ISSN 2395-8618

Collaboration and Content-Based Measures to Predict Task Cohesion in Global Software Development Teams

A. Castro-Hernández, V. Pérez-Rosas, and K. Swigger

***Abstract*—Task cohesion is a key component of team performance. This paper explored the use of collaboration and content-based measures to examine task cohesion within global software development teams. The study aimed to predict the perception of task cohesion among teams involving students from two different countries. The study applied collaboration from previous work and also proposed new metrics such as Reply Similarity and Reply Rate. In addition, a machine learning classifier is used to derive content measures by categorizing teams' message interactions as social, planning, or work. Correlation analyses are conducted to examine whether collaboration and metrics are predictive of task cohesion. The analyses are conducted at the individual and group levels and used the culture factor as a control variable since cohesion has been found previously affected by location. The research findings suggest that content-based measures were more effective in predicting individual-level cohesion while collaboration-based metrics were more effective at the group-level.**

***Index Terms*—Component, formatting, style, styling, insert.**

1. Introduction

WORKING with global teams presents important challenges associated with managing time, distance, and communication technologies. Most of the information is, in many cases, stored in databases. However, for a user to obtain information from a database (DB), he must have knowledge of a query language for databases (such as SQL). [1].

Particularly, in educational settings, global teams provide students with many valuable experiences but also pose several challenges such as having to deal with people from diverse cultures, coping with different perceptions of time and relationships, and finding effective communication tools that allow distributed groups to work together.

Needless to say, any one of these challenges can have a significant effect on team performance. Among several team processes affected by computer-mediated communication (such as cohesiveness, status, and authority relations), cohesion has remained a critical issue for all types of work teams. In global teams, positive cohesion levels have been directly linked to group performance [1]. However, when compared with co-located teams, virtual teams tend to be less cohesive [2]. Thus, making it important to develop effective methods to measure cohesion that allow opportune team interventions.

While researchers have proposed several ways to measure task cohesion levels within groups, very few have been tested on or designed for virtual learning teams.

In this paper, we examine several existing cohesion metrics that characterize different degrees of similarity among group members (e.g., word category usage, reply behavior, etc.). We also proposed new measures that capture interaction aspects such as word and reply rates. We explored the question of whether quantitative group measures are better at predicting cohesion than those that are associated with the individual. Thus, our main objective was to determine whether similarity measures are better at predicting cohesion levels than quantity-based measures and whether individual measures are more related to cohesion perception than group measures.

2. Related Work on Cohesion Measures

Cohesion is usually defined as "a dynamic process which is reflected in the tendency for a group to stick together and remain united in the pursuit of its goals and objectives" [3]. It has been studied at both the individual and group levels [4] and has been linked to group performance [5]. Group cohesiveness in any type of team seems to increase over time, particularly when there is a leader in the group [6]. Other elements that appear to affect a group's cohesiveness include team size, degree of democratic behavior within a group, participation, and satisfaction [7].

The Group Environment Questionnaire (GEQ) [8] is a survey instrument commonly used to measure individual perceptions of group cohesion. GEQ consists of 18 items that measure group and individual factors, i.e., group integration and individual attraction to the group. These factors are further divided into tasks and social dimensions, which describe general motivation toward achieving group objectives and developing social relationships.

Manuscript received on October 21, 2020, accepted for publication on November 25, 2020, published on December 30, 2020.

A. Castro-Hernández is with the Department of Computer Science and Engineering, Wayne State University, Detroit, USA (e-mail: hg3246@wayne.edu).

V. Pérez-Rosas is with the Department of Computer Science and Engineering, University of Michigan, Ann Arbor, USA (e-mail: vrncapr@umich.edu).

K. Swigger is with the Department of Computer Science and Engineering, University of North Texas, Denton, USA (e-mail: kathy@cse.unt.edu).

ISSN 2395-8618

In addition to surveys, researchers have also developed more objective measures for characterizing a group's cohesiveness. These measures can be helpful in the analysis of team cohesion in remote work, which has particularly increased due to the impact of COVID-19 on our society [9]. For instance, [10] described group cohesion in terms of a density measure expressed as a ratio of the number of connections among members over all possible links. However, this approach did not take into account the weight or intensity of those connections. Thus, [11] calculated group cohesion (density) by looking at only the links that have weights higher or equal to a pre-defined number, which may vary according to different contexts.

Another type of cohesion measure, called Linguistic Style Matching (LSM), was developed by [12]. This particular measure is based on the similarity of the use of function words between two individuals. Once all paired similarities among group members are computed, the paired values in the group are then averaged, and this number becomes the group's cohesiveness score. Using this technique, the researchers found a correlation between LSM and a cohesion construct (obtained through a survey) and a limited relation between LSM and performance. This particular study tested the LSM measure using chat communications generated during a one-hour session from single-gender teams. However, researchers who applied LSM to the analysis of email messages among team members over an extended period of time were unable to duplicate the significant relationships between cohesion and performance [13]. Moreover, the use of function words in a non-native speaker group setting might also affect LSM's predictive capabilities.

In another study, [14] proposed a measure called *Individual cohesion* that is designed to predict group cohesion. The *Individual cohesion* measure is calculated by summing messages between each pair of individuals on a team and then averaging those counts. Members' individual cohesion scores are then correlated with task cohesion. Although the study found a significant relationship between *Individual cohesion* and a group's cohesion level, it was noted that this similarity approach might be affected by individuals who perform poorly but have similar interaction scores. Thus, the authors suggested that a measure based on communication intensity might be a better predictor of cohesion in a virtual setting because of the low interaction rates often found among group members in this type of setting.

The above research served as an important tool for defining the major factors that were deemed important for this study. These factors include both similarity and quantitative measures. The general question asked was which of these factors tends to be a better predictor of cohesion levels among global software learners. The measures mentioned above represent some of the factors that were included in the experiments for this study.

TABLE I
ACTIVITIES PER PROJECT AND INSTITUTION.

	US institution	MX institution
Project 1	Museum website	Database functionality
Project 2	Museum website redesign	Database functionality
Project 3	Learning website	Learning website functionality

3. DATASET

Our data is drawn from a global software development study involving students from three higher education institutions located in the US and Mexico (MX). Data collection occurred from 2014 to 2015 and included software development projects from three different undergraduate-level courses. Participants consisted of 116 males and 62 females who were 22-years old on average.

All participant's communications happened in English and used an online collaborative tool called Redmine [15]. Teams were formed by pairing students from the American institution and either of the Mexican institutions. Participants were randomly assigned to their respective groups. The teams worked on three different projects involving either the design of a museum or a learning website as well as implementing their functionality. Activities were distributed as shown in Table III.

We created a project management web application in Redmine that allowed us to record communications occurring among the team's members. These include participation in chats, forums, and wikis, as well as file sharing. We also recorded the date and time in which each activity occurred as well as the author of each online activity.

A. Collaboration Measures

We explored two main team factors in relation to cohesion: communication similarity and communication processes. We obtained assessments of *Task Cohesion* using individual questionnaires completed by most of the team members. The questions in the survey are derived from the multidimensional cohesion model developed by [8] and also from work by [16].

B. Communication similarity metrics

We derive communication similarity metrics from the graph representation of communication replies among team members. Figure 1 shows a sample graph derived from the interaction shown in Table II. In this figure, the conversation is shown as a directed graph, where vertices represent the different team members, and edges represent the number of replies received by participants. For example, u_2 replied to u_1's initial message, while u_1 replied to u_2's message about studying engineering. The reply counts are extracted from the forum and chat exchanges and reflect when a message is delivered to a specific participant (i.e., the last participant in the communication).

ISSN 2395-8618

TABLE II
EXAMPLE OF COMMUNICATION AMONG THREE TEAM MEMBERS.

User	Message
$u1$	Hello. My name is Bryan and I am a student at UNT...
$u2$	Hi, I am Carlos.
$u2$	I live in Panama.
$u2$	I study Computer Engineering.
$u1$	Nice two meet you.
$u1$	I just read the project description and it seems that...
$u3$	Hello I live in Texas, and I study Computer Science.

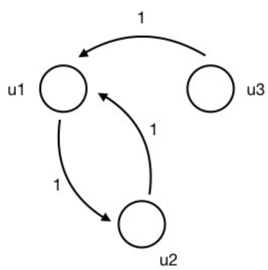

Fig. 1. Reply graph of conversation in Table II

We generate reply graphs for each team and project and then derive three similarity metrics as described below. Density is a group measure that describes the degree of interconnectedness among group members, with high density indicating a high degree of interconnectedness and low density indicating a low degree of interconnectedness [10]. The original density metric is calculated at the group level; however, we redefine it at the individual level. Where the *individual density* of a team participant (D_i) is the count of people with whom the participant interacted with (*actual_links$_i$*) over the number of people in the rest of the team (*possible links*).

$$D_i = \frac{actual\_links_i}{possible\_links} \quad (1)$$

Reply similarity We also measure the similarity on reply behaviors at the individual and group levels. These measures are derived from the cohesion metric proposed by [17] and are based on the idea that people view their group's cohesiveness as being a combination of their own participation as well as from others. We thus calculate the reply rate at the team level (*reply similarity*, where $reply_{ij}$ is the number of replies from person i to person j, and $reply_{ji}$ is the number of replies from person j to person i.

$$replysimilarity_i = 1 - \frac{|reply_{ij} - reply_{ji}|}{reply_{ij} + reply_{ji}} \quad (2)$$

The reply similarity metric produces a score between 0 and 1, indicating the degree of reply similarity among team members. The *individual reply similarity* indicates whether participation in group discussions increased the individual's perception of group cohesion and was simply calculated as the total number of replies that an individual sent.

Linguistic Style Matching (LSM) is a metric used to determine whether individuals match their language use. Its use as a cohesion measure was first proposed by [12]. We calculated LSM using the formula shown in Equation 3, where x is a function word from the LIWC lexicon (e.g., auxiliary verbs, articles, common adverbs, personal pronouns) and x_i denotes their use frequency by the i team member (or j). Using this formula, we determined the similar usage of function words between member i and each of the other members in the team.

$$LSMx_{ij} = 1 - \frac{|x_i - x_j|}{x_i + x_j} \quad (3)$$

Information Exchange We derived two metrics for information exchange among team members: similarity and rate. The first (*Information Exchange Similarity*) is calculated by simply counting the number of words typed by each participant, with the assumption that these words were being transmitted to every other team member in the group (i.e., each word in each message is perceived as some type of participatory exchange). This calculation was computed as follows:

$$informationexchange_i = 1 - \frac{|wc_i - wc_j|}{wc_i + wc_j} \quad (4)$$

where wc_i and wc_j are the numbers of words used by user i and j, respectively.

The second metric, called Information Exchange Rate, is calculated as the total word count by a team participant.

C. Content Measures

Previous research showed that words associated with word categories such as contribution, seeking input, reflection, social, and planning are highly related to cohesion perception [18]. For example, the use of words related to social content seems to influence the level of trust among team members as well as create a more pleasant environment within a team. Although these word categories may not have a direct effect on *Task Cohesion*, it is natural to assume that they may have some indirect effect on *Task Cohesion*; for example, the absence of social behavior may result in a decrease in *Task Cohesion* among group members.

We thus derived a set of content measures that capture the use of these categories.

Social: number of messages using language related to social interactions.

Planning number of messages containing language related to organizing the project development, i.e., verbs, nouns, or dates. Work messages containing information about the project organization and management. Borrowing from the original research [19], the work category included words related to *Contribution* and *Seeking input*.

4. RESULTS

A total of 5,583 messages were transmitted during the three projects. A total of 167, out of a possible 180, *Task Cohesion*

ISSN 2395-8618

surveys were collected. Since we had 23 missing surveys, we kept only messages by students who had completed the questionnaire; thus, our dataset included a total of 5446 messages.

A. The Culture Effect

A previous study on *Task cohesion* found that location (or country of birth) affects *Task cohesion* perception [14]; and since our experiments included participants studying in different locations (i.e., US or Mexico) but born in different countries, we anticipated that culture might have an effect on individuals' perceived cohesion.

TABLE III
TASK COHESION VALUES BY CULTURE. *$p < 0.05$.

Country	n	mean	India	US
India	55	8.01		
US	20	6.21	1.807*	
Mexico	78	6.48	1.5310*	0.2746

TABLE IV
PARTIAL CORRELATIONS OF SIMILARITY MEASURES AND DENSITY TO TASK COHESION, CONTROLLED BY CULTURE. †$p < 0.10$., *$p < 0.05$.

Measure	Task cohesion
Individual Density	0.037
Reply similarity	0.060
Linguistic Style Matching	0.114†

Our data contained survey responses from people who were born in eleven different countries. However, most of the surveys were completed from students born in India, the US, and Mexico (i.e., $n > 20$), while the rest of the countries were represented by only a few surveys (i.e., $n < 4$). As a result, we reduced our dataset even further and used data from only students born in any of the three main countries represented in our study; thus, ending up with a final count of 4,849 messages sent by 153 participants. We then conducted a preliminary analysis to evaluate whether *Task Cohesion* assessments differed among respondents. We compared the Task cohesion mean values between countries and found that students from India tended to have higher *Task Cohesion* perceptions than either US or Mexican students (see Table III). Thus, we decided to use the *Culture* factor as a control variable in our analyses.

B. Similarity Metrics in Task Cohesion

In order to determine the relationship between our similarity measures and *Task Cohesion*, we computed a partial correlation between each measure (i.e., *Individual Density, Reply Similarity, Linguistic Style Matching,* and *Task Cohesion*), controlling for the *Culture* factor. The results of these correlations are reported in Table IV.

Results suggest that the *Density* measure at the individual level is unrelated to *Task Cohesion* (r=0.037, p=0.325). Since weights assigned to different edges in the reply graphs tended to vary widely (ranging from 1 to 70), we suspect that the large variance among participants' exchanges may have affected the correlation between *Density* and *Task Cohesion*. Similarly, we found no correlation between *Reply Similarity* and *Task Cohesion* (r=0.060, p=0.230). The lack of a correlation between these two variables may be explained by the inactivity of one or two group members. For example, we noted that in cases where at least one of the group members was not participating in group discussions, the *Reply Similarity* score was low.

TABLE V
PARTIAL CORRELATIONS OF QUANTITY-BASED MEASURES TO TASK COHESION, CONTROLLED BY CULTURE. †$p < 0.10$., *$p < 0.05$.

Measure	Similarity	Rate
Reply	0.060	0.108†
Information exchange	0.152*	0.175*

On the other hand, *Linguistic Style Matching* has a positive, although low, significant correlation with *Task Cohesion* (r=0.114, p=0.081). An analysis of the use of function words by culture for the first week of each project shows students who participated in the Spring 2014 project had a significant difference in their use of personal pronouns (p=0.027) and quantitative words (p=0.054); also, students who participated in the Spring 2015 project show a significant difference in the use of impersonal pronouns (p=0.055); however, all 2014 participants showed no difference in function-word usage by country. It is important to note that the Fall 2014 project consisted largely of students who were born in either Mexico or the US, while the other two projects consisted mainly of students born in either Mexico or India. Thus, it appears that US and Mexican students have more similar linguistic patterns than Mexican and Indian students.

Finally, *Information Exchange Similarity* has a positive and significant correlation with *Task Cohesion* (r=0.152, p=0.031). Although *Information Exchange Similarity* and *Reply Similarity* are somewhat related, in terms of what they are measuring, the simple word-based metric of *Information Exchange Similarity* shows a better correlation with cohesion perception, possibly because words rather than replies (which tend to be sentences) produce more data. The amount of data that is used to analyze the relationship among variables may affect the degree of significance. Word counts may also be a better metric of different levels of interaction since individuals who tend to be more engaged in the project will probably communicate more, which in turn may affect the perception of the group's cohesiveness.

C. The Effect of Quantity-Based Measures in Task Cohesion

We next tried to determine if the intensity of the team's communications, as measured by *Reply Rate* and *Information Exchange Rate*, has an effect on Task cohesion. Again, we

ISSN 2395-8618

conducted a partial correlation analysis controlled by the *culture* factor between these two variables and Task Cohesion. Results are shown in Table V. The first column shows the correlation scores for both *Reply Similarity* and *Information Exchange Similarity*, while the second column shows the correlation scores for *Reply Rate* and *Information Exchange Rate*.

Our results indicate that *Reply Rate* has a positive, although low, significant correlation with *Task Cohesion* (r=0.108, p=0.093). *Information Exchange Rate* also has a significant positive correlation with *Task Cohesion* (r=0.175, p=0.016). In comparison with the regular similarity measures, the more quantitative-based measures have higher correlation values.

TABLE VI
PARTIAL CORRELATIONS OF GROUP MEASURES TO TASK COHESION, CONTROLLED BY CULTURE. †P < 0.10.

Measure	Task cohesion
Group reply rate	0.113†
Group information exchange rate	0.112†

This seems to suggest that quantity-based measures that capture amounts, and perhaps engagement, may be better predictors of group cohesion than measures that try to assess different similarity constructs in a team's exchanges.

D. Effect of Group Measures

We then took the two variables that appeared to be significantly related to task cohesion and calculated group-level scores for each of these two measures. Thus, we created a *Group Reply Rate* and *Group Information Exchange Rate* variable and examined the relationship between these two factors and *Task Cohesion*. Table VI shows the correlations for each of these two variables.

Group Reply Rate and *Group Information Exchange Rate* show a positive but low, significant correlation with *Task cohesion*, i.e., r=0.113, p=0.083 and r=0.112, p=0.085, respectively. These results suggest that an individuals' and a group's engagement in communication is a predictor of task cohesion perception.

E. Content-based Features for Task Cohesion

To enable these analyses, we devised a data-driven approach to extract content features from student's messages. We thus aimed to categorize participant messages into the social, planning, and work categories. Note that we used this approach as custom lexicons for these categories are not readily available. We thus developed an automated text analysis program that could classify students' messages into three different categories. To conduct our experiments, we used a dataset consisting of 1,866 messages that had been manually annotated in previous work [19], [20] with the planning, contributing, seeking input, reflections, monitoring, and social categories.

Our study only used four of these categories (i.e., social, planning and work) with the work category including the contribution and seeking input categories. Thus, messages that contained the *agreement* label were removed so that these types of short messages were not included in any of our counts for the proposed categories. Agreement messages are those that confirm or deny some previous message such as "ok," "sure," "good," etc. This resulted in a dataset consisting of 305 messages under the social category, as well as 166 and 1279 messages in the planning and working categories, respectively.

The features used to help seed the classification process were those found in research related to the LIWC software tool. Using this tool, we investigated 73 features. We also computed the unigrams of each message, obtaining a dictionary of more than 3000 entries. We tested the ability of both feature sets (i.e., Unigrams & LIWC) to predict the target label. Similarly, we tested the use of LIWC features only during the classification task.

We compared the performance of three classifiers: Support Vector Machines, Random Forest, and Naive Bayes. The performance metric used for comparison was the F1-score, which is the harmonic mean of precision and recall. Also, we conducted 10-fold cross-validation during these experiments.

TABLE VII
F-SCORE VALUES FOR CLASSIFICATION ALGORITHMS.

Feature set	Support Vector Machines	Random Forest	Naive Bayes
LIWC+Unigrams	0.821	0.777	0.522
LIWC	0.752	0.776	0.495

As seen in Table VII, all classifiers obtained better results by using both the LIWC and Unigrams features as compared to using only LIWC features. In addition, the best performances were obtained by the Support Vector Machines classifier (F1score=0.821). Hence, we used the Support Vector Machine classifier to label messages in our dataset. The result of applying the classifier to the messages in the experimental dataset can be found in Table VIII.

TABLE VIII
DISTRIBUTION OF MESSAGE CONTENT CATEGORIES IN THIS STUDY

Class	Instances
Social	936
Planning	4528
Work	392

Some of the features that provided more information for classifying each message into a category are shown in Table IX. The Work category was found to be related to unigrams that are sometimes associated with performing activities such as give, talk, data, instructor, right. However, the

ISSN 2395-8618

Work label was also found to be related to three LIWC categories: 1) function words *LIWC.FUNCTION* (e.g., prepositions, articles, and common adverbs), which are often related to the idea of formal thinking [21]; 2) work *LIWC.WORK* (e.g., accomplish, work, and success), which are often related to the idea of well-performing teams [14]; 3) Interrogatives *LIWC.INTERROG*, (e.g., how, when, what, which are often used to represent the students' seeking input process.

TABLE IX
RELEVANT WORDS BY CATEGORY, ACCORDING TO CLASSIFIER.

Work	Social	Planning
Right	LIWC.INFORMAL	meeting
LIWC.FUNCTION	fun	night
LIWC.WORK	Lets	tonight
LIWC.INTERROG	Thanks	Thursday
data	Oh	whenever
talk	LIWC.FOCUSPRESENT	must
instructor	later	yet
could	LIWC.SOCIAL	schema
modification	Nice	yours
give	hello	Monday

Similarly, words related to the Social category included words such as fun, lets, thanks, hello, later, nice - all of which seem to suggest characteristics related to social interaction. The Social category also included more concrete relations with specific LIWC categories: 1) informal *LIWC.INFORMAL*, a category that consists of words such as netspeak (lol, btw, thx), swear words (fuck, damn, shit), non-fluencies (err, mmm); 2) *Present tense* (*LIWC.FOCUSPRESENT*), a category that includes words such as today, is, now, which are words related to more personal information sharing; 3) social (*LIWC.SOCIAL*), a category that consists of words such as mate and they. Again, LIWC provided the research with relevant key categories related to specific conversation types.

On the other hand, the main word features that comprised the Planning label did not include any specific LIWC category. Instead, the classifier produced unigrams related to project management activities such as night, whenever, days of the week (Thursday, Friday, Monday). We also saw patterns that included words related to specific management activities such as meeting and schema (which was probably because of the type of projects that were assigned, e.g., database schema). The Planning label also included words related to future events, which are often used in planning tasks. Given this particular list of words, it should be possible, at some time in the future, to create a LIWC category that would automatically identify these types of communication.

These results show the importance of the LIWC tool and the use of unigrams to obtain the appropriate label for the message's exchanges by team members.

After placing the messages into their various content categories, we computed the correlations between *Social Similarity*, *Planning Similarity*, and *Work Similarity* and *Task Cohesion*. In addition, we computed correlations between *Social Rate*, *Planning Rate*, and *Work Rate* and the same target variable.

TABLE X
CORRELATIONS OF CONTENT-BASED VARIABLES WITH TASK COHESION. †P < 0.1, *P < 0.05

Content type	Similarity	Rate
Social	0.052	0.136*
Planning	0.001	0.060
Work	0.101†	0.063

Table X shows the correlations of these variables when controlled by team size and culture. The *Work Similarity* variable shows a nearly statistically significant correlation with *Task Cohesion* (r=0.101). However, neither *Social Similarity* or *Planning Similarity* are correlated with the cohesion construct.

On the other hand, there exists a statistically significant correlation (r=0.136) between the *Social Rate* variable and *Task Cohesion*. But, again, *Work Rate* and *Planning Rate* do not show a statistically significant correlation.

TABLE XI
AVERAGE AND SUM OF SIMILARITY AND RATE CONTENT MEASURES

Content type	Similarity (average)	Rate (sum)
Social	0.216	559
Planning	0.081	203
Work	0.408	2046

The statistically significant *Work Similarity* correlation can be explained by looking at previous results that show the perception of Group Cohesiveness is often affected by a group's perception that all members are doing their fair share of the work. So, if evidence shows that all members are participating in the communication, then an individual's perception of the group's cohesiveness should be higher. Moreover, the similarity between members' conversations about work (i.e., *Work Similarity*) seems to be more important than the rate at which these exchanges occur.

Interestingly, *Social Rate* shows a stronger correlation with *Task Cohesion* than *Work Rate*. This result may demonstrate the importance of having some social communications among group members within a virtual environment. Despite the possibility that participants' social interactions may sometimes affect the accurate assessment of *Task Cohesion*, as discussed in [7], we believe that this did not occur in the context studied in this research, as evidenced by a large number of work-related communications as compared to social messages that were transmitted among group members, as shown in Table XI.

ISSN 2395-8618

In contrast, there were no correlations between *Planning Similarity* and *Task Cohesion*, nor between *Planning Rate* and *Task Cohesion*. This lack of statistical significance between *Task Cohesion* and the planning construct may be due to the relatively small number of messages that were generated in this category.

We also evaluated the performance of these same metrics at the group-level, i.e., we calculated *Group Social Similarity*, *Group Work Similarity*, *Group Planning Similarity*, *Group Social Rate*, *Group Work Rate*, *Group Planning Rate*. We first removed teams that had one or more participants who were not identified with one of the major countries, i.e., the United States, India, and Mexico. Also, we computed the *Group Task cohesion* measure by averaging the *Task Cohesion* perception of the team members. When a survey was not completed by a student, *Group Task Cohesion* was estimated using a missing data technique based on systematic non-response [22]. A total of 18 teams (out of 35) were used in the group-level analysis.

TABLE XII
CORRELATIONS OF CONTENT-BASED VARIABLES AT THE GROUP-LEVEL WITH TASK COHESION.

Content type	Similarity	Rate
Social	–0.235	0.088
Planning	–0.124	0.192
Work	–0.062	–0.038

Results shown in Table XII indicate that none of the group variables had a statistically significant correlation with *Task Cohesion*. This result may have occurred because of the small sample size for the comparison. Thus, more data might be required to determine whether the individual-level metrics, which we found to be related to *Task Cohesion*, i.e., whether (*Work Similarity* and *Social Rate*), have similar effects when aggregated at the group level.

5 CONCLUSIONS

This research examined a number of different measures to determine whether certain types of factors were better predictors of *Task Cohesion* than others. Similarity measures, drawn from previous research, and new quantitative measures created for this study were examined. The measures were applied to data generated from three global software learning projects that took place between students located in Mexico and the US.

Individual similarity measures intended to predict *Task Cohesion* had mixed results. For example, *Individual Density* was not correlated with *Task Cohesion*. At the same time, *Reply Similarity* had no significant relation to cohesion. The lack of significant results for these two variables can be explained by looking at the number of exchanges between members in the co-located teams. Students seemed to have fewer exchanges between team members in their own country as opposed to team members in the remote country. Possibly, co-located team members may have had offline conversations that were not measured. It may also be the case that the measure itself needs to be redefined since scores for both *Density* and *Reply similarity* are highly affected by inactive members. On the other hand, *Linguistic Style Matching* had a small but significant relationship to group cohesiveness. We speculate that the significance was low because of the large difference between the use of function words by Indian students as compared to Mexican students. *Information Exchange Similarity*, which is a word-based calculation, also achieved a significant correlation with *Task Cohesion*. This seems to suggest that a simple measure of the number of words exchanged provides a better representation of cohesion than either messages or replies, perhaps because such a measure generates more data for the analysis.

Since communication within a virtual learning team tends to vary, metrics based on interaction intensity were also proposed, i.e., *Reply Rate* and *Information Exchange Rate*.

Both seemed to predict *Task Cohesion* much better than the similarity version of these two variables. It should be noted that the word-based factor of *Information Exchange Rate* was a better predictor of *Task Cohesion* than *Reply Rate*. Although we computed group versions of *Reply Rate* and *Information Exchange Rate*, we found only low significant relationships between either one of the variables and *Task Cohesion*.

These results seem to indicate that previously cited similarity measures used to predict task cohesion may be limited to analyzing groups that are highly interactive and that work on short-term tasks. Data from this study suggests that similarity measures may be affected by less-active members and the presence of cross-cultural teams, both of which can impact message length and word usage. Since global software teams have both of these characteristics, similarity measures may have limited value in computing cohesiveness for distributed learning projects. However, this study also found that several quantitative measures that captured both reply and word rates were useful predictors of a group's cohesiveness within a global software development learning team.

Furthermore, measures based on the content within the communications were also developed. First, we generated a message classifier that performed accurately. Using output from the classifier, we were able to analyze their relationship with *Task Cohesion*. These measures showed mixed results. Only *Work Similarity* and *Social Rate* were found correlated to *Task Cohesion*. The statistically significant relation between *Work Similarity* and *Task Cohesion* can be explained by looking at previous results that show that the perception of cohesiveness is often affected by a group's perception that all members are doing their fair share of the work. So, if evidence shows that all members are participating in the communication, then an individual's perception of the group's cohesiveness should also be higher. On the other hand, the strong

ISSN 2395-8618

correlation between *Social Rate* and *Task Cohesion* shows that it is important to have at least some social communications during a software development project since these types of interactions can lead to an increase in trust within the group, which may result in an increase in group cohesiveness [13].

Further research is needed to determine whether a temporal approach (which may help to identify when good-performance teams start working) can produce better predictions of the perceived cohesiveness within a distributed team. Until that research is completed, we believe that similarity and, to a larger extent, content measures can be used to predict group cohesiveness within a global software student project.

REFERENCES

[1] P. Goodman, E. Ravlin, and M. Schminke, "Understanding Groups in Organizations," *Tepper School of Business*, Jan. 1987.

[2] K. Furumo and J. Pearson, "An Empirical Investigation of How Trust, Cohesion, and Performance Vary in Virtual and Face-to-Face Teams," in *Proceedings of the 39th Annual Hawaii International Conference on System Sciences, 2006. HICSS '06*, vol. 1, Jan. 2006, pp. 26c–26c.

[3] A. V. Carron, "Cohesiveness in sport groups: Interpretations and considerations." *Journal of Sport psychology*, 1982.

[4] S. M. Gully, D. J. Devine, and D. J. Whitney, "A meta-analysis of cohesion and performance effects of level of analysis and task interdependence," *Small Group Research*, vol. 26, no. 4, pp. 497–520, 1995.

[5] S. S. Webber and L. M. Donahue, "Impact of highly and less job-related diversity on work group cohesion and performance: A meta-analysis," *Journal of management*, vol. 27, no. 2, pp. 141–162, 2001.

[6] V. L. Schwanda, K. Barron, J. Lien, G. Schroeder, A. Vernon, and J. T. Hancock, "Temporal Patterns of Cohesiveness in Virtual Groups," in *Proceedings of the ACM 2011 Conference on Computer Supported Cooperative Work*, ser. CSCW '11. New York, NY, USA: ACM, 2011, pp. 709–712.

[7] E. Rovio, J. Eskola, S. A. Kozub, J. L. Duda, and T. Lintunen, "Can High Group Cohesion Be Harmful? A Case Study of a Junior Ice-Hockey Team," *Small Group Research*, vol. 40, no. 4, pp. 421–435, Aug. 2009.

[8] A. V. Carron, W. Widmeyer, and L. R. Brawley, "The development of an instrument to assess cohesion in sport teams: The Group Environment Questionnaire." *Journal of sport psychology*, 1985.

[9] A. Ozimek, "The future of remote work," *Available at SSRN 3638597*, 2020.

[10] S. Wasserman, *Social network analysis: Methods and applications*. Cambridge University Press, 1994, vol. 8.

[11] C. Reffay and T. Chanier, "How Social Network Analysis can help to Measure Cohesion in Collaborative Distance-Learning," in *Designing for Change in Networked Learning Environments*, ser. Computer-Supported Collaborative Learning, B. Wasson, S. Ludvigsen, and U. Hoppe, Eds. Springer Netherlands, Jan. 2003, no. 2, pp. 343–352.

[12] A. L. Gonzales, J. T. Hancock, and J. W. Pennebaker, "Language Style Matching as a Predictor of Social Dynamics in Small Groups," *Communication Research*, vol. 37, no. 1, pp. 3–19, Feb. 2010.

[13] S. A. Munson, K. Kervin, and L. P. Robert Jr, "Monitoring Email to Indicate Project Team Performance and Mutual Attraction," in *17th ACM conference on Computer supported cooperative work & social computing*. New York, NY, USA: ACM, 2014, pp. 542–549.

[14] A. Castro-Hernandez, K. Swigger, and M. P. Ponce-Flores, "Effects of´ cohesion-based feedback on the collaborations in global software development teams," in *Collaborative Computing: Networking, Applications and Worksharing (CollaborateCom), 2014 International Conference on*. IEEE, 2014, pp. 74–83.

[15] Redmine. (2021) Redmine overview. [Online]. Available: www.redmine.org

[16] S. A. Carless and C. D. Paola, "The Measurement of Cohesion in Work Teams," *Small Group Research*, vol. 31, no. 1, pp. 71–88, Feb. 2000.

[17] A. Castro-Hernandez, K. Swigger, M. P. Ponce-Flores, and J. D.´ Tern-Villanueva, "Measures for Predicting Task Cohesion in a Global Collaborative Learning Environment," in *Global Software Engineering Workshops (ICGSEW), 2016 IEEE 11th International Conference on*. IEEE, 2016, pp. 31–36.

[18] K. Swigger, M. Hoyt, F. C. Sere, V. Lopez, and F. N. Alpaslan, "The temporal communication behaviors of global software development student teams," *Computers in Human Behavior*, vol. 28, no. 2, pp. 384–392, Mar. 2012. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0747563211002160

[19] D. D. Curtis and M. J. Lawson, "Exploring collaborative online learning," *Journal of Asynchronous learning networks*, vol. 5, no. 1, pp. 21–34, 2001.

[20] M. R. Hoyt, "Automatic Tagging of Communication Data," Ph.D. dissertation, University of North Texas, Denton Texas, Aug. 2012. [Online]. Available: http://digital.library.unt.edu/ark:/67531/metadc149611/

[21] J. Pennebaker, *The Secret Life of Pronouns: What Our Words Say About Us*. Bloomsbury USA, 2013.

[22] D. A. Newman and H.-P. Sin, "How Do Missing Data Bias Estimates of Within-Group Agreement? Sensitivity of SD WG, CVWG, rWG(J), rWG(J) * , and ICC to Systematic Nonresponse," *Organizational Research Methods*, vol. 12, no. 1, pp. 113–147, Jan. 2009.