

Research Report
ETS RR-17-59

Network Analysis of Conversation Data for Engineering Professional Skills Assessment

Mengxiao Zhu

Mo Zhang

December 2017

ETS Research Report Series

EIGNOR EXECUTIVE EDITOR

James Carlson
Principal Psychometrician

ASSOCIATE EDITORS

Beata Beigman Klebanov
Senior Research Scientist

Heather Buzick
Senior Research Scientist

Brent Bridgeman
Distinguished Presidential Appointee

Keelan Evanini
Research Director

Marna Golub-Smith
Principal Psychometrician

Shelby Haberman
Distinguished Presidential Appointee

Anastassia Loukina
Research Scientist

John Mazzeo
Distinguished Presidential Appointee

Donald Powers
Principal Research Scientist

Gautam Puhan
Principal Psychometrician

John Sabatini
Managing Principal Research Scientist

Elizabeth Stone
Research Scientist

Rebecca Zwick
Distinguished Presidential Appointee

PRODUCTION EDITORS

Kim Fryer
Manager, Editing Services

Ayleen Gontz
Senior Editor

Since its 1947 founding, ETS has conducted and disseminated scientific research to support its products and services, and to advance the measurement and education fields. In keeping with these goals, ETS is committed to making its research freely available to the professional community and to the general public. Published accounts of ETS research, including papers in the ETS Research Report series, undergo a formal peer-review process by ETS staff to ensure that they meet established scientific and professional standards. All such ETS-conducted peer reviews are in addition to any reviews that outside organizations may provide as part of their own publication processes. Peer review notwithstanding, the positions expressed in the ETS Research Report series and other published accounts of ETS research are those of the authors and not necessarily those of the Officers and Trustees of Educational Testing Service.

The Daniel Eignor Editorship is named in honor of Dr. Daniel R. Eignor, who from 2001 until 2011 served the Research and Development division as Editor for the ETS Research Report series. The Eignor Editorship has been created to recognize the pivotal leadership role that Dr. Eignor played in the research publication process at ETS.

RESEARCH REPORT

Network Analysis of Conversation Data for Engineering Professional Skills Assessment

Mengxiao Zhu & Mo Zhang

Educational Testing Service, Princeton, NJ

In this paper, we examine the student group discussion processes in a scenario-based assessment of engineering professional skills called Engineering Professional Skills Assessment (EPSA). In the assessment, the students were evaluated through a discussion on a scenario related to an engineering problem with no clear-cut solution. We applied various social network analysis (SNA) techniques and constructed 2 types of networks to model the patterns of the communication process in the student discussion and the connections among engineering professional skills (EPS) during the discussion process. We found that some network statistics were statistically related to team scores (or performance), such as the density of the EPS networks. We also investigated the temporal pattern of the discussion process by comparing the network statistics across 3 different time points. Results showed that the middle and end of the discussion processes differed significantly on 2 network measures: communication density and EPS connectedness. Results of the study have implications for instruction and scoring of such performance tasks. Limitations and next steps are also discussed.

Keywords Collaborative problem solving; engineering education; group discussion; network analysis

doi:10.1002/ets2.12191

In the 21st century global working environment graduating engineering students are expected to possess not only solid technical knowledge and skills but also strong professional “soft” skills that enable them to collaborate effectively with others in any problem-solving and teamwork situation (American Association of Colleges and Universities, 2013; American Society for Engineering Education & National Science, 2013; Seat, Parsons, & Poppen, 2001; Sheppard, Macatangay, Colby, Sullivan, & Shulman, 2008). Among the important engineering professional skills (EPS) considered by employers are collaborating with coworkers applying technical knowledge in real-world scenarios making ethical judgments and having the ability for continued learning. A recent survey (Passow, 2012) of more than 2,000 alumni from 11 engineering disciplines showed that teamwork communication data analysis and problem-solving skills are the most sought-after skills in engineering professional practices at workplaces. These EPS are all represented in the Accreditation Board for Engineering Technology (ABET) Criterion 3 Student Outcomes (ABET, 2011)

In the present study, we examined communication processes and integration of EPS during student group discussions. Traditionally, assessments of communication and other soft skills are conducted holistically, meaning that teachers assign a final score based on their overall evaluation of the student performance represented through the whole discussion session. In contrast to using holistic scores to describe students’ performance or proficiency, this study applied a network analysis approach for analyzing the communication process among students and the connections among different EPS. Analyses conducted on data at different time points of a discussion session (e.g., first half vs. second half) can further help identify temporal patterns of the problem-solving process.

This study is built upon Zhu and Zhang (2016), in which the authors found that the information captured by social network analysis (SNA) techniques provided useful evidence of group performance. In that investigation, the authors observed drastic differences in both communication and EPS networks using two student group discussions. For the communication networks, students in the high-performing group had more balanced and more frequent conversations among themselves with, on average, shorter sentences in their statements. The low-performing group tended to show less balanced and less frequent conversations among the students, though with relatively longer sentences among subgroups of students rather than all students in the discussion process. The authors concluded that SNA techniques have a great potential to uncover patterns in the collaborative problem-solving processes that are not obvious from traditional holistic ratings of student performance.

Corresponding author: M. Zhu, E-mail: mzhu@ets.org

Using a larger data set than in the previous study, we explored the following two research questions.

Research Question 1: What is the relationship between the student team scores and their network statistics?

Research Question 2: How do the network statistics change longitudinally over the course of the assessment?

Engineering Professional Skills Assessment

Our assessment instrument was the Engineering Professional Skills Assessment (EPSA), which has been developed under the support of the National Science Foundation. EPSA is conceptualized to be a direct measure of five skills identified in the ABET Criterion 3 Student Outcome (Ater Kranov et al., 2011, 2013; Ater Kranov, Hauser, Olsen, & Girardeau, 2008). These criteria as developed by ABET are intended to gauge the readiness of the graduating engineering students and are used in the accreditation process for colleges of engineering. The validity and psychometric properties of EPSA have been examined in previous research, such as Ater Kranov et al. (2011) and Zhang et al. (2015).

EPSA is a scenario-based performance assessment. Each EPSA consists of a scenario describing an engineering problem that does not have a clear solution and a scoring rubric containing five aspects. The unit of analysis is a group of four to six students. In the assessment, students are instructed to engage in a 45-minute discussion around the issue given in the scenario.

The EPSA scenarios were written to meet several parameters to ensure their comparability with one another and to maximize students' engagement in meaningful discussions (McCormack, Beyerlein, Ater Kranov, Pedrow, & Schmeckpeper, 2014). As for the scoring rubric, it has five skill dimensions corresponding to the skills 3f to 3j in the ABET Criterion 3 Student Outcome. Further, three of the five dimensions have subdimensions to assess more fine-grained aspects of the skill. Each dimension is graded on a scale of integers, 0 through 5: 0 (*missing*), 1 (*emerging*), 2 (*developing*), 3 (*practicing*), 4 (*maturing*), and 5 (*mastering*). Both the complete version and an abbreviated one-page version of the EPSA scoring rubric are discussed in Schmeckpeper, Kelley, and Beyerlein (2014). Here, we provide high-level descriptions of the dimensions and their subdimensions. The coding scheme described in the following section was developed specific to these dimensions.

- 3f. Understanding of professional and ethical responsibility (three subdimensions): Evaluates the extent to which the group (a) shows perspective on stakeholders, (b) identifies and frames the problems, and (c) attends to ethical considerations.
- 3g. Ability to communicate effectively: Evaluates the level and productiveness of group interaction and group self-regulation with respect to achieving consensus among members.
- 3h. Broad understanding of the impact of engineering solutions in global, economic, environmental, and cultural/societal contexts: Evaluates the extent to which the group examines and weighs its proposed solutions with respect to the impact on major relevant contexts, and justifies those possible solutions with reasonable accuracy.
- 3i. Recognition of the need for and ability to engage in lifelong learning (two subdimensions): Evaluates the extent to which the group (a) utilizes past personal experiences or historical events to inform analysis, check assumptions, and examine the credibility of the sources, and (b) identifies the boundaries of its knowledge related to the issue and discusses how those limitations affect its analysis of the issue.
- 3j. Knowledge of contemporary issues (two subdimensions): Evaluates the extent to which the group (a) gives consideration to contemporary political and geopolitical issues, and (b) considers modern methods, technologies, and tools in solving the problems.

Related Work

Social Network Analysis in Education

As a research method first initiated in the social sciences, SNA is often used to model the connections among human beings (Wasserman & Faust, 1994). The nodes are individuals, and the links are various relations among these individuals, such as friendship, work, and marriage (Guimera, Uzzi, Spiro, & Amaral, 2005; Padgett & Ansell, 1993). SNA also extends beyond modeling humans to model complex connections in other fields outside of social science, including computer science, information science, and biology (Newman, 2003). The nodes analyzed in these complex networks may not be

human beings, but computer routers, websites, or neurons, and the links would represent the wiring among the routers, the hyperlinks between websites, and the connections among neurons. Related studies have covered topics ranging from the topology and structural features of the networks (e.g., Newman, 2003) to the relationship between network features and nodal attributes such as individual performance in working teams (e.g., Ahuja, Galletta, & Carley, 2003).

In education, SNA is used to model both human systems and systems of nonhuman entities. The human systems in education can be either student interactions in the classroom and the online environment (Grunspan, Wiggins, & Goodreau, 2014; Ruane & Koku, 2014) or friendship and peer influence among students (Snijders & Baerveldt, 2003; Valente, Fujimoto, Unger, Soto, & Meeker, 2013). In such systems, nodes represent students, and links represent interactions, friendship, or peer influence among students. In addition to connections among students, researchers have also used SNA to model the interactions among teachers, teaching coaches, and administrators to understand, for example, the impacts of interventions on school systems (Sweet, Thomas, & Junker, 2013). SNA has also been used to model nonhuman systems in education, such as the transitions among student activities recorded in the log data for simulation-based assessment (Zhu, Shu, & von Davier, 2014) and students' eye movements while they were solving mathematics problems in scenario-based assessment (Zhu & Feng, 2015). In both studies, network statistics were found to be closely related to student performance.

Analysis of Conversation Data

Research on conversation and discourse analysis can be conducted either by focusing on the linguistic features or by coding the text generated by students. At the low level, the linguistic features can be frequency of words and collocation analysis. Semantic analysis and other natural language processing (NLP) methods can also be applied to generate more complicated linguistic features (Dönmez, Rosé, Stegmann, Weinberger, & Fischer, 2005; Gee, 2013). For instance, Crossley et al. (2015) studied the NLP features of the student-generated text in the forum related to student retention in massive open online courses (MOOCs). Another approach is to code the conversation data, either automatically or manually, using external categories to capture the conversation process or other features of interest. For instance, Liu and Burn (2007) used the TEMPO coding system (Futoran, Kelly, & McGrath, 1989) to capture the problem-solving process and the communication patterns in team discourse data and studied how these features are related to team performance.

In this study, we coded the team conversation data from engineering education, with a focus on the communication processes among the students and the connections among the assessed engineering professional skills. The former networks modeled student interactions, and the latter modeled the interactions among nonhuman entities. Through the network analysis, we explored how network statistics are related to the performance of the teams being assessed and how network statistics change over time.

Method

Participants

We used a sample of 33 student groups consisting of more than 130 individual students. The data were collected from three engineering colleges in the United States. All participants were undergraduate students majoring in various disciplines of engineering, including electrical engineering, civil engineering, mechanical engineering, and computer science. A total of six EPSA scenarios were used, and each student group was given one of six scenarios. Short descriptions of the scenarios are provided in Table 1. Note that the descriptors in no way capture the complexity of the issue that the EPSA scenario text provides; rather, they give a very high-level idea of the issue presented in the scenario.

All the group discussions were audio-recorded and professionally transcribed. Students' names were anonymized to protect their privacy. Each student was assigned an ID such as A, B, or C in each group transcript, which allowed us to conduct discourse and network analyses. A team of three experienced engineering faculty and one researcher in engineering education graded each student group performance using the EPSA rubric. Each transcript was evaluated by at least two graders, with each grader providing a score on each EPSA dimension.

Table 1 Descriptions of Engineering Professional Skills Assessment Scenarios

Scenario	Description
Energy critical elements	Discuss the supply and demand situation of rare earth materials.
Lithium mining	Discuss lithium mining to supply Li-ion electrical vehicle batteries.
Offshore wind farms	Discuss the development of offshore wind resources.
Oil spill	Discuss the environmental and social impacts of an oil spill and its cleanup.
Power grid vulnerabilities	Discuss security issues related to U.S. electric power grids.
Coal ash spill	Discuss the impacts and solution for the toxic coal ash spill in Tennessee.

Student A: *So like the only two issues that was talked about were the dredging and the oil spill, and I just don't see dredging being near as big of a problem as this massive oil spill. It seems like—I don't know, they're just like playing that up so that BP doesn't really have to fix the problem to the full amount that they should.*

Student B: *Which is the oil mats that were left undisturbed by the cleanup crews since the dredging would amplify it. So essentially what they're doing is I'm guessing like the oil is just down there, kind of sitting. Is that what that means?*

Student C: *Yeah.*

Student A: *Where's that?*

Student B: *This last sentence. And so I don't know, I think that they—that would be a good place for the pump, because I kind of see what they're going—they don't want to, you know, dredge through and . . .*

Figure 1 An example of who-is-talking-to-whom in discussion.

Coding of the Transcripts

Coding of the transcripts included two steps. The first step was to identify who was talking to whom at different time points of the discussion process. The second step was to identify whether one or more EPSs were reflected at any time point of the discussion. To this end, we coded each transcript at the utterance level, in which an utterance is defined as a statement made by a student after the previous student finishes talking and before the next student talks. An utterance can be a word (e.g., *okay*), a phrase (e.g., *to some extent*), a sentence (e.g., *Yeah, that's true*), or a paragraph.

For the first step of the coding, each utterance was coded such that the initiator of that utterance was identified as well as the target audience (who the statement is intended for). Given the nature of the group discussion, the majority of the utterances were intended for everyone (indicated by *All* in our coding system). It was nonetheless not unusual for a student to speak to another specific student. In addition to students, a moderator was part of each discussion group; though for all transcripts, the moderators intervened only minimally to avoid interfering with the students' discussion. Using this who-was-talking-to-whom information, we were able not only to construct an overall communication network for each group discussion, but also to examine the patterns of those interactions among group members in the discussion process.

Figure 1 shows an excerpt of a transcript (from the oil spill scenario), which consists of five utterances made by three students. In this example, Student A was coded to be talking to everyone in the team. Student B responded to Student A's comments by adding some information. Student C briefly responded to Student B. Student A asked a follow-up question of Student C. Student B then replied with more information. For this short example, the corresponding coding of the communication pattern is given in Table 2.

For the second step of the coding, we looked for evidence of whether one or more EPSs were reflected in the utterances. During the coding, the EPSA rubrics were used as the guideline. In Figure 2, we provide two examples of utterances that exhibit EPSs. Example 1 is the case in which we identified evidence of the skill of *ProblemID* (3f), and Example 2 is the case in which we identified two skills: *Impact* (3 h) and *NonTechIssue* (3j).

The coding of the transcripts was carried out by four researchers who have education and linguistic backgrounds. The most senior and experienced researcher in the coding team provided adjudications to all observed disagreements. We report results based on the final adjudicated codes.

Table 2 Coding of Who-Is-Talking-to-Whom in the Example

Who	ToWhom
Student A	All
Student B	Student A
Student C	Student B
Student A	Student C
Student B	Student A

Example 1

Student: *That's kind of the problem that I would have with the chemical treatment that they proposed, was that they didn't know if it was going to hurt worse than the oil, and so that fact . . .*

Example 2

Student: *I think if somebody did like shine the spotlight on them a little bit more, like saying, "This is still a big issue" and putting out there to the public that they would probably be more inclined to fix it and come up with new solutions and all of that.*

Figure 2 Examples of utterances.

Constructing Communication and Engineering Professional Skills Networks

The construction of the communication networks is intuitive. The nodes in the networks are the students in each group. Directed links are formed, in which the arrow suggests the direction from the speaker to the target. Each link is assigned a weight, which equals to the frequency of the communication between the two nodes. Besides *Students*, there are two additional nodes: *All* representing everyone in a group, and the *Moderator*. We intentionally kept *All* in the network instead of deleting it or separating it into links to all individual student nodes. There were two reasons for making this decision. First, removing *All* would result in a loss of a substantial amount of information, in that ability to address the group is an important aspect of communication skill. Second, compared to talking to a specific person, talking to the whole group entails a different type of information sharing and should not be confounded with the former. For instance, a group with members talking only to each other individually or in subgroups is different from another group in which its members are addressing everyone else. Removing the *All* node would make these two phenomena indistinguishable.

To construct the EPS networks, we drew ideas from studies on epistemic network analysis (e.g., Shaffer et al., 2009). Epistemic network analysis is an application of SNA in understanding and assessing learning. Epistemic network analysis is based on an epistemic frame containing skills, knowledge, identity, values, and epistemology. Relevant to this study, we focus on the skills. In the EPS networks, the nodes are the nine EPSs (or subdimensions) as introduced in the coding section. The links represent conceptual connections among the EPSs, which we consider as coappearance in the same utterance. Of note is that the links are undirected and connect two or more EPSs only if they coappear in the same utterance. The links are also weighted by the frequency of the observed connections. The resulting networks can, therefore, in some sense inform on the extent to which students master the skills as a system instead of as isolated cases.

Using the coded transcripts, we constructed the overall communication and EPS networks for each student group discussion. To examine how communication patterns among students and connections among EPSs changed over the course of the discussion, we evenly divided each student discussion process into three time periods using the total number of utterances. The first third of the utterances in a transcript was treated as the beginning of the discussion process (*T1*). By the same token, the next third was considered as the middle, or *T2*, and the last third was the end, or *T3*, of the discussion process. We constructed the communication and EPS networks separately for each time period for each transcript and compared the patterns across these three time periods on the various measures described in the next section.

Variables

Team Scores and Team Size

In our data set, each transcript was evaluated by at least two graders, and each grader provided a score on each EPSA dimension. Team scores are calculated by averaging the scores provided by the graders over all EPSA dimensions. *Team size* is the number of students participating in each team.

Network Measures for Communication Networks

By definition, communication networks are directed and weighted. The links are from the initiator of the utterance to the target of the utterance, and the weight of the links indicates the frequency of conversation between two students. Formally, we defined the communication network C as a set of nodes $S = \{s_1, s_2, \dots, s_n\}$, a set of links $\mathcal{L} = \{l_{1,2}, \dots, l_{ij}, \dots, l_{n-1,n}\}$ where $l_{ij} \in \{0, 1\}$, and a set of link weights $\mathcal{W} = \{w_1, w_2, \dots, w_m\}$. We considered the following four network measures for communication networks.

Comm density. The density of the communication network captures the average weight of the links and is calculated as the sum of the link weights divided by the number of possible links in a directed network of the same number of nodes. For a directed network with n nodes, there are $n * (n - 1)$ possible links. Thus, comm density $D_c = \frac{\sum_{i=1}^n \sum_{j=1, j \neq i}^n l_{ij} w_{ij}}{n * (n - 1)}$. This network measure captures how active the teams are in taking turns during the discussion. A higher number indicates more iterations of conversation.

Comm reciprocity. Reciprocity captures the network structure at the dyadic level. For each pair of nodes a and b , the tie between them is called reciprocal if both link l_{ab} and link l_{ba} exist. The reciprocity of a directed network is calculated as the number of mutual dyads divided by the total number of nonnull dyads (Wasserman & Faust, 1994), $R_c = \frac{\sum_{i=1}^n \sum_{j=1, j \neq i}^n l_{ij} l_{ji}}{\sum_{i=1}^n \sum_{j=1, j \neq i}^n l_{ij} |l_{ji}|}$. Comm reciprocity captures the extent to which students in a team talk to each other during the discussion. A higher number indicates that a higher percentage of student pairs talked to each other.

Comm transitivity. Transitivity is a network measure that captures the triadic level structures (Wasserman & Faust, 1994). A set of three nodes $\{a, b, c\}$ is considered, and if the links exist among nodes ab , bc , and ac , it is called a transitive triad. The transitivity measure for the network is calculated as the total number of transitive triads divided by the number of triads with at least two of the three needed links. A similar measure to capture the closure of triangles is a clustering coefficient for undirected networks (Watts & Strogatz, 1998). Comm transitivity captures the extent to which three individuals communicate among all members in the transitive sense. Previous research has suggested an association between high transitivity and more effective communication (Holland & Leinhardt, 1971).

Comm outdegree centralization. The outdegree of nodes in the communication network captures the frequency of utterances for each student, which is also called *outdegree centrality*. At the network level, the outdegree centralization measures the variance of the outdegree over all nodes in the network. A higher value indicates a higher variance in the outdegree. This study adopted Freeman's (1979) definition of centralization. Comm outdegree centralization captures how balanced students are in terms of actively participating in the discussion. A higher value indicates that students are not equally active during the discussion.

Network Measures for Engineering Professional Skills Networks

EPS networks are undirected and weighted, and the links among different EPS are coappearance of these skills in the same utterance. Formally, we define the EPS G as the node set $N = \{n_1, n_2, \dots, n_9\}$, the link set $E = \{e_{1,2}, \dots, e_{ij}, \dots, e_{8,9}\}$ where $e_{ij} \in \{0, 1\}$, and the weight set for the links $\mathcal{W} = \{w_1, w_2, \dots, w_m\}$.

We considered the following three network measures for EPS networks.

EPS density. Similar to the definition of the density for the communication network, the density for the EPS network is defined as the sum of the link weights divided by the number of possible links. For an undirected network with the same number of nodes, n , there are $n*(n-1)/2$ possible links. And, EPS density $D_{eps} = \frac{2*\sum_{i=1}^m w_m}{n*(n-1)/2}$. This network measure captures how EPSs are closely connected with each other as they cooccur in the same utterance.

EPS degree centralization. Because EPS networks are undirected networks, node degrees are not directional. The degree of a node captures the extent to which that skill coappears with other skills in the same utterance. For the EPS network, degree centralization measures how the degrees of different skills vary. A higher value indicates higher variance of the degree. Again, similar to the outdegree centralization for the communication network, we followed Freeman's (1979) definition of centralization.

EPS connectedness. To measure how well all EPSs are connected to each other, we adopted Krackhardt's definition of connectedness (1994). It is defined as the fraction of all dyads such that there is an undirected path between the two nodes in each dyad. A higher value indicates that a higher fraction of nodes are connected, and a lower value indicates a higher fraction of isolated nodes in the network.

Results

Coding Agreement

Consistent with the results in Zhu and Zhang (2016), the interrater agreements for coding were high for all EPSs except *communication* (3 g). This aspect posed the most challenge to the coding team, with exact percentage agreement values ranging from .46 to .82 across all transcripts. The remaining EPSs achieved reasonably high intercoder agreements. The average exact percentage agreements for the other EPSs were .73 (*problem identification*), .88 (*stakeholder identification*), .94 (*ethics*), .92 (*impact and context*), .96 (*scrutinize information*), .96 (*knowledge status*), .96 (*nontechnical issue*), and .98 (*technical issue*). Also worth noting is that, even though the *To Whom* category in practice appeared to be a difficult aspect for the coders to agree on, it achieved an average exact percentage agreement of .81 across all the transcripts.

Descriptive Statistics of Variables and Network Visualization

Table 3 shows the mean and the standard deviation for each variable and the correlations between variables. Even though different network measures capture different network statistics, they can be highly correlated to each other. For instance, the reciprocity of the communication network is highly correlated with both the transitivity and the outdegree centralization of the network.

We note that comm reciprocity, comm transitivity, EPS density, and EPS centralization showed the highest associations with team scores, with correlation coefficients of 0.35, 0.41, 0.37, and -0.35 , respectively. This finding is expected, according to Zhu and Zhang (2016), and is further discussed in the regression analyses in the next section. As examples, we present the overall communication and EPS networks for two selected, typical high- and low-scoring groups (Figures 3 and 4). The high-scoring group had five students, and the low-scoring group had four. The performance statistics of the two student groups are given in Table 4.

As can be seen in Figure 3, in general the high-scoring group showed more communication among team members than the low-scoring group, and all team members participated in the discussion despite some talking more (e.g., Students 2 and 4) than others. In contrast, the low-scoring group had Students 1 and 2 dominating the discussion most of the time, whereas the other three students did not participate in the discussion much at all.

Figure 4 also shows the typical contrasts between low- and high-scoring groups in the EPS networks. For the high-scoring group, the links between EPSs are generally thicker, and more connections are observed among different EPSs. The low-scoring group generally lacked connections among EPSs and exhibited less evidence of EPSs in their group discussion process.

Table 3 Descriptive Statistics and Correlations of Variables

Statistics	Mean	SD	1	2	3	4	5	6	7	8
0. Team score	2.09	0.55	-0.18	0.17	0.35	0.41	0.02	0.37	-0.35	0.31
1. Team size	6.27	0.67		-0.47	-0.01	-0.25	-0.37	0.02	0.28	-0.07
2. Comm density	8.66	4.74			0.43	0.55	0.77	0.22	-0.10	0.40
3. Comm reciprocity	0.61	0.11				0.84	0.31	0.18	0.27	0.38
4. Comm transitivity	0.88	0.19					0.41	0.32	0.05	0.42
5. Comm outdegree centralization	9.99	4.75						0.40	-0.08	0.47
6. EPS density	3.87	2.19							-0.44	0.44
7. EPS centralization	0.37	0.10								-0.04
8. EPS connectedness	0.86	0.16								

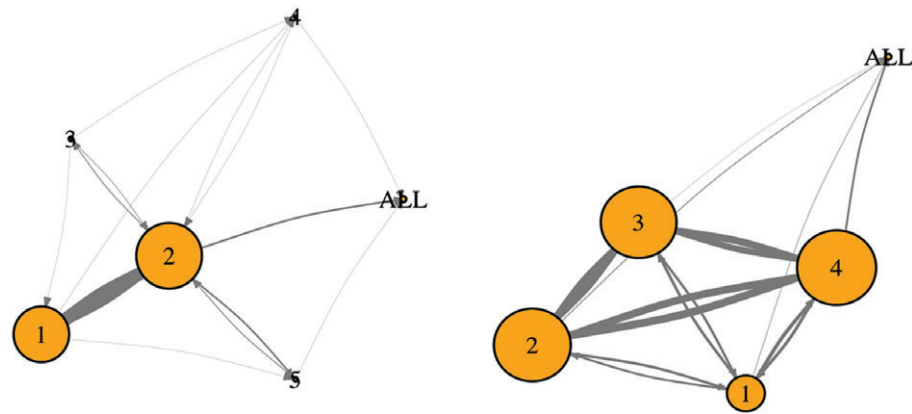


Figure 3 Overall communication networks. *Note.* Left panel: low-scoring group; right panel: high-scoring group.

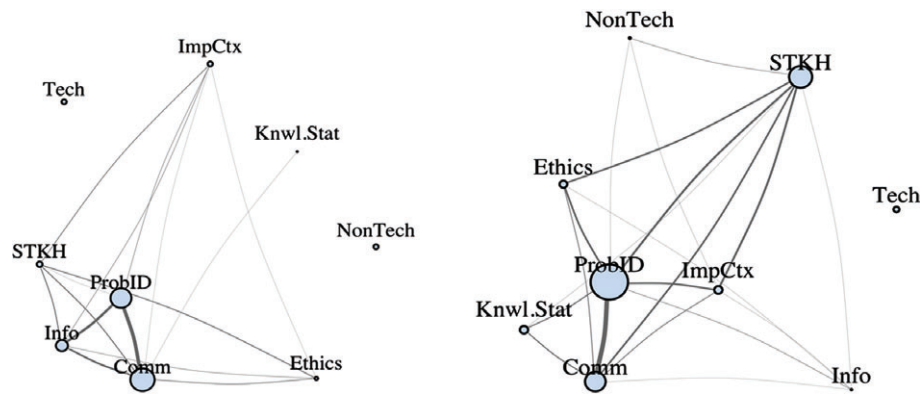


Figure 4 Overall engineering professional skills networks. *Note.* Left panel: low-scoring group; right panel: high-scoring group.

Network Statistics and Team Performance

In order to find out how network statistics for both the communication networks and the EPS networks are related to the performance of the teams, we ran ordinary least-squares regression analysis with the team scores as the dependent variable and all other variables as the independent variables. To build the regression models, we adopted a hierarchical analysis method (Aiken & West, 1991) and entered variables into the multiple regression models in blocks. This method enabled us to evaluate the effects of only the control variable, the team size (Model 1), the densities of both networks (Model 2), and other network statistics (Model 3). While building the models, we found that some network statistics that are highly correlated with each other cannot be included in the same model. Thus, we decided to include only comm

Table 4 Engineering Professional Skills Assessment Scores for the Two Selected Groups

Group	Engineering Professional Skills Assessment skill					Overall score
	3f	3g	3h	3i	3j	
Low scoring	1.4	1.2	1	1	1	1.12
High scoring	2	3.5	3	3	2	2.7

Table 5 Regression Analysis Results

Effect	Model 1 estimate (SE)	Model 2 estimate (SE)	Model 3 estimate (SE)
(constant)	2.99 (0.91)	2.65 (1.09)	1.98 (1.06)
Team size	-0.14 (0.14)	-0.15 (0.16)	-0.14 (0.16)
Comm density		0.001 (0.02)	-0.03 (0.02)
Comm reciprocity			2.36 (0.95)*
EPS density		0.09 (0.04)*	0.03 (0.05)
EPS centralization			-2.14 (1.07)†
EPS connectedness			0.51 (0.67)
R ²	0.03	0.17	0.38

Note. EPS = engineering professional skills.

* $p \leq .05$. † $p \leq .1$.

reciprocity, because comm transitivity and comm outdegree centralization were both found to have effects similar to comm reciprocity.

The results are shown in Table 5. Model 1 included only team size as the control variable, and the results showed that the sizes of the teams were not related to the team scores. Also included in Table 1 are the R -squared values of the regression models.

In Model 2, we included two network variables, the densities for the communication networks and the densities for the EPS networks. The results showed that the densities of the communication networks were not related to the team scores. Because this variable captures the frequencies of the turn-taking during the conversation, the results indicated that teams that had more iterations of conversation did not necessarily have higher scores. However, the densities of the EPS networks were found to be positively related to team scores. Higher values for EPS density indicated higher frequency of cooccurrence of different EPSs during the conversation. Together, the results suggest that it is the content of the conversation that matters and not the frequency of the communication.

Next, in Model 3, three more network variables were included. For the communication network, we included the reciprocity measure, which captured the extent to which the conversations were mutual between team members. For the EPS networks, we considered the network degree centralization, which captured the variance of the node degrees for the skills, and network connectedness, which captured how well the EPSs were connected. The results showed that two of these three network variables were significantly related to the team scores, with EPS centralization being marginally significant at the $p < 0.1$ level. Comm reciprocity was positively related to team scores, indicating that teams with more of their members talking with each other tended to receive higher scores. This result can be partly explained by the fact that “3 g. Ability to communicate effectively” is one of the EPS dimensions. However, combined with the nonsignificant results for the density of the communication, it seems that effective communication was less represented by having more frequent conversations among the team members but more by mutual information exchange among them.

The second network variable entered into Model 3 was EPS centralization, which was negatively related to team scores and was marginally significant. This result indicated that teams that exhibited more balanced and connected EPSs in their discussion process received higher performance scores. This result is understandable because the final scores reflected the overall EPS levels while considering all dimensions. The third network variable was EPS connectedness, and it was not significantly related to the team scores. This indicates that how well the EPSs were connected was not well captured by the final scores.

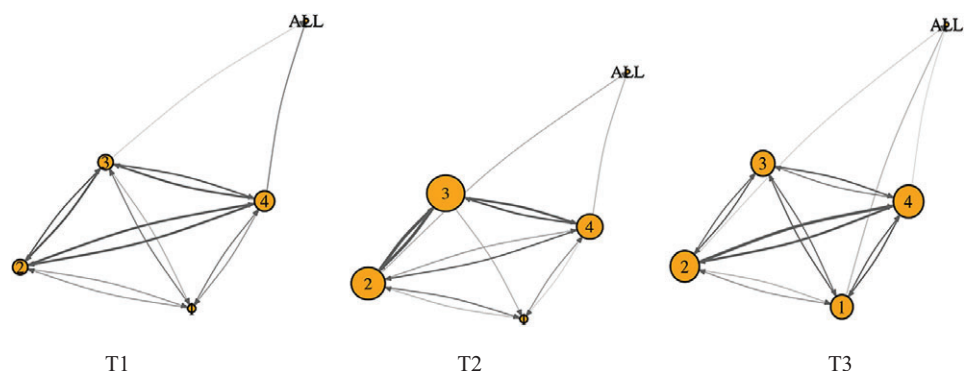


Figure 5 Communication networks over time.

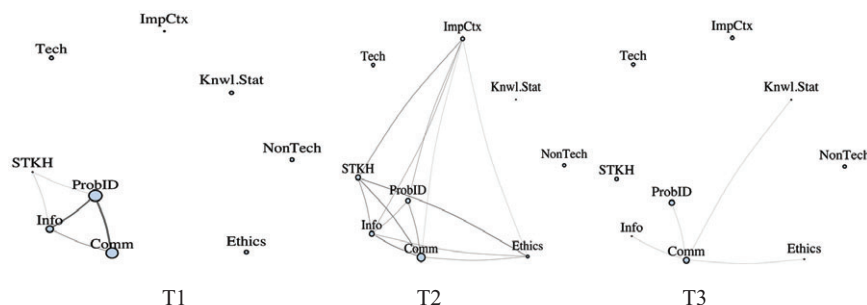


Figure 6 Engineering professional skills networks over time.

Network Statistics Over Time

To study the longitudinal trend of the network measures over the assessment session, we compared the values of the network variables included in the regression models discussed in the previous section. In our data set, there were three time periods: to capture the beginning, the middle, and the end of the conversations, labeled as T1, T2, and T3. We built the network separately for each of the time periods for all groups. As an illustration, Figures 5 and 6 show examples of the communication networks and the EPS networks for one group across the three different time points. To compare the measures over three time points, we first ran repeated measure ANOVA on the network measures over time. We then ran paired t -tests to compare the network measures in adjacent time periods.

For the communication networks, the repeated measures ANOVA results showed that time had a statistically significant effect on the network density, $F(2, 64) = 12.57, p < .001$. The paired t -tests result showed that comm density in T1 ($M = 2.90, SD = 1.59, N = 33$) was not significantly different from comm density in T2 ($M = 2.89, SD = 1.59, N = 33$), $t(32) = 1.58, p = .12$. However, comm density in T2 was significantly higher than comm density in T3 ($M = 2.86, SD = 1.57, N = 33$), $t(32) = 3.77, p < 0.05$. No significant time effects were observed for comm reciprocity, $F(2, 64) = 0.31, p = .73$. The results from the paired t -tests confirmed the finding. Comm reciprocity in T1 ($M = 0.58, SD = 0.15, N = 33$) was not significantly different from comm reciprocity in T2 ($M = 0.60, SD = 0.17, N = 33$), $t(32) = -0.55, p = .48$. Comm reciprocity in T2 was also not significantly different from comm reciprocity in T3 ($M = 0.58, SD = 0.15, N = 33$), $t(32) = 0.96, p = .34$.

For the EPS networks, no significant time effects were observed for the density measure, $F(2, 64) = 1.32, p = .27$. Again, the results from the paired t -tests were consistent with the repeated measure ANOVA. EPS density in T1 ($M = 1.28, SD = 0.97, N = 33$) was not significantly different from EPS density in T2 ($M = 1.40, SD = 0.85, N = 33$), $t(32) = -1.27, p = .21$. And EPS density in T2 was not significantly different from EPS density in T3 ($M = 1.19, SD = 0.70, N = 33$), $t(32) = 1.50, p = .14$.

We also did not observe significant time effects on EPS centralization, $F(2, 64) = 0.34, p = .71$. EPS centralization in T1 ($M = 0.41, SD = 0.08, N = 33$) was not significantly different from EPS centralization in T2 ($M = 0.40, SD = 0.10, N = 33$),

$t(32) = 0.09, p = .93$. EPS centralization in T2 was also not significantly different from EPS centralization in T3 ($M = 0.39, SD = 0.11, N = 33$), $t(32) = 0.61, p = .55$.

For EPS connectedness, we observed that time had marginal significant effect, $F(2, 64) = 2.45, p = 0.09$. The paired t -tests showed that the significant difference was observed again for the last two time periods. EPS connectedness in T1 ($M = 0.59, SD = 0.27, N = 33$) and EPS connectedness in T2 ($M = 0.66, SD = 0.23, N = 33$) were not significantly different from each other, $t(32) = -1.33, p = .19$. However, EPS connectedness in T2 was significantly higher than EPS connectedness in T3 ($M = 0.55, SD = 0.25, N = 33$), $t(32) = 2.36, p < .05$.

To sum up, significant time effects were observed on two network measures. The first one was the density of the communication networks between the second and the third periods. The density decreased during the end of the conversation session. The second network measure that significantly changed over time was the connectedness for EPS networks. We also observed a decrease for connectedness from the second to the third time periods. One potential explanation is that toward the end of the assessment, it is more likely that the students had already reached agreements on the discussion items and were working more toward summarizing and concluding the discussions. Thus, the frequency of turn-taking among team members decreased a little, and so did the cooccurrence of different EPSs.

Discussion and Future Directions

In this study, we used SNA techniques to better understand the patterns of the communication process and connections of EPSs during student group discussions. Compared to a previous study by Zhu and Zhang (2016), we used a much larger sample of student groups to investigate two research questions.

For the first question about the relationship between network statistics and team performance, we found that team size was not statistically related to team scores, meaning that a larger discussion team does not necessarily perform better on the EPSA assessment. The EPS network's densities were found to be positively and significantly related to team scores, suggesting that higher performing student groups exhibited more cooccurrence of different EPSs during the conversation. With regard to the communication network, the extent to which the conversations were mutual among team members is also a significant predictor of team scores, as one might expect. This result is also consistent with findings in the previous study, in which members of a higher performing student group generally had more balanced and more frequent conversations among themselves, and a lower performing group tended to show less balanced and less frequent conversations among the students.

As for the second question, we examined whether and how the network statistics changed longitudinally over time across the discussion. We found that most of the communication and EPS network patterns were not statistically different across time points, except for the communication density and EPS connectedness between T2 and T3. The results suggest that among the three time periods as defined in this study, the first two were consistent, and toward the end of the conversation, students tended to participate less actively. Our speculations were that students were close to reaching an agreement on the discussion question, thus being less involved and less active. Further analysis on the content of the conversation data may provide evidence to explain the observed phenomena.

This study also has some limitations, most of which can lead to important future directions. For instance, the current data set included only 33 teams, even though there were more than 130 students participating in these teams, and each team had a discussion session of around 45 minutes. More teams can increase the reliability and the generalizability of current findings. On the other hand, the current analysis adopted the coding method of analyzing the conversation data generated during the assessment. Another important direction is to analyze the content of the conversations using NLP methods. For instance, one potential direction is to explore the relationships between the individual words composing each conversation and the skill ratings provided by the expert human graders.

Acknowledgments

Our thanks go to coders at Educational Testing Service: Zhitong Yang, James Bruno, Chelsea Ezzo, and Adam Bacall. We thank Ashley Ater Kranov and the EPSA leadership team for providing access to the data. The data used in this study are supported by the National Science Foundation Grant No. DUE DRL 1432997. Any opinions expressed are those of the authors and do not necessarily reflect the views of the NSF.

References

- Accreditation Board for Engineering Technology. (2011). *Criteria for accrediting engineering programs*. Baltimore, MD: Author. Retrieved from http://www.abet.org/uploadedFiles/Accreditation/Accreditation_Process/Accreditation_Documents/Current/eac-criteria-2012-2013.pdf
- Ahuja, M. K., Galletta, D. F., & Carley, K. M. (2003). Individual centrality and performance in virtual R&D groups: An empirical study. *Management Science*, 49(1), 21–38. Retrieved from <http://www.jstor.org/stable/822582>
- Aiken, S., & West, L. (1991). *Multiple regression: Testing and interpreting interactions*. Newbury Park, CA: Sage.
- American Association of Colleges and Universities. (2013). *It takes more than a major: Employer priorities for college learning and student success*. Retrieved from <https://www.aacu.org/publications-research/periodicals/it-takes-more-major-employer-priorities-college-learning-and>
- American Society for Engineering Education & National Science. (2013, May). *Transforming undergraduate education in engineering. Phase I: Synthesizing and integrating industry perspectives*. Paper presented at the workshop of the American Society for Engineering Education, Arlington, VA. Retrieved from http://www.asee.org/TUEE_PhaseI_WorkshopReport.pdf
- Ater Kranov, A., Hauser, C., Olsen, R., & Girardeau, L. (2008). A direct method for teaching and assessing professional skills in engineering programs. In *Proceedings of the ASEE Annual Conference and Exposition* (pp. 13.32.1–13.32.200). Retrieved from <https://peer.asee.org/a-direct-method-for-teaching-and-assessing-professional-skills-in-engineering-programs.pdf>
- Ater Kranov, A., Williams, R. L., Pedrow, P. D., Schmeckpeper, E. R., Beyerlein, S., & McCormack, J. (2013). A direct method for teaching and measuring engineering professional skills: A validity study for the National Science Foundation's research in evaluation of engineering and science education (REESE). In *Proceedings of the ASEE International Forum* (pp. 21.2.1–21.2.18). Retrieved from <https://peer.asee.org/a-direct-method-for-teaching-and-measuring-engineering-professional-skills-a-validity-study-for-the-national-science-foundation-s-research-in-evaluation-of-engineering-and-science-education-reese.pdf>
- Ater Kranov, A., Zhang, M., Beyerlein, S. W., McCormack, J., Pedrow, P. D., & Schmeckpeper, E. R. (2011). A direct method for teaching and measuring engineering professional skills: A validity study. *Proceedings of the ASEE Annual Conference and Exposition* (pp. 22.38.1–22.38.21). Retrieved from <https://peer.asee.org/a-direct-method-for-teaching-and-measuring-engineering-professional-skills-a-validity-study.pdf>
- Crossley, S., Mcnamara, D. S., Baker, R., Wang, Y., Paquette, L., Barnes, T., & Bergner, Y. (2015). Language to completion: Success in an educational data mining massive open online class. In *Proceedings of the 8th International Conference on Educational Data Mining* (pp. 388–391). Retrieved from <https://www.upenn.edu/learninganalytics/ryanbaker/CrossleyCorrected.pdf>
- Dönmez, P., Rosé, C., Stegmann, K., Weinberger, A., & Fischer, F. (2005). Supporting CSCL with automatic corpus analysis technology. In T. Koschmann, D. Suthers, & T. W. Chan (Eds.), *Proceedings of the 2005 Conference on Computer Support for Collaborative Learning Learning 2005: The next 10 years!* (pp. 125–134). <https://doi.org/10.3115/1149293.1149310>
- Freeman, L. (1979). Centrality in social networks I: Conceptual clarification. *Social Networks*, 1, 215–239.
- Futoran, G. C., Kelly, J. R., & McGrath, J. E. (1989). TEMPO: A time-based system for analysis of group interaction process. *Basic and Applied Social Psychology*, 10(3), 211–232.
- Gee, J. P. (2013). *An introduction to discourse analysis: Theory and method*. London, England: Routledge.
- Grunspan, D. Z., Wiggins, B. L., & Goodreau, S. M. (2014). Understanding classrooms through social network analysis: A primer for social network analysis in education research. *Cell Biology Education*, 13(2), 167–179. <https://doi.org/10.1187/cbe.13-08-0162>
- Guimera, R., Uzzi, B., Spiro, J., & Amaral, L. A. N. (2005). Team assembly mechanisms determine collaboration network structure and team performance. *Science*, 308, 697–702.
- Holland, P. W., & Leinhardt, S. (1971). Transitivity in structural models of small groups. *Comparative Group Studies*, 2(2), 107–124.
- Krackhardt, D. (1994). Graph theoretical dimensions of informal organizations. In K. M. Carley & M. J. Prietula (Eds.), *Computational organization theory* (pp. 89–111). Hillsdale, NJ: Lawrence Erlbaum.
- Liu, Y. C., & Burn, J. M. (2007). Improving the performance of online learning teams: A discourse analysis. *Journal of Information Systems Education*, 18(3), 369–379.
- McCormack, J. P., Beyerlein, S. W., Ater Kranov, A., Pedrow, P. D., & Schmeckpeper, E. R. (2014). Scenario and scoring sheet development for engineering professional skill assessment. In *Proceedings of the ASEE Annual Conference and Exposition* (pp. 24.1070.1–24.1070.18). Retrieved from <https://peer.asee.org/scenario-and-scoring-sheet-development-for-engineering-professional-skill-assessment.pdf>
- Newman, M. E. J. (2003). The structure and function of complex networks. *SIAM Review*, 45, 167–256.
- Padgett, J. F., & Ansell, C. K. (1993). Robust action and the rise of the Medici. *American Journal of Sociology*, 98(6), 1400–1434.
- Passow, H. J. (2012). Which ABET competencies do engineering graduates find most important in their work? *Journal of Engineering Education*, 101, 95–118.
- Ruane, R., & Koku, E. F. (2014). Social network analysis of undergraduate education student interaction in online peer mentoring settings. *MERLOT Journal of Online Learning and Teaching*, 10(4), 577–589.

- Schmeckpeper, E. R., Kelley, M., & Beyerlein, S. (2014). Using the EPSA rubric to evaluate student work on ethics case studies in a professional issues course. In *Proceedings of the 2014 Zone 1 Conference of the American Society for Engineering Education*. <https://doi.org/10.1109/ASEEZone1.2014.6820686>
- Seat, E., Parsons, J. R., & Poppen, W. A. (2001). Enabling engineering performance skills: A program to teach communication, leadership, and teamwork. *Journal of Engineering Education*, 90, 7–12.
- Shaffer, D. W., Hatfield, D., Svarovsky, G. N., Nash, P., Nulty, A., Bagley, E., ... Mislavy, R. J. (2009). Epistemic network analysis: A prototype for 21st-century assessment of learning. *International Journal of Learning and Media*, 1(2), 33–53. <https://doi.org/10.1162/ijlm.2009.0013>
- Sheppard, S. D., Macatangay, K., Colby, A., Sullivan, W. M., & Shulman, L. S. (2008). *Educating engineers: Designing for the future of the field*. San Francisco, CA: Jossey-Bass.
- Snijders, T. A. B., & Baerveldt, C. (2003). A multilevel network study of the effects of delinquent behavior on friendship evolution. *The Journal of Mathematical Sociology*, 27(2), 123–151. Retrieved from <http://www.informaworld.com/10.1080/00222500305892>
- Sweet, T. M., Thomas, A. C., & Junker, B. W. (2013). Hierarchical network models for education research: Hierarchical latent space models. *Journal of Educational and Behavioral Statistics*, 38(3), 295–318. <https://doi.org/10.3102/1076998612458702>
- Valente, T. W., Fujimoto, K., Unger, J. B., Soto, D. W., & Meeker, D. (2013). Variations in network boundary and type: A study of adolescent peer influences. *Social Networks*, 35(3), 309–316. <https://doi.org/10.1016/j.socnet.2013.02.008>
- Wasserman, S., & Faust, K. (1994). *Social network analysis: Methods and applications*. Cambridge, England: Cambridge University Press.
- Watts, D. J., & Strogatz, S. H. (1998). Collective dynamics of “small-world” networks. *Nature*, 393(6684), 440–442.
- Zhang, M., Ater Kranov, A., Beyerlein, S. W., McCormack, J. P., Pedrow, P. D., & Schmeckpeper, E. (2015). Investigating a scenario-based performance assessment of engineering professional skills. *Proceedings of the IEEE STEM Education Conference* (pp. 230–235). Princeton, NJ: IEEE.
- Zhu, M., & Feng, G. (2015). An exploratory study using social network analysis to model eye movements in mathematics problem solving. In *Proceedings of the 5th International Conference on Learning Analytics and Knowledge (LAK '15)* (pp. 383–387). Poughkeepsie, NY: ACM.
- Zhu, M., Shu, Z., & von Davier, A. A. (2014, April). *Using networks in representing and analyzing process data for educational assessment*. Paper presented at the 34th International Sunbelt Social Network Conference, St. Petersburg, FL.
- Zhu, M., & Zhang, M. (2016). Examining the patterns of communication and connections among engineering professional skills in group discussion: A network analysis approach. In *Proceedings of the 2016 IEEE Integrated STEM Education Conference* (pp. 181–188). Princeton, NJ: IEEE.

Suggested Citation:

Zhu, M., & Zhang, M. (2017). *Network analysis of conversation data for Engineering Professional Skills Assessment* (Research Report No. RR-17-59). Princeton, NJ: Educational Testing Service. <https://doi.org/10.1002/ets2.12191>

Action Editor: James Carlson

Reviewers: Jessica Andrews and Qiwei He

ETS, the ETS logo, and MEASURING THE POWER OF LEARNING. are registered trademarks of Educational Testing Service (ETS). All other trademarks are property of their respective owners.

Find other ETS-published reports by searching the ETS ReSEARCHER database at <http://search.ets.org/researcher/>