Predicting Individual Performance in Student Project Teams

Matthew Hale  
University of Tulsa  
matt-hale@utulsa.edu

Noah Jorgenson  
University of Tulsa  
nnoah-jorgenson@utulsa.edu

Rose Gamble  
University of Tulsa  
gamble@utulsa.edu

Abstract

Due to the critical role of communication in project teams, capturing and analyzing developer design notes and conversations for use as performance predictors is becoming increasingly important as software development processes become more asynchronous. Current prediction methods require human Subject Matter Experts (SME) to laboriously examine and rank user content along various categories such as participation and the information they express. SEREBRO is an integrated courseware tool that captures social and development artifacts automatically and provides real-time rewards, in the form of badges and titles, indicating a user's progress towards predefined goals using a variety of automated assessment measures. The tool allows for instructor visualization, involvement, and feedback in the ongoing projects and provides avenues for the instructor to adapt or adjust project scope or individual role assignments based on past or current individual performance levels. This paper evaluates and compares the use of two automated SEREBRO measures with SME content-based analysis and work product grades as predictors of individual performance. Data is collected from undergraduate software engineering teams using SEREBRO, whose automated measures of content and contribution perform as well as SME ratings and grades to suggest individual performance can be predicted in real-time.

1. Introduction

Bringing current development methodologies into an undergraduate software engineering curriculum requires a framework that provides for rapid discussion, design, and documentation surrounding implementation. From the student team perspective, this means that they have a wide range of discussion opportunities with automated information capture, that design and documentation is facilitated, and that coding styles, such as individual, team-oriented, or “bursty” (determined by the team makeup, the milestone to be achieved, and competing work from other classes) are supported. From the instructor perspective, deliberate and quick performance assessment is needed throughout the process to guide and encourage teams to successful milestone achievements, while at the same time helping individuals manage their different skill sets, work habits, and project commitments. In this paper, we discuss our courseware, called SEREBRO, and the performance indicators used to predict the success of an individual on a software project team.

Communication is an important part of the group software development process [1-3]. Research into qualitative [4] and quantitative [5] communication metrics shows how communication can be effectively measured. Other research links these metrics with performance statistics related to a success factor, supporting their use as performance indicators [6, 7]. Knowledge management research [8] suggests that both explicit (know-what) and tacit (know-how) knowledge capture take place for better shared information among software teams. The use of incentives to capture knowledge [9, 10] mitigate some of the overhead placed on development teams to perform knowledge capture.

At the University of Tulsa, Computer Science majors take a full academic year of software engineering in consecutive courses – Software Projects I taught in the fall semester and
Software Projects II taught in the spring semester. Software Projects I immerses students in the software engineering process, basic software artifact construction, web application development, and dealing with team-related issues. Software Projects II allows the students to form stable teams for the semester, deepen their understanding of the software development process and associated artifacts, and concentrate on a non-trivial software project with a designated customer.

The class uses the SEREBRO courseware throughout project development, which employs a diverse set of indicators to examine participation in team discussion, responsibility taken for a certain team role, and tangible artifact generation. Because SEREBRO captures tacit knowledge implicitly in real time, it allows online metrics to be calculated dynamically as a team communicates. We show how two of these online metrics can be used as predictors of individual performance in a team software project by correlating them to offline metrics that include project grades and subject matter expert (SME) ratings that measure the project based content and contribution of individuals in team discourse.

2. The Courseware

SEREBRO was initially developed as part of a pilot study grant from NSF’s CreativeIT program to recognize creativity in software development teams within an undergraduate class [11]. The goal was to implement a Web application tool for capturing and relating ideas among team members to foster creative design processes and to reward those who contribute to the project’s creative elements. Each year we receive approval from our Internal Review Board to collect student data for analysis and publication. Though SEREBRO serves as the communication and process medium, students are informed of the research taking place and voluntarily sign consent forms to contribute personal information via surveys taken during the class. This paper represents our first assessment of individuals and not just the collective team use. In this section, we overview SEREBRO in its courseware form, as it is used by the Software Projects classes, and how it relates to our assessment of performance indicators.

The core facility for expression and team discussion is an idea network that combines aspects of idea management systems and social networking. Figure 1 shows a sample idea network in SEREBRO 3.0 [12]. Because SEREBRO is modeled as an online forum, asynchronous postings, along with email alerts of posts, give students freedom to work on the project from anywhere and at anytime. Figure 1 is a view inside a topic of a conversation among a team discussing web interface ideas for their Web application. Topics are often loosely based on the required artifacts designated for each project milestone. An idea can be a belief statement, a problem solving approach, a solution to a problem, or even discussion related to any previously expressed idea. Topic discussions begin when someone posts a brainstorm node (blue circle in Figure 1). A team member can agree (green triangle) with a post to continue the discussion with his/her own ideas, disagree (orange inverted triangle) and add a counter argument, or comment (talking bubble) with questions or neutral statements. Multiple brainstorms can be present within a single topic, started by any team member, and produce independent trees. A post appears when a user hovers over a node. Clicking on a node shows the post on the right of Figure 1 and allows the user to respond.

When users log in, the home page lists their development projects and the class discussion area. The class area allows questions about requirements, programming issues, and SEREBRO usage, to be answered directly by peers, the instructor, and the SEREBRO support team. All users have profiles where they can input images to be displayed with their posts and their contact information. The profile also contains current performance status based on SEREBRO’s online measures and is publically displayed in each post. In Figure 1, the status information about the person posting is shown on the left of the post.
Project milestones direct activity through a set of development tasks which are completed during a specific time period. The milestones mimic a generic iterative and incremental development process and the associated work products. The set of tabs grouped at the top left show where the user is within SEREBRO, e.g. home page (Home), project (Titanium Team), workspace (Forum), development milestone (Challenge 2), and topic (Website Ideas). The last tab, Post View, displays the idea network as a thread with indented child nodes. The tabs on the right hand side toggle posts as meeting minutes from face-to-face communication (Flashback Mode) and allow a search of the tags used by the team (Tag search).

As SEREBRO use continued, it became increasingly clear, through surveys taken and posts to the class discussion area, that students desired it to be courseware, with embedded project management facilities within SEREBRO to retain their conversations and link their communication content to artifacts. Therefore, as seen in the left-hand side menu in Figure 1, SEREBRO 3.0 now incorporates a Wiki for project documentation, file sharing (Uploaded Files), a Calendar (Schedule) for meeting scheduling, a custom Gantt chart (Tasks) and Subversion (SVN) repository for revision control of project source code.

### 3. Performance Measures

In this section, we describe two primary measures, SEREBRO Score and SEREBRO Events that SEREBRO calculates in real time, denoting information specific to our assessment results in this paper. Given that content based assessment is a clear metric of performance [6], we show the process by which Subject Matter Experts (SMEs) use SEREBRO to rate content and contribution of individuals. Grades are another standard performance metric used.

#### 3.1. SEREBRO Score

An initial SEREBRO Score was used in the early work to assess creativity [13]. Revised in Spring 2009, the current scoring algorithm better reflects conversation stemming from disagreement and its contribution to the quality of the discourse. We use this algorithm in Section 4 to assess its efficacy as a performance indicator. The Badges I and II (there are 5 total) accumulated by Burkmaster in Figure 1 are derived from his current SEREBRO Score, a real-time calculation for each user over the idea network using the following rules.
1. When a user posts a Brainstorm, SEREBRO allocates $p$ points to its author.
2. When a user Agrees with a post $X$, SEREBRO allocates $p\times k$ points to $X$’s author.
3. When a user Disagrees with a post $Y$, SEREBRO allocates $p/k$ points to $Y$’s author.
4. When $t$ points are received at a post $Z$, SEREBRO allocates $t/k$ points to the author of $Z$’s parent, in order to propagate the score to the progenitors of the thread.

The rules concentrate points on ideas that are well-received by the team and generate further discussion. Brainstorms immediately allocate points to their authors because of thread start. Agree and Disagree nodes do not directly provide points to their authors but instead allocate points to the author of their parent node. Subsequent ancestors receive a reduced point award equal to the child node’s allocated points $t$ divided by a factor $k$. This point distribution effectively propagates points over the network, while discounting them as the distance from the considered node increases. We use a value of 10 for $p$ and a value of 2 for $k$, which we experimentally found rewarded students for reasonable levels of contribution given course requirements and time constraints. Though students can manipulate posts in SEREBRO to achieve higher scores, they are aware that the instructor can manually discount posts with the archive button in the lower left of the post in Figure 1.

3.2. SEREBRO Events

The events logged by SEREBRO include contributions to the idea network, wiki, calendar, Gantt chart, uploads and SVN. Events are tagged by users to achieve relative connectivity between ideas and artifacts [14]. Figure 2 depicts the timeline event capture by SEREBRO to determine the SEREBRO Events per individual. Scrolling (not shown) allows movement through the timeline. Each event is hyperlinked to the actual event artifact or node. We examine SEREBRO Events as a performance measure in Section 4. Current work targets to the visual representation of the events to facilitate instructor feedback to the teams and also to understand how team work patterns vary during long and short milestone deadlines.

![Figure 2: SEREBRO Event timeline](image)

Other measures used by SEREBRO include Role status, a weighted score derived from work product activity related to a particular responsibility, and reputation, accumulated by star (upper right of post in Figure 1) and ‘like’ ratings (‘thumbs up’ in lower left of post in Figure 1). At this time, we do not assess these measures as performance indicators.

3.3. PM Idea Network Evaluation

In this section, we discuss the methods SMEs use to rate the participation and content of posts on an individual basis within a particular topic and the calculation of individual grades. The Project Management (PM) assessment tool under Staff Tools in Figure 1, provides SMEs with a visual interface to explore and rate each team member's contribution to a topic thread as seen in Figure 3.
We partially obfuscate user names, since post content is not altered. Red circles appear on nodes attributed to the selected team member. A message indicates when a team member does not contribute to the topic and no scoring is performed. Hovering over nodes brings up each post for context and quick assessment. Qualities are rated on a scale of Poor, Neutral, Good, Very Good, and Excellent relative to other team members for the topic. *Initiation* refers specifically to brainstorm contributions. *Participation* refers to the amount of contribution over all posts by the team member within the topic. *Flexibility* examines participation specific to agree/disagree/comment posts. *Discussion* assesses the depth of information contributed in agree/disagree posts. For instance, a collection of simple posts with messages such as “That looks great” would likely be rated to have less depth than posts which contain meaningful content like the post shown in Figure 1. *Content* judges the accumulated relevant information provided across all of the individual’s contribution to the rated topic. Hence, a higher content score means better overall content.

The SME raters are faculty and graduate students related to the research project. For the data presented we performed a variance assessment across all 50 individual rankings. Variance, $\sigma^2$, was calculated per user per dataset across all SMEs and averaged. Over all datasets, regardless of the tasks required or their duration, SMEs rated individual users similarly with a low degree of variance on all of the PM subscores, all $\sigma_{avg}^2 < 0.25$. The average variances for each PM subscore are shown in Table 1.

![Figure 3: PM view for rating content](image)

**Table 1: Average variances of SME ratings across datasets**

<table>
<thead>
<tr>
<th></th>
<th>Initiation</th>
<th>Participation</th>
<th>Frequency</th>
<th>Depth</th>
<th>Content</th>
<th>Average</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0.184</td>
<td>0.130</td>
<td>0.242</td>
<td>0.248</td>
<td>0.187</td>
<td>0.110</td>
</tr>
</tbody>
</table>

Team grades are based on the delivery of milestone requirements. Detailed rubrics per requirement give performance expectations. The individual grade is the product of the team grade and the *evaluation* of the individual by team members. Evaluations are based on a scale of 1-10 over factors such as ideas, effort, results, and resourcefulness as related to individual contributions.

**4. Experimental Environment**

SEREBRO use, experimentation, and data collection began in Fall 2008. In a typical fall semester Software Projects I class, students alternate among teams performing a series of
short (1 week) and long (2-3 week) challenges to reach a particular milestone for the same application requirements, which culminates with team presentations of results that are graded by the entire class across a set of predetermined rubrics. There is an average of 6 challenges per fall semester. In the spring semester Software Projects II class, the teams and projects are chosen for the duration of the semester with 4 Builds that have work products specific to the project. Builds are generally 4-6 weeks long and culminate in a progress presentation and product demonstration by all teams.

To analyze SEREBRO performance measures for their validity in predicting individual and team success, we compiled datasets from one short challenge in Fall 2009, one build in Spring 2010, and one short and one long challenge in Fall 2010. The length, tasks and project team members across the challenges and build varied to affect distinct project outcomes on performance with respect to different milestone requirements.

4.1. Datasets

The first dataset from Challenge 3 in Fall 2009 (Ch-3-09) lasted one week. Challenge 3 had 4 teams, each with 3 team members. Work products were a marketing slogan for their product, essential use cases, and a mashup within a prototype gardening application. The difficulty was moderate on a scale of hard, moderate, or easy. Challenge 1 in Fall 2010 (Ch-1-10) also had 4 teams, three with 3 members and one with 4 members. Its work products were vision, scope and feasibility documents for a web based gardening tool. The difficulty was easy and the project duration was nine days. Challenge 2 in Fall 2010 (Ch-2-10) also consisted of 3 teams with 3 members each and one team with 4 members. Challenge 2 required use cases, a use case diagram, content models, a content architecture diagram, and a navigational semantics diagram reflected in the developed prototype and the product specifications for the gardening software investigated in Challenge 1. Specific functionality was required in the web application that included service mashups, user and plant profiles, and a recommendation engine for gardening activities. The difficulty was considered hard, especially for the project duration of only 21 days. Build 1 (B-1-10) was performed the previous semester, Spring 2010, with students from Challenge 3 (described first). This build had 4 teams with 3 self-selected members. Build 1 required teams to document the scope and feasibility of their software, develop a product brand, identify a target market, define use cases and initial functional requirements, draft a Software Requirements Specification document, and produce a functioning prototype. The difficulty was considered moderate because the students had been trained in the fall course and the project duration was 43 days. The raw data for Ch-1-09 is shown in Table 2. Due to space requirements, data for Ch-1-10, Ch-2-10 and B-1-10 are included on our website [15]. Note that PM Avg, PM P, and PM C, refer to the average, Participation, and Content scores from Figure 3.

Table 2: Raw data collection for Ch-3-09

<table>
<thead>
<tr>
<th>User</th>
<th>Team ID</th>
<th>SEREBRO Score</th>
<th>Team Score</th>
<th>Eval Grade</th>
<th>PM Avg</th>
<th>PM P</th>
<th>PM C</th>
<th>SEREBRO Events</th>
<th>Artifact Count</th>
<th>Ideas</th>
</tr>
</thead>
<tbody>
<tr>
<td>CH-3_User-1</td>
<td>1</td>
<td>554.00</td>
<td>0.86</td>
<td>0.93</td>
<td>79.71</td>
<td>2.57</td>
<td>3.06</td>
<td>2.49</td>
<td>20</td>
<td>1</td>
</tr>
<tr>
<td>CH-3_User-2</td>
<td>1</td>
<td>797.50</td>
<td>0.86</td>
<td>0.95</td>
<td>81.43</td>
<td>3.38</td>
<td>3.87</td>
<td>3.88</td>
<td>62</td>
<td>14</td>
</tr>
<tr>
<td>CH-3_User-3</td>
<td>1</td>
<td>315.00</td>
<td>0.86</td>
<td>0.94</td>
<td>80.57</td>
<td>2.78</td>
<td>3.51</td>
<td>2.19</td>
<td>61</td>
<td>14</td>
</tr>
<tr>
<td>CH-3_User-4</td>
<td>2</td>
<td>741.75</td>
<td>0.92</td>
<td>0.95</td>
<td>87.76</td>
<td>2.93</td>
<td>3.14</td>
<td>2.73</td>
<td>52</td>
<td>2</td>
</tr>
<tr>
<td>CH-3_User-5</td>
<td>2</td>
<td>488.75</td>
<td>0.92</td>
<td>0.94</td>
<td>86.84</td>
<td>2.69</td>
<td>3.37</td>
<td>2.54</td>
<td>48</td>
<td>18</td>
</tr>
<tr>
<td>CH-3_User-6</td>
<td>2</td>
<td>530.00</td>
<td>0.92</td>
<td>0.93</td>
<td>85.91</td>
<td>3.00</td>
<td>3.71</td>
<td>2.81</td>
<td>52</td>
<td>1</td>
</tr>
<tr>
<td>CH-3_User-7</td>
<td>3</td>
<td>505.52</td>
<td>0.97</td>
<td>0.95</td>
<td>92.29</td>
<td>3.25</td>
<td>3.41</td>
<td>3.39</td>
<td>61</td>
<td>26</td>
</tr>
<tr>
<td>CH-3_User-8</td>
<td>3</td>
<td>468.33</td>
<td>0.97</td>
<td>0.90</td>
<td>87.43</td>
<td>2.69</td>
<td>3.43</td>
<td>2.70</td>
<td>25</td>
<td>2</td>
</tr>
<tr>
<td>CH-3_User-9</td>
<td>3</td>
<td>306.09</td>
<td>0.97</td>
<td>0.92</td>
<td>89.37</td>
<td>2.72</td>
<td>3.15</td>
<td>2.36</td>
<td>27</td>
<td>15</td>
</tr>
<tr>
<td>CH-3_User-10</td>
<td>4</td>
<td>825.23</td>
<td>0.92</td>
<td>0.98</td>
<td>90.53</td>
<td>3.16</td>
<td>3.64</td>
<td>2.93</td>
<td>51</td>
<td>7</td>
</tr>
<tr>
<td>CH-3_User-11</td>
<td>4</td>
<td>975.39</td>
<td>0.92</td>
<td>0.99</td>
<td>91.46</td>
<td>3.26</td>
<td>3.65</td>
<td>2.99</td>
<td>109</td>
<td>20</td>
</tr>
<tr>
<td>CH-3_User-12</td>
<td>4</td>
<td>455.94</td>
<td>0.92</td>
<td>0.83</td>
<td>76.68</td>
<td>2.72</td>
<td>3.05</td>
<td>2.57</td>
<td>30</td>
<td>8</td>
</tr>
</tbody>
</table>
4.2. Research Questions

We pose two research questions for our data analysis. The first question is concerned with how accurately SEREBRO Score and SEREBRO Events measure the content and contribution of an individual on a project within a team.

**RQ1: Do SEREBRO measures reflect the content and contribution of team members?**

Consistent with previous research into content based assessment [6], we use PM scores as measures of content and contribution. We normalize the PM scores according to the number of ideas an individual has by multiplying the number of ideas (Ideas from Table 2) by PM scores of interest (PM Avg, PM P, PM C from Table 2), for a particular user. The result of this product is the weighted PM score, or *WPM score*. Our second research question concerns which of the automated SEREBRO measures, from Section 3, predict student performance measured against a student's final project grade.

**RQ2: Which automated SEREBRO measures accurately predict final performance?**

We first validate the use of WPM scores, previously shown to correlate with performance in [6, 16], as predictors and then use them as baseline values of comparison.

5. Results

In this section, we answer RQ1 and RQ2 using Pearson's correlation *r*. Correlation values were calculated using sum of squares, as part of linear regressions between measure/metric pairs per challenge. Table 3 shows the average correlation, rounded to two digits or precision, across datasets. Values highlighted in blue represent the measures deemed significant for RQ1, while values in green show the significant measures for RQ2. Note that *WPM Participation*, *WPM Content* and *WPM AVG* are the *WPM Scores* calculated from *PM P*, *PM C*, and *PM Avg* respectively. For RQ1, we consider a correlation $|r| > 0.59$ to be significant (df=13, p=0.02) to say the measure $x$ under consideration accurately reflects the content based analysis measure *WPM score* ($y$). For RQ2, we consider correlations $|r| > 0.40$ to be significant (df=13, p=0.15) to classify the measure $x$, as a performance indicator using grade ($y$) as the metric of performance. Correlation values were also examined for variance, we found a low degree of variance across challenges, all $\sigma^2 < 0.149$. The full result of the variance analysis is presented with the individual challenge datasets on our website [15].

<table>
<thead>
<tr>
<th></th>
<th>Serebro Score</th>
<th>Grade</th>
<th>WPM AVG</th>
<th>WPM Participation</th>
<th>WPM Content</th>
<th>SEREBRO Events</th>
<th>Artifact Count</th>
<th>Idea Count</th>
</tr>
</thead>
<tbody>
<tr>
<td>Serebro Score</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Grade</td>
<td>0.41</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>WPM AVG</td>
<td>0.83</td>
<td>0.46</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>WPM Participation</td>
<td>0.81</td>
<td>0.44</td>
<td>0.99</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>WPM Content</td>
<td>0.84</td>
<td>0.43</td>
<td>0.99</td>
<td>0.98</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SEREBRO Events</td>
<td>0.63</td>
<td>0.40</td>
<td>0.84</td>
<td>0.84</td>
<td>0.82</td>
<td>1.00</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Artifact Count</td>
<td>0.22</td>
<td>0.24</td>
<td>0.39</td>
<td>0.38</td>
<td>0.39</td>
<td>0.75</td>
<td>1.00</td>
<td></td>
</tr>
<tr>
<td>Idea Count</td>
<td>0.75</td>
<td>0.43</td>
<td>0.93</td>
<td>0.94</td>
<td>0.91</td>
<td>0.82</td>
<td>0.30</td>
<td>1.00</td>
</tr>
</tbody>
</table>

5.1. Addressing RQ1

We found that both SEREBRO Score and SEREBRO Events are correlated with WPM Scores. SEREBRO Score correlated best with WPM Content, which leads us to believe it is indeed a reasonable automated measure of the content of the idea networks. Similarly the number of SEREBRO Events was positively correlated with WPM Content, which suggests that a team member's presence on the system is related to the quality of the end result. This is
substantiated by other research [5]. Unsurprisingly SEREBRO Events, which represents the sum of a user's contribution across all areas logged by the system, correlates best with WPM Participation. All correlations were very strong, \( r > 0.8 \), which coupled with the low degree of correlation variances between measures \( \sigma^2 < 0.15 \), high SME agreement evidenced by the low degree of variance among SME responses \( \sigma_{avg}^2 < 0.25 \) and a 98% confidence interval gives significant weight to using SEREBRO measures as content and contribution assessment mechanisms.

5.2. Addressing RQ2

In RQ2, we found that WPM Scores performed the best as performance indicators, with a correlation \( 0.43 < r < 0.46 \), validating previous work by others [6, 7, 16] showing content quality and participation to be indicative of performance. SEREBRO measures, excluding artifact count which we found to not be indicative of performance, performed well \( 0.40 < r < 0.43 \), given the desirable automated nature of the measures. Overall through examination of RQ2, we found that measuring content, through both automated and manual methods, could be used to predict performance with 85% confidence in the result.

6. Related Work

Idea management platforms allow idea submission, refinement and alerts, though they generally have focus on a few specific qualities. They differ from SEREBRO in how the idea is refined, credited, and visualized. Some idea management platforms allow a variety of commenting [17-20], rating [17, 21] and voting [19, 21]. Some platforms provide participation metrics [17-19, 22, 23], though none directly track individual progress, provide productivity tools, or allow external content mapping of ideas to tangible artifacts [14].

Issue-Based Information Systems (IBIS) are a concept structure which utilizes a set language to simulate a dialogue on an issue, the basic structure of which is a tree. IBIS look at solving 'issues' by organizing the communication of stakeholders within a project [24]. In short an IBIS attempts to capture fragmentary collaboration as a problem, or issue, which is proposed then supported, or weakened, by arguments (pro or con) [18, 21, 25].

It is well known that communication is a critical component of successful team-based software development projects [1-3]. Various metrics have been used to measure the content, type and quality of communication. Early work by Dutoit and Bruegge [5] demonstrate the effectiveness of quantitative communication metrics, such as word count and number of messages sent by a team per day, as analysis tools indicative of team performance. Such coarse group level quantitative measurements give performance appraisers a rough feel for overall work activity but do not capture the semantics of the communication artifacts.

Content based qualitative analysis as applied to communication artifacts, such as threads, posts, ratings, and alerts, allow them to be classified into various content categories according to the type, intent-of or goal. The artifacts can be sorted and tabulated by content category [4]. Examining the expressions given a category can produce communication patterns that correspond to high or low performing teams. For example, Serce et al. [6] identify five team content categories: social interaction, planning, contributing, seeking input and reflection. They show that teams judged to be very contributive received higher project grades than teams with increased levels of social interaction or with low levels of planning. While categorical content-based metrics give a detailed view of team communication over rough quantitative measurements, such as number of messages, they can be more time consuming to calculate, requiring "labor intensive" assessment of each message [4]. Automated methods
using natural language processing (NLP) and word identification can reduce the work involved [16], but the quality of ratings is not as good as instructor in the loop.

Wolf et al. [7] use social network analysis to suggest developers and managers can use communication metrics to predict final performance. By measuring the levels of social connectedness among team members in project groups using IBM's Jazz [26], and correlating the quality of communication artifacts to their corresponding development artifacts they are able to predict failed software module integration of a subgroup’s code with the larger software product [7]. Thus, it is possible for managers to "adjust collaborative tools or processes towards improving the [quality of] integration [7]."

Prior to analysis, communication artifacts must first be extracted from documents and email for examination. The extracted information is entered into a Knowledge Management (KM) system, which may be in the form of a wiki or forum of some kind. Janz et al. [8] partition knowledge into two types, tacit – knowing-how and explicit – knowing-what. KM environments typically already support explicit knowledge capture [8]. Without an integrated development tool, capturing tacit knowledge can add significant overhead to the development process and is often lost if busy software teams neglect to forcibly codify it, potentially resulting in poorly documented software systems [27]. KM research also examines the role of individual incentive based rewards (IBR) as motivators for effective knowledge sharing [9, 10]. Since IBRs are organizationally defined, they result in both an individual's increased utility through the achievement of reward (e.g. pay, grade, reputation, etc) and increased knowledge sharing as a team [28]. Overall, KM provides the design and implementation rationale of the team while providing for easier reuse, revision, and re-engineering.

7. Discussion and Conclusion

SEREBRO as courseware provides a broad set of features targeted toward studying software engineering. Tied to these features are indicators that cue the instructor and team members as to their performance with respect to their collaboration, contribution, and progress toward stated milestones. We show that these indicators correlate to SME ratings of content and contribution of an individual in idea networks and to instructor project grades on work products associated with milestones. Thus, automatic SEREBRO assessment mechanisms are able to predict an individual's grade and contribution to a project team.

We are currently pursuing additional types of analysis to examine an individual's performance at filling particular roles on the team, such as lead, analyst, and programmer, and team dynamics over the project milestone period, such as how "bursty" vs. consistent communication relates to milestone success. Similar to other researches [7, 16], we are interested in how performance indicators might be combined to yield real time predictions of impending build or challenge shortcomings. Our goal is to derive a weighted product which can be used to calculate an individual's likelihood of failure given their progress towards milestone completion and mitigate that failure before it manifests itself. While we are several steps away from this, deriving performance metrics from existing real time system data lends support to our continued pursuit of performance classification.

Acknowledgement. This material is based upon work supported by the National Science Foundation under Grant No. 0757434. Any opinions, findings and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation (NSF).
8. References


