asked to carbon-copy that address on any email communication they sent off-list, including communication between individuals on the team. Researchers and instructors informed the students that the instructors would never have access to the email sent to this address, and that copying this list on email between team members would earn students one point of extra credit in the course.

There were 44 project teams in this course. 30 project teams, including a total of 137 subjects, participated in this study. One of these teams did not continue automated email address in their communication after the second week, and so they are excluded from the data analysis and result. Additionally, in two further teams, only one or two team members appeared to reliably use the automated email address in their communication (with less than four emails from some team members), so we excluded those teams as well. This left 27 project teams and a total of 124 students (50 men and 74 women). Most teams were gender-balanced or approximately gender-balanced, but two teams were all-female, two teams were 75% female, and two teams were 25% female. From these 27 teams, we collected and analyzed 6993 emails (after excluding automated messages, messages from the team's clients, and messages from instructors), an average of 259 emails per team. We were unable to collect emails that team members sent to each other individually rather than by using the distribution list. Typically, email collection began late in the second week or early in the third week of the semester.

We were also concerned that there may be a self-selection confound in this study: perhaps deciding to participate in the study in a timely manner was indicative of a well-functioning team, which would mean that our data did not include LSM scores from low-performing teams. To check for this confound, we compared the course grades from participating teams to the grades for teams not participating and found no differences in either the mean or distribution of scores.

Measures

Our analysis was conducted at the team level, using demographics of the teams, the emails to the distribution list, a survey that included measures of trust and shared understanding, and their assignment grades. Participants completed surveys individually. We distributed the survey electronically after the end of the semester; students had received their final grades by the time they took the survey.

Linguistic Mimicry

In addition to our focus on real–world, 14-week semester–long projects, a notable difference from Gonzales et al.'s study is that they used transcripts of synchronous communication as an input in both the CMC and FTF conditions, while we use asynchronous communication (email) as the input. The emails were analyzed according to the LSM approach described in [13]. To summarize, Gonzales et al. input transcripts of interactions (in our case, emails) into the Linguistic Inquiry and Word Count (LIWC) language analysis tool [23], which is capable of outputting word counts for specific categories of words, including function words. LIWC is used to measure the frequency with which each group member uses nine types of function words: auxiliary verbs (e.g., can, has, am), articles (e.g., a, an, the), personal pronouns (e.g., her, I, we, they, you), indefinite pronouns (e.g., anyone, someone, others), prepositions (e.g., about, at, unless, till), negations (e.g., not, never, nor, nowhere, without), conjunctions (e.g. also, though, but, while), and quantifiers (e.g., all, besides, best, worst, some). For each of the nine categories c, the percentage of an individual n's total words ($p_{c,n}$) was calculated, as well as the percentage of the group's total words (p_{Gc}). This allows the calculation of an individual's similarity against the group, per word category, as

$$LSM_{c,n} = 1 - \frac{|p_{c,n} - p_{Gc}|}{p_{c,n} + p_{Gc}}$$

The group G's average for category c is the average of the individual scores:

$$LSM_{Gc} = \frac{\sum_{n \in G} LSM_{c,n}}{|G|}$$

And the LSM across the nine categories is simply:

$$LSM_G = \frac{\sum_{c=1}^{9} LSM_{Gc}}{9}$$

Moving from face-to-face or computer mediated chats to emails required deciding which text to input into this analysis process. We excluded automated emails sent from collaboration tools (e.g., Google Docs or the University's course management system), emails sent from the client organization, and emails sent from the instruction team. Within each email, we excluded quoted reply text and signature lines. Emails sent after the last assignment was due were excluded.

Mutual Attraction

To measure team cohesion, prior researchers [13, 31] have used the Interaction Rating Questionnaire (IRQ). The IRQ was developed for experiments studying linguistic style matching [21] in specific, temporally-bounded interactions (e.g., a chat). Because we were assessing interactions over the course of a semester, we did not feel the IRQ was the most appropriate measure for our study.

We instead looked to other measures of team cohesion or mutual attraction. We also wanted to be able to distinguish which components of mutual attraction LSM indicates. LSM is thought to be a useful indicator in part because mimicry represents improved shared understanding [12, 13]. Researchers have also commonly used team trust as a measure of mutual attraction [26, 27], and it has been found to be a primary factor in team cohesion [8]. Thus, we chose to include short measures for team trust and shared understanding in the survey. We selected:

- **Team Trust**: four items from Simons and Peterson [30].
- **Shared Understanding**: five items from Ko, Kirsch, and King [17].

We had also intended for the survey to include a three item measure of team cohesion, not differentiated as team trust and shared understanding. A survey configuration error prevented these questions from being displayed.

We also added disposition to trust as a control variable for use in models indicating trust. This was based on a six-item measure from Schoorman, Mayer, and Davis [28].

Task Performance

To measure task performance, we collected the team's grades on each of the four group assignments on which there was variance in grading (i.e., not the assignments that were simply credit/no credit). These assignments included a model of the process the team was studying, an affinity diagram and walkthrough, a final presentation, and a final report. All analyses were conducted without including the extra credit received for participation in this study.

RESULTS

In all of the following analyses, we controlled for group size and sex (measured as percentage of the team who was male), which were also control variables in [13]. Because grades and instruction were determined in part by the teaching assistant who taught the team's course section, we evaluated whether we should control for each team's assigned teaching assistant. After finding no differences in the distributions of grades between teaching assistants, we elected to drop this control variable out of concern that its inclusion (there were six teaching assistants and thus five dummy variables) would lead to overfitting of the models.

Linguistic Style Matching

We evaluated LSM as an indicator for both mutual attraction and team performance. Teams' LSMs varied from 0.822 to 0.94 ($\mu = 0.90$, median $= 0.90$).

H1. Mutual attraction

Based on Gonzales et al.'s findings, we expected to find that LSM was indicative of the two measures for mutual attraction we collected: shared understanding and team trust. We constructed OLS regression models with shared understanding and trust as dependent variables and our control variables and LSM as the independent variables. As one would expect, a greater disposition to trust indicated higher reported trust among teams ($\beta = 0.79$ ($p < 0.001$); $F = 3.854$ ($p < 0.05$) adjusted $R^2 = 0.26$), and so we also included disposition to trust as a control variable in our model for LSM based on team trust.

LSM was not a statistically significant indicator for either shared understanding or team trust, and observed effects were in the opposite direction of what was hypothesized. Thus, we did not find support for H1a or H1b.

H2. Performance

We next assessed whether a group's LSM score is indicative of their overall performance. We operationalized each group's performance as the overall score of all of the group's assignments ($p_{overall}$), weighted according to how course grades were determined. Participating teams' overall scores varied between 48.2 and 53.1 ($\mu = 50.8$, stdev$= 1.3$); the maximum possible was 55). We constructed OLS linear regression models for performance as a function of the LSM score, controlling for team size and sex. LSM was not a significant indicator of performance in either of these models. Furthermore, the observed effect of a 0.01 increase in LSM was -0.04 (95% confidence interval: -0.27 to 0.18).

Other Linguistic Indicators

We then tested the other linguistic indicators that Gonzales et al. [13] identified for cohesion and performance.

We first tested the indicators [13] identified for cohesion. We generated separate OLS models that indicated shared understanding and team trust based on word count, word count per person (since our teams varied in size), and proportion of first person plural pronouns. There did not appear to be effects of any meaningful size and none were statistically significant, and so we find no support for hypotheses H3a, H3b, H4a, or H4b.

We then tested the indicators that Gonzales et al. [13] identified for team performance. In their study, a higher proportion of achievement words negatively indicated performance and that a higher proportion of future-oriented words positively indicated performance.

Among the teams in our study, the proportion of future-oriented words was a positive indicator for the overall score (Table 1). A 1% increase in use of future-oriented words would correspond to a 1.6 point increase in overall score (95% confidence interval: 0.5 - 2.25 points). We also checked this effect on each of the assignments that comprise the overall score. It was a positive indicator for the modeling assignment (A1) and affinity diagram and walkthrough grades (A2). It was not an indicator for scores on the final report (A3)

Table 1. OLS models for team grades on different assignments as indicated by proportion of future-oriented words used.

	$\mathbf{S_{overall}}$			$\mathbf{S_{A1}}$			$\mathbf{S_{A2}}$			$\mathbf{S_{A3}}$			$\mathbf{S_{A4}}$		
	B	SE	p	B	SE	p	B	SE	p	B	SE	p			
Intercept	51.30	2.44		95.10	7.22		90.60	5.17		92.18	6.27		90.86	7.11	
Team size	-0.43	0.39	n.s.	-1.89	1.77	n.s.	-0.81	0.78	n.s.	0.19	1.02	n.s.	0.17	1.16	n.s.
Prop. Male	-3.38	1.26	n.s.	-8.25	3.72	0.037	-4.63	2.49	0.080	-6.23	3.23	0.07	-7.25	3.67	0.06
Prop. Future-oriented words	163.90	57.56	0.01	381.74	164.22	0.029	351.44	112.47	0.005	60.48	142.80	n.s.	186.09	161.88	n.s.
Adjusted R^2	0.36			0.27			0.54			0.03			0.05		
F-statistic on 3 & 23 dof	2.90		0.06	4.28		0.015	4.92		0.008	1.25		n.s.	1.46		n.s.

and presentation (A4). Additionally, a more conservative approach of applying a family-wise correction (and inflating the standard errors) would cause all confidence intervals to include zero, eliminating this effect. Thus, we find limited support for H5, that use of a higher proportion of future-oriented words indicates higher team performance, but use of this metric in this context merits further investigation.

In our data, the proportion of achievement words was not a indicator of performance, and thus we find no support for H6.

LIMITATIONS

Our context deviated in many ways from that tested by Gonzales et al. [13]. Their study used synchronous communication in a short-term (20 minute) task with answers that were verifiably correct or incorrect. All groups were same sex. Teams were 4-6 people. While our teams were similarly sized, they were mixed-sex. The project was a semester (14 weeks) in duration. The instruction team evaluated performance (according to rubrics and example assignments), and performance was subject to outside factors such as the specific problem and each team's clients' support and cooperation. We also were only able to capture a small portion (the email) of each team's communication. We chose this context as it was highly authentic and similar – apart from the performance evaluation – to how teams in many organizations function: similarly sized, working on similar but non-identical tasks, for different clients or stakeholders.

Because we deviated in many ways from the original experiment, it is not possible to determine which of these dimensions cause LSM to be a less effective indicator in this context. Future work may wish to systematically explore different dimensions – variance in specific project, task duration, whether the task is verifiably true or false or allows for more creativity, team size, team composition, communication synchronicity, and completeness of captured communication – and when LSM is or is not a indicator for team process and performance, either in the lab or in the wild. Finally, by having only measures for two aspects of mutual understanding or team cohesion – team trust and shared understanding – and not an overall measure for team cohension, it is possible that our study failed to measure some element of team cohesion for which LSM would have been an indicator.

We also chose student teams as a proxy for workplace teams. This allowed us to use grades as a standardized performance measure and offered considerable control for factors such as project duration and specific deliverables (since all had the same assignments and schedule). While our student teams had many similarities to many workplace teams – they worked with different, client organizations to make recommendations addressing clients' real problems – they are not the same as workplace teams. Our teams, for example, had their deliverables determined by an outside, third party (the instructors), and their primary goal was learning, with solving clients' problems as a secondary goal. This is a limitation of this study, but one that makes us still more skeptical that LSM and the other measures in [13] are appropriate for longer-term, real-world teams. If these measures are not good indicators even among such controlled teams and tasks, what potential do they have for more diverse projects?

DISCUSSION

Our work does not suggest that the combination of LSM and automatic email harvesting from project teams is an effective tool for monitoring teams' cohesion (H1a, H1b) and indicating performance (H2). We do find some support for using the proportion of future oriented words to indicate team performance. Further work may still be warranted to see whether our results hold in other settings, such as project teams in actual workplaces. Additionally, we believe that this approach should be explored for indicating the performance of open source software teams, for whom the majority of communication occurs electronically, and would be less subject to our inability to collect face-to-face or phone communication.

LSM

We found no support for using LSM to indicate mutual attraction or performance among student teams engaged in real-world projects over the duration of a semester. This suggests at least one important limit to the practical applications of LSM.

While our study had limited power and we cannot reject the idea that LSM measured in asynchronous communication correlates with team performance or mutual attraction in longer-term teams, we are skeptical that LSM is a practical measure for the type of data and teams that we examined. A study with more teams would give more statistical power, but even if it were to identify a correlation between LSM and performance, it is likely that this correlation is too small to be practically useful. Furthermore, in even less-controlled settings, it seems like any effect would be lost among other variables for which we cannot adequately control. While past work suggests that LSM is a valuable indicator in many types of interactions, the teams' work on projects in our study does not appear to be one of them. Future research may inform whether this limitation is because this context has too many uncontrolled factors (e.g., differences in clients being studied), email logs are too limited of a signal when teams are also communicating through other channels, a semester is too

long of a period for LSM to be applicable, or some other reason.

Exploration: LSM over time

Seeking explanations for our negative result, we were curious whether a semester was too long of a period for LSM to be valuable. Perhaps all teams eventually converge, and what matters is their rate of convergence, or how much the team's language converges. Thus, we explored whether mimicry increased over time, operationalized as whether teams' LSM scores were higher at the end of the semester than at the start. In a paired t-test, LSM scores were no higher in the second half of the semester than the first half, or in the last quarter of the semester than the first quarter. Observed differences were very small (0.005 and 0.008, respectively, and with little variance), so even if our test was under-powered, any actual change over time is likely negligible. This would be consistent with findings that style matching occurs very quickly (in a matter of seconds in face-to-face interactions) [22].

Had we observed changes in LSM over the semester, we would have been curious if the changes (rather than raw score) indicated our measures of mutual attraction or performance. As a further check, we analyzed the emails from the two weeks before each group assignment's due date to see if the LSM score from this more limited set of messages was an indicator of short-term performance: the grade on that particular assignment. These models were also not indicative of performance.

Future-oriented words

A higher proportion of future-oriented words used in team messages was indicative of higher team performance, as it was in [13]. This is consistent with other findings that future-oriented attitudes or scenarios can indicate performance (e.g. [25, 33]), but the ability to detect this in limited, text-based behavior traces is exciting. Though we do not fully understand its limits, the robustness of this result across two different studies and contexts is notable. With further refinement, this measure may be suitable for use in team dashboards or to provide visibility of potential performance issues to management.

There are, however, many questions that remain to be answered. How robust is this result? Why was it a indicator for early assignments but not later assignments? Our study does not offer an explanation, but earlier work encourages some speculation. The theory of achievement motivation [24, 25] predicts that the effects of future-oriented scenarios are greater when success at the current task is perceived to be an immediate prerequisite for future success. Our teams may have perceived success on the earlier assignments to be a prerequisite for success on the latter assignments - i.e., they would not be able to produce good quality final reports if they were not able to produce good process models (A1) or affinity diagrams (A2) - but not have perceived success on the final assignments as a prerequisite for any future success. This does not explain the indicative ability of future oriented words in [13], though, as their tasks had no dependencies. We hope that future research will address these questions.

Obtrusiveness

We selected email because it was practical to collect and fairly unobtrusive on teams' work practices. Beyond adding the collection email address to their distribution list and remembering to use that distribution list, this approach required no additional work from team members. Of the 30 teams that opted-in to the study, 27 appear to have remembered to use the distribution list in at least a substantial portion of their email communication, suggesting that our assumptions were reasonable.

Our study does not, however, address how intrusive such a monitoring system is from a privacy or autonomy perspective. While a production-system could further protect users' privacy by consuming emails as they are received – storing only word counts – rather than storing messages to be analyzed post-hoc, as they were in our study, team members may still resist such a system. There is a long history of workers perceiving performance monitoring as intrusive or diminishing their autonomy [2]. If dashboards are to be built to indicate team performance or process based on email or other communication traces, there remains a question of whether the project team members will be willing to allow their email to be monitored in this way, or whether such an application is "snoopware" [1] that people would choose to avoid. Future research should address this question before such systems are built and deployed.

CONCLUSION

This study contributes to the literature by evaluating indicators previously found to be valuable for indicating team performance and cohesion in the lab in a field setting, among real-world teams and over a three-month project. We were optimistic that LSM and the other indicators identified in [13] would prove to be useful indicators of team performance, shared understanding, or trust in the field. Had this turned out to be the case, it would have opened the door for management dashboards or other tools to monitor team performance and functioning based on these measures.

Instead, we find no support for using the majority of indicators identified by Gonzales et al. [13]. There are many possible reasons for this, some of which we discuss above. While this is disappointing from a practical standpoint, it helps to better understand in which situations and for what LSM and other metrics may or may not be useful indicators.

We also found that one measure – proportion of future-oriented words used – which indicated team performance in a short-term, lab setting may also indicate team performance in a semester-long educational setting, but not for all assignments. Due to the limitations of this study, these results should receive further scrutiny in future research. Future work should also explore conditions in between those we tested and those tested by Gonzales et al. to better understand the limits of these measures and why they do or do not work in different contexts.

ACKNOWLEDGEMENTS

We thank to the students and instructors who helped make this experiment possible. This study also benefited from feedback

from participants in the Social Computing seminar in Fall 2009, especially Paul Resnick and Eytan Bakshy. Anonymous reviewers' comments and suggestions substantially improved this manuscript.

REFERENCES

1. Allen, J. Groupware and social reality. *Computers Society 22*, 1-4 (1992), 24–28.
2. Ball, K., and Wilson, D. C. Power, control and computer-based performance monitoring: Repertoires, resistance and subjectivities. *Organization Studies 21*, 3 (2000), 539–565.
3. Biehl, J. T., Czerwinski, M., Smith, G., and Robertson, G. G. Fastdash: a visual dashboard for fostering awareness in software teams. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '07, ACM (New York, NY, USA, 2007), 1313–1322.
4. Brannick, M. T., and Prince, C. W. An overview of team performance measurement. In *Team Performance Assessment and Measurement: Theory, Methods, and Applications*, M. T. Brannick, E. Salas, and C. W. Prince, Eds. Lawrence Erlbaum, Mahwah, New Jersey, 1997.
5. Brennan, S. E., and Clark, H. H. Conceptual pacts and lexical choice in conversation. *Journal of Experimental Psychology: Learning, Memory, and Cognition 22*, 6 (1996), 1482.
6. Chen, G., Wang, C., and Ou, K. Using group communication to monitor web-based group learning. *Journal of Computer Assisted Learning 19*, 4 (2003), 401–415.
7. Chioccio, F. Project team performance: A study of electronic task and coordination communication. *Project Management Journal 38*, 1 (2007), 97–109.
8. Cohen, S. G., and Bailey, D. E. What makes teams work: Group effectiveness research from the shop floor to the executive suite. *Journal of Management 23*, 3 (1997), 239–290.
9. Danescu-Niculescu-Mizil, C., Lee, L., Pang, B., and Kleinberg, J. Echoes of power: language effects and power differences in social interaction. In *Proceedings of the 21st international conference on World Wide Web*, WWW '12, ACM (New York, NY, USA, 2012), 699–708.
10. DiMicco, J., and Bender, W. Group reactions to visual feedback tools. In *Persuasive Technology*, Y. Kort, W. IJsselsteijn, C. Midden, B. Eggen, and B. Fogg, Eds., vol. 4744 of *Lecture Notes in Computer Science*. Springer Berlin Heidelberg, 2007, 132–143.
11. Fischer, U., McDonnell, L., and Orasanu, J. Linguistic correlates of team performance: Toward a tool for monitoring team functioning during space missions. *Aviation, Space, and Environmental Medicine 78*, 1 (2007), 86–B95.
12. Giles, H., and Coupland, N. *Language: Contexts and consequences*. Thomson Brooks/Cole Publishing Co, 1991.
13. Gonzales, A. L., Hancock, J. T., and Pennebaker, J. W. Language style matching as a predictor of social dynamics in small groups. *Communication Research 37*, 1 (2009), 3–19.
14. Grudin, J. Groupware and social dynamics: Eight challenges for developers. *Communications of the ACM 37*, 1 (1994), 92–105.
15. Grudin, J., and Poole, E. S. Wikis at work: success factors and challenges for sustainability of enterprise wikis. In *Proceedings of the 6th International Symposium on Wikis and Open Collaboration*, WikiSym '10, ACM (New York, NY, USA, 2010), 5:1–5:8.
16. Hacker, M. E., and Lang, J. D. Designing a performance measurement system for a high technology virtual engineering team a case study. *Integrated Manufacturing Systems 2*, 3 (1999), 225–232.
17. Ko, D.-G., Kirsch, L. J., and King, W. R. Antecedents of knowledge transfer from consultants to clients in enterprise system implementations. *MIS Quarterly 29*, 1 (2005), pp. 59–85.
18. Leshed, G., Perez, D., Hancock, J. T., Cosley, D., Birnholtz, J., Lee, S., McLeod, P. L., and Gay, G. Visualizing real-time language-based feedback on teamwork behavior in computer-mediated groups. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '09, ACM (New York, NY, USA, 2009), 537–546.
19. MacBryde, J., and Mendibil, K. Designing performance measurement systems for teams: theory and practice. *Management Decision 41*, 8 (2003), 722–733.
20. Meyer, C. How the right measures help teams excel. *Harvard Business Review*, 5 (1994), 95–103.
21. Niederhoffer, K. G., and Pennebaker, J. W. Linguistic style matching in social interaction. *Journal of Language and Social Psychology 21* (2002), 337–360.
22. Pennebaker, J. W. *The Secret Life of Pronouns: What Our Words Say About Us*. Bloomsbury Press, 2013.
23. Pennebaker, J. W., Booth, R. J., and Francis, M. E. Linguistic inquiry and word count (LIWC): A computerized text analysis program, 2007.
24. Raynor, J. O. Future orientation and motivation of immediate activity: An elaboration of the theory of achievment motivation. *Psychological Review 76*, 6 (1969), 606–610.
25. Raynor, J. O., and Rubin, I. S. Effects of achievement motivation and future orientation on level of performance. *Journal of Personality and Social Psychology 17*, 1 (1971), 36–41.

26. Robert, L. P., Dennis, A. R., and Ahuja, M. Social capital and knowledge integration in digitally enabled teams. *Information Systems Research 19*, 3, 314–334.

27. Robert, L. P., Dennis, A. R., and Hung, C. Individual swift trust and knowledge-based trust in face to face and virtual team members. *Journal of Management Information Systems 26*, 2 (2009), 241–279.

28. Schoorman, F. D., Mayer, R. C., and Davis, J. H. Organizational trust: Philosophical perspectives and conceptual definitions. *The Academy of Management Review 21*, 2 (1996), 337–340.

29. Scissors, L. E., Gill, A. J., and Gergle, D. Linguistic mimicry and trust in text-based cmc. In *Proceedings of the 2008 ACM conference on Computer supported cooperative work*, CSCW '08, ACM (New York, NY, USA, 2008), 277–280.

30. Simons, T. L., and Peterson, R. S. Task conflict and relationship conflict in top management teams: The pivotal role of intragroup trust. *Journal of Applied Psychology 85*, 1 (Feb. 2000), 102–111.

31. Tausczik, Y. R., and Pennebaker, J. W. Improving teamwork using real-time language feedback. *CHI 2013* (2013), 459–468.

32. Taylor, P. J., and Thomas, S. Linguistic style matching and negotiation outcome. *Negotiation and Conflict Management Research 1*, 3 (2008), 263–281.

33. Weitzenkorn, S. D. An adjusted measure of achievement motivation for males and females and effects of future orientation on level of performance. *Journal of Research in Personality 8*, 4 (1974), 361–377.