NERA Conference Proceedings 2018

Northeastern Educational Research Association (NERA) Annual Conference

2018

The Influences of Communication and Group Dynamics on Collaborative Problem Solving Task Performance

Jonathan Steinberg
Educational Testing Service, jsteinberg@ets.org

Jessica Andrews-Todd
Educational Testing Service, jandrewstodd@ets.org

Carol Forsyth
Educational Testing Service, cforsyth@ets.org

The Influences of Communication and Group Dynamics on Collaborative Problem Solving Task Performance

Jonathan Steinberg

Jessica Andrews-Todd

Carol Forsyth

Paper presented at the 49th annual meeting of the Northeastern Educational Research Association (NERA) held October 17-19, 2018, in Trumbull, CT.

The research reported here was supported by the National Science Foundation through Grant DUE 1535224 to Educational Testing Service in conjunction with The Center for Occupational Research & Development (CORD), and Concord Consortium. The opinions expressed are those of the authors and do not represent views of Educational Testing Service or the National Science Foundation.

The Influences of Communication and Group Dynamics on Collaborative Problem Solving Task Performance

Abstract

This study relates collaborative problem solving (CPS) behavior and background characteristics of three-person student teams completing tasks in an online electronics environment to task performance. Task performance was primarily predicted by classroom membership and minimally impacted by CPS communication types and group dynamics. The online environment and process data measuring CPS behavior substantially add to the field.

The Influences of Communication and Group Dynamics on Collaborative Problem Solving Task Performance

Study Purpose

In this study we examined how the composition of three-person student teams completing a series of tasks in an online electronics environment relates to their collaborative problem solving (CPS) behavior and associated task performance. The work is motivated by research suggesting that a more highly technological workplace requires mastery of this complex skill as a determinant of success (Burrus, Jackson, Xi, & Steinberg, 2013). The use of an online environment and associated process data to measure CPS behavior (Kerr, Andrews, & Mislevy, 2016) in a single content domain may have benefits over the use of traditional assessments such as multiple choice questions because it can provide finer-grained information about students. Online environments allow for the capturing of all actions and discourse as individuals solve problems. This can provide substantial progress in the field, as we can capture and evaluate the processes that individuals use to solve problems rather than just their final answer choices.

Theoretical Framework

The collaborative environment literature focuses on the medium and effectiveness of associated interactions that occur in small groups (Bergner, 2018). From an educational context, as discussed in Lau (2003), collaborative environments can enrich learning and provide new insights on the content area of study (Brace & Roberts, 1997), as well as foster mutual learning in a group setting (Dougherty et al., 1995). Effective communication in online environments is influenced more by meaningful information exchanges than the frequency of communication (Zhu & Zhang, 2017). However, Kreijns, Kirschner, and Jochems (2003) cautioned that types and interpretations of exchanged communications can influence both interpersonal communication (Rice, 1992) and relationships (Walther, 1996). This demonstrates that the *quality* of communications in collaborative environments seems to matter more than the *quantity* of those communications (Forsyth, et al., 2013).

The formulation of groups engaging with these types of environments is also important to acknowledge. According to Bergner (2018), it would be ideal to allow the instructor or developer to condition the grouping of participants on demographic variables or prior performance measures. This would enable the instructor to explore the effects of group composition in the context of collaborative assessment. However, Soller (2001) discussed that the simple placement of students in groups alone by an instructor is insufficient to ensure collaboration, namely that while in some groups interaction may develop naturally, others may encounter difficulties balancing elements such as participation, leadership, understanding, and encouragement. Furthermore, assessing students for proper placement invokes many practical concerns. Nevertheless, the urge for such placement suggests that the types of communication and the group's dynamics may be extremely important.

CPS is a complex construct involving social and cognitive skills making it difficult to be assessed by traditional methods (e.g., multiple choice or constructed response items) (Davey et al., 2015). Technology (e.g., simulation-based items) facilitates a more in-depth assessment of CPS skills (von Davier & Halpin, 2013) because of the interactions allowed and the vast amount of process data collected. However, there are challenges associated with mapping CPS skills to specific actions in an environment and generalizing these to CPS skill mastery based on performance within a specific environment. Thus, one must properly account for students' CPS skill characteristics as these play out in the environment to effectively measure whether or not students demonstrate CPS skill mastery.

Furthermore, the interaction between the group members' characteristics likely also influence their ability to complete a task. For example, there may be concerns around those who engage in "free-riding" behavior (Kerr & Bruun, 1983) in which team members reduce their effort in the task and allow other team members to carry the load. Analysis of process data can provide insights into how each team member contributes to the success of the collaboration through translating event logs into meaningful variables for analysis. Thus, the research questions for this study were generated by utilizing the vast amount of process data to examine how CPS skills and behaviors emerge in a collaborative environment and how these relate to task performance and account for group dynamics.

Research Questions

In this study, we examined the following questions:

1. How do participant characteristics predict task performance?

2. How do types of CPS communication and group dynamics additionally influence performance?

Methodology

Participants

There were 129 electronics and engineering program students across eight community college or university classes participating in the online study (average age = 23). The students worked in groups of three assigned by their instructors, thereby comprising 43 groups. Instructors were asked to randomly assign their students to groups, but it is unclear to what extent that was actually done. The students were predominantly male (81%) and of those reporting their race/ethnicity (2% unreported), the students were predominantly White (51%)

followed by Hispanic (22%), one or more race (10%), Black or African American (7%), Asian (6%), American Indian or Alaskan Native (2%), and the remainder reported Other (2%).

Instrumentation

Students were first given a background survey to collect demographic information, and using five-point Likert scales, preferences about working in groups or alone and attitudes toward the importance of collaboration. The primary task, known as the Three-Resistor Activity (see Figure 1), is a simulation-based task measuring concepts associated with Ohm's Law. In the task, students worked in groups of three, each on a separate computer, and each running a simulation of a portion of an electronic circuit. Students were tasked with reaching a specified goal voltage on their respective circuits. The task consists of four levels, with each successive level being more complex by providing less information upfront (e.g. external voltage and resistance), therefore requiring more collaboration among team members to successfully achieve targeted goal voltage values. Table 1 provides an overview of the task levels, including information about how each task level differed. As students worked with their teams, the system logged their relevant actions, including measurements, resistor changes, calculations, communications through text chat, and submission of work.

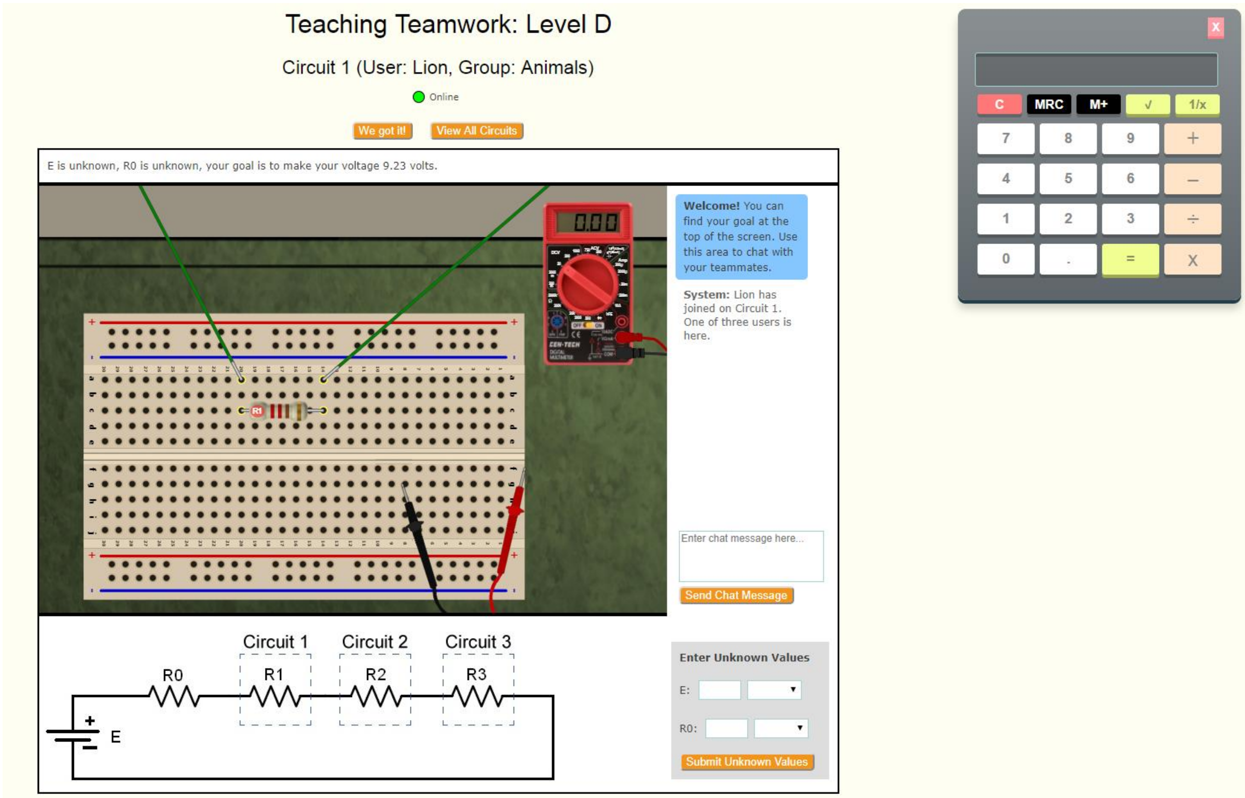

Figure 1. Screenshot of Three-Resistor Activity

Table 1.

Summary of Task Levels

Task Level	**External Voltage (E)**	**External Resistance (R0)**	**Goal Voltages**
1	Known by all teammates	Known by all teammates	Same for all teammates
2	Known by all teammates	Known by all teammates	Different for each teammate
3	Unknown by teammates	Known by all teammates	Different for each teammate
4	Unknown by teammates	Unknown by teammates	Different for each teammate

Communication through text chat was used to measure nine CPS skills. Four skills correspond to the social dimension of CPS (maintaining communication, sharing information, establishing shared understanding, and negotiating) and five skills correspond to the cognitive dimension of CPS (i.e., exploring and understanding, representing and formulating, planning, executing, and monitoring). *Maintaining communication* refers to content irrelevant, social communication whereas *sharing information* refers to content relevant information shared in the service of solving the problem. *Establishing shared understanding* refers to communication used to learn the perspective of others and ensure what has been said was understood. *Negotiating* refers to communication used to express agreement or disagreement or resolve conflicts that arise.

In the cognitive dimension, *exploring and understanding* corresponds to actions used to explore the environment and build a mental representation for components of the problem. *Representing and formulating* refers to communication used represent the problem and formulate hypotheses. *Planning* refers to communication used to develop a strategy for solving the problem. *Executing* corresponds to actions taken to carry out a plan and communication to let teammates know the actions being taken to carry out a plan. *Monitoring* corresponds to actions and communication used to monitor progress toward the goal and the team organization. Given that two of the skills (executing and monitoring) occurred in both actions and chats, these were split into separate CPS skills, resulting in a total of 11 CPS skills for analysis.

These 11 skills are part of an ontology, a theory-driven representation of the specific CPS skills to be measured through engagement with the task. The CPS ontology includes the high-level CPS skills, subskills, the relationships between the skills, and observable behaviors that would indicate evidence of each skill. For more in-depth discussion of the CPS ontology, see Andrews-Todd, Forsyth, Steinberg, & Rupp, (2018) and Andrews-Todd & Kerr (in press).

Instructors also provided ratings of their students' teamwork skills and electronics content knowledge on a five-point Likert scale. In a post-survey, students were asked about their preferences for working alone relative to working with others and their experiences with their teammates during the activity.

Analyses

In answering Research Question 1 to determine how students' characteristics affect task performance, a linear regression model was developed with the team's completed levels (0-4) as the dependent variable and two sets of independent pretest student variables: fixed effects (gender, race, and class) and covariates (attitudes on work styles and collaboration):

$$\text{Score} = \beta_0 + \beta_1\,(\text{Gender}) + \beta_2\,(\text{Race}) + \beta_3\,(\text{Class}) + \beta_4\,(\text{WorkStyle}) + \beta_5\,(\text{Collaboration}) + \varepsilon \quad (1)$$ [1]

In answering Research Question 2 to explore how the types of CPS communication and group dynamics additionally influenced task performance, the regression model residuals[2] were explored utilizing three group-level and three individual-level variables we thought would secondarily relate to performance. The first group-level variable consisted of four CPS skill clusters previously derived from existing CPS frequency data, and consistent with CPS theory. The first, *Chatty Doers*, demonstrated a great deal of off-topic communication, but executed actions at a high level. The second, *Social Loafers*, demonstrated lower levels of CPS skills overall compared to other students. The third, *Group Organizers*, contrasted from the *Chatty Doers*, in that their communications and executed actions were more relevant to completing the task. Finally, the *Active Collaborators* were those demonstrating much higher levels of CPS skills compared to other students. Please refer to Andrews-Todd et al., (2018) for more information about these groups. The previous analysis showed that there was a relationship between skill cluster and performance, such that Active Collaborators had the highest average performance on the task while Social Loafers had the lowest average performance.

---

[1] Respective reference groups in the above model for fixed effects were males (gender), White students (race), and highest value for class designation (class). Three cases were removed for which gender information was not available.

[2] Calculated as observed value – predicted value.

The second group-level variable represented an estimated profile on the social and cognitive dimensions of the ontology using the frequency data from the 11 CPS skills described earlier. Andrews-Todd and Forsyth (2018) used the proportion of skills on each higher-level construct and found relationships to task performance based on profiles derived from median splits of the calculated proportions. However, the previous analysis did not account for the inherent latent relationship between the higher-level social and cognitive dimensions as laid out in the ontology. In this paper, an exploratory factor analysis (EFA) on the frequency data from the 11 CPS skills using maximum likelihood extraction of a two-factor solution with promax rotation was performed, based on the presumption that the two resulting dimensions would be correlated, but not too strongly to produce distinct factors. This was shown to be the case (r = 0.756; $p < 0.001$). A median split on the resulting factor scores was used to create high and low groups on each dimension as inputs to examine model performance. The estimated median for the social factor scores was -0.25 (range = -1.16 to 3.62) and the estimated median for the cognitive factor scores was -0.20 (range = -1.20 to 3.44).

The third group-level variable was the gender composition of the team. This was self-reported by the students with three missing cases, all within different teams, so these three teams were removed for this particular analysis. Of the 40 remaining teams, 24 were exclusively male, 11 had one female team member, and 5 were majority female. The reason for studying this is rooted in an over-arching ongoing concern about females being underrepresented in STEM fields, including engineering (Noonan, 2017), which may be a “missed opportunity” (Milgram, 2011) in the development of human capital, such as women holding senior engineering management positions (Corbett & Hill, 2015) and the resulting potential for innovation. Thus, we investigated gender differences to address this point.

The first of the individual-level variables was level of engagement using observed task times. This was done qualitatively based on the proportion of time each individual spent within and across the different task levels working within their teams. It is important to note that not all teams completed or even were exposed to all task levels and even with individuals working on their own computers, the team advanced to each successive level as time allowed. There were 15 profiles initially identified which could be categorized into three primary types: highly engaged, moderately engaged, and less engaged. Those highly engaged typically spent 40-50% of the team's time on one or more levels, collaborators typically spent 30-35% of the team's time on one or more levels, and those less active spent 20-30% of the team's time on one or more levels. This was worth exploring in the context of the regression results because previous unpublished analyses showed a relationship between membership in one of the four skill clusters referenced in Andrews-Todd et al. (2018) and level of engagement, using chi-square analysis ($\chi^2 = 14.063$, $df = 6$; $p = 0.029$). This additionally allowed us to examine the potential influence of "free riding" (Kerr & Bruun, 1983) within the context of the less engaged participants.

The second of the individual-level variables was based on the instructor's evaluation of the student's teamwork skills prior to engagement with the task. This was presented on a five-point scale (1 = low teamwork skills and 5 = high teamwork skills). Accounting for missing data from six students, the correlation between instructor responses and student task performance was significant ($r = 0.221$; $p = 0.014$) and therefore worthy to examine, even though Andrews-Todd et al., (2018) found no significant relationship between student responses regarding collaboration and CPS skill cluster membership ($p = 0.465$). It should be noted that all but two students received scores of 3 or higher from their instructors on this question. The final individual-level variable was based on the instructor's evaluation of the student's electronics skills prior to engagement with the task. This was presented on a five-point scale (1 = low electronics skills and 5 = high electronics skills). Accounting for missing data from six students, the correlation between instructor responses and student task performance was significant ($r = 0.371$; $p < 0.001$) and therefore likewise worthy to investigate.

Even though there was no association between instructors' evaluation of students' teamwork skills and electronics skills ($\chi^2 = 12.250$, $df = 8$; $p = 0.140$), for examining the regression results, median splits were applied to each variable such that those at or above the median were considered high and those below the median were considered low. The respective medians on both the teamwork question and the electronics skills questions were 4 on the respective five-point scales. Therefore, levels of 1, 2, or 3 were considered low and levels of 4 or 5 were considered high. In essence, this would produce a seemingly different social-cognitive skill profile based on the instructors' perceptions relative to those generated from the 11 CPS skills analyzed through the EFA described earlier.

Results

Research Question 1

The Appendix shows that the independent variables in the regression explained just over 70% of the variance in successfully completed task levels. It should be noted that not all teams were able to attempt, let alone complete, successive task levels if the lower-level tasks could not be completed. Among the 43 teams starting at the first level, only 34 attempted the second level, 27 attempted the third level, and 20 attempted the last level. For purposes of our analyses, any non-attempted levels were automatically coded as being not completed rather than missing.

Only class membership was a significant predictor of levels completed, our proxy measure for performance ($p < 0.05$) with similar results in the majority of individual classes.

Therefore, there was quite a bit of variation in team performance across different classes as shown in Table 2, with mean levels completed ranging from 2.09 to 4. None of the teams in Class G completed the first level, whereas all but one team in each of Classes B and C completed all levels.

Table 2.

Summary of Task Performance by Class

				Levels Completed				
Class	Number of Teams	Mean	SD	0	1	2	3	4
A	6	2.67	0.97	0	1	1	3	1
B	3	3.67	0.50	0	0	0	1	2
C	5	3.80	0.41	0	0	0	1	4
D	4	1.75	1.36	1	1	0	2	0
E	6	3.17	0.71	0	0	1	3	2
F	5	1.80	1.01	0	3	0	2	0
G	7	0.00	0.00	7	0	0	0	0
H	7	1.29	0.72	1	3	3	0	0
Total	43	2.09	1.47	9	8	5	12	9

Research Question 2

Table 3 displays mean task levels completed by the four skill clusters and associated model residuals. While most cluster means are similar, the Active Collaborator profile suggests a positive relationship to performance. Model residuals indicate that on average, Group Organizers may have underperformed, while Active Collaborators may have over-performed, although sample sizes are small in these two groups.

Table 3.

Summary of Task Performance by Skill Frequency Cluster

		Task Performance		Levels Completed					Model Residuals	
Cluster	*N*	Mean	*SD*	0	1	2	3	4	Mean	*SD*
Chatty Doers	33	2.18	1.33	4	8	5	10	6	-0.05	0.75
Social Loafers	67	1.91	1.54	18	13	8	13	15	0.04	0.70
Group Organizers	16	2.06	1.39	3	3	2	6	2	-0.38	0.80
Active Collaborators	10	3.30	0.48	0	0	0	7	3	0.50	0.79

Table 4 shows mean task levels completed across the four social and cognitive skill component score profiles and associated model residuals. Given this CPS task with an emphasis on social interaction, it was expected that those with lower average social skill component scores would complete fewer levels than those with above average social skill scores. While cognitive skills are also important in our definition of CPS and shown to be moderately correlated with social skills through the EFA, we did not necessarily presume cognitive skill differences would be as evident in this analysis. Nonetheless, the residuals suggest those with below average social scores but above average cognitive scores may have underperformed.

Table 4.

Summary of Task Performance by Social and Cognitive Skill Component Profile

		Task Performance		Levels Completed					Model Residuals	
Profile	*N*	Mean	*SD*	0	1	2	3	4	Mean	*SD*
Low Social - Low Cognitive	55	1.35	1.29	19	15	7	11	3	-0.15	0.70
Low Social - High Cognitive	10	1.30	1.34	4	2	1	3	0	-0.31	0.79
High Social - Low Cognitive	9	1.78	1.20	1	3	3	1	1	0.09	0.90
High Social - High Cognitive	55	3.04	1.14	3	4	4	21	23	0.19	0.74

Table 5 shows mean task levels completed based on team gender profile and associated model residuals. While the majority male or exclusively male teams on average performed as expected given the regression model, those on majority female teams on average did not complete even the first level and showed evidence of underperforming based on the regression model. A follow-up investigation revealed that the majority of these teams were in Classes G and H, which underperformed relative to those from other classes.

Table 5.

Summary of Task Performance by Team Gender Profile

		Task Performance		Levels Completed					Model Residuals	
Team Gender Profile	*N*	Mean	*SD*	0	1	2	3	4	Mean	*SD*
All Male	72	2.42	1.36	6	18	9	18	21	0.02	0.78
Majority Male	33	2.18	1.29	6	3	6	15	3	0.06	0.84
Majority Female	15	0.80	1.21	9	3	0	3	0	-0.27	0.48

Table 6 displays mean task levels completed across three levels of engagement (highly engaged, moderately engaged, less engaged) and associated model residuals using task times. No substantive differences in average task performance or model residuals were detected. However, it is evident that at least on average, those in the less engaged group had the highest number of levels completed, which may indicate some notion of "free-riding" behavior (Kerr & Bruun, 1983).

Table 6.

Summary of Task Performance by Engagement Level

		Task Performance		Levels Completed					Model Residuals	
Engagement Level	*N*	Mean	*SD*	0	1	2	3	4	Mean	*SD*
Highly Engaged	42	2.10	1.43	9	7	3	17	6	0.04	0.72
Moderately Engaged	51	2.02	1.52	12	9	8	10	12	0.01	0.79
Less Engaged	33	2.27	1.40	4	8	4	9	8	-0.07	0.76

Table 7 displays mean task levels completed across the social-cognitive profiles derived from instructors' evaluations of students' teamwork (social) and electronics (cognitive) skills and associated model residuals. Inspection of the means looks different compared to the one presented in Table 4 in two distinct ways. First, there appears to be a more direct relationship on the electronics skills (cognitive) level to task performance, particularly from the Low Social – High Cognitive and High Social – Low Cognitive groups. Also, even with a relatively low perception of students' teamwork skills, on average one additional level is completed compared to the orientation based on CPS skills. Further investigation between our qualitative ratings of student skills and instructor-derived ratings may suggest that instructors could interpret CPS differently than the researchers (for more information see Andrews-Todd & Forsyth, 2018)

Table 7.

Summary of Task Performance by Instructor Rating Profile

		Task Performance		Levels Completed					Model Residuals	
Instructor Rating Profile	*N*	Mean	*SD*	0	1	2	3	4	Mean	*SD*
Low Social - Low Cognitive	15	2.27	1.16	2	1	4	7	1	-0.16	0.98
Low Social - High Cognitive	15	2.60	0.99	0	3	2	8	2	0.05	0.84
High Social - Low Cognitive	37	1.46	1.48	13	11	1	7	5	-0.09	0.71
High Social - High Cognitive	56	2.48	1.48	9	7	7	14	19	0.13	0.68

Conclusions and Future Directions

These results show that CPS skill profiles and group dynamics obtained from students' behavior or their instructors' ratings minimally influenced task performance beyond pretest individual demographic and attitudinal factors. However, there does appear to be some evidence that estimates of group performance on this task are affected by certain CPS skill profiles (Group Organizers underperform and Active Collaborators over-perform), ontological profiles derived through EFA (Low Social – High Cognitive students underperform), and gender composition (majority female teams underperform). In the case of the ontological profiles derived through EFA, better understanding of the relationships among the higher-level social and cognitive dimensions becomes important. In a future study we will attempt to identify further lower-level skills to help solidify the foundations for these higher-level dimensions. In the case of gender composition, more exploration of the underlying characteristics of the exclusively female teams may be warranted. If there are deficits with respect to underlying electronics knowledge and/or teamwork skills, instructors may need to more purposefully assign female students to teams to avoid lack of engagement or improve chances of success with these types of tasks.

While the results should be interpreted cautiously, our approach may represent an important and interesting aspect of CPS at least within the electronics domain. Given the small group sizes, these results did not account for nesting of students within classes. This will be considered in a future study with a larger sample size since class membership was a significant predictor of task performance in this study. Also specifically by gender, the relative sparseness of female participants may have presented difficulties in uncovering more realistic group dynamic effects due to team gender composition. A larger sample size will also allow for potential examination of team composition by race/ethnicity given underrepresentation issues exist at this level as well within institutions and in the workforce (National Science Foundation, 2017). Additionally, it is clear from some of the results that there may be a misalignment between instructor perceptions of the underlying social (teamwork) and cognitive (electronics) skills of their students relative to what the students themselves produced within the task. This may or may not have been a result of team assignment, even though instructors were told to randomly assign students to teams. In a future study we will examine instructor practices, student background knowledge, and student teamwork skills more in-depth to understand where students may be starting out on these dimensions prior to engaging with the task. However, the rationale for how teams were formed by instructors was not given and may be explored in future research.

There are multiple issues to consider with this line of research on a broader scale that revealed limitations with the results shown in this paper and require further analysis and study. The first is more of a fundamental question around how success with this task is defined, both at the group and individual levels. The current task assigns group-level scores based on solving for a goal voltage, but no standards currently exist as to the optimal mix of individual or group-level CPS behaviors that are required or the amount of time needed to solve for those goal voltages. It may also be important to note that the need for exhibiting CPS skills becomes progressively more important as students complete each level, thus the difficulty of each successive level is certainly not equivalent to the preceding level.

As discussed earlier, it was revealed from examining timing data on the task that some team members did not fully engage at certain levels, whether consistently or in shifting from being more highly engaged to allowing others to become more engaged, which may have affected their profile on the 11 CPS skills at each level. Therefore, while we plan empirical investigations to further examine the skill profiles, development of formal individual scoring rules would be an extremely complex undertaking at this stage.

The level of fidelity of implementation is uncertain given that teams are composed of students within the same class, who may or may not have prior familiarity with each other and/or their respective levels of content knowledge, which may be necessary in order to effectively collaborate on the task. Additionally, we currently are not able to determine whether there were interactions taking place within the in-class environment that were not captured in the chat transcripts. A corresponding grant led by one of the co-authors with a middle school population in a different content area utilizes video to more directly capture aspects of CPS behavior among team members not directly evident from the chats, such as body language and vocal inflections. Additionally, this grant will allow for the comparison of results using dyads of students compared to triads in this study.

Recruitment is ongoing to validate the underlying ontology and associated assessment with a larger sample. In addition, the study design will be strengthened in two ways. One is through the use of concept inventories related to CPS (Anderson & West, 1998; Cegala, Savage, Brunner, & Conrad, 1982). These inventories will allow participants to self-report their evaluations of individual and team CPS behaviors which is a step above the current study which only relies on the information from the chats and associated time on task and instructor ratings collected prior to student engagement with the task. Secondly, the effect of personality in relation to group dynamics will be studied more closely through the use of established instruments measuring the Five Factor Model of Personality (e.g. John & Srivastava, 1999). Thus, future research will be able to further examine, expand upon, and provide further validation for the results found in the current paper as well as the underlying theory with a larger sample and the additional instrumentation.

References

Anderson, N. R., & West, M. A. (1998). Measuring climate for work group innovation: Development and validation of the team climate inventory. *Journal of Organizational Behavior: The International Journal of Industrial, Occupational and Organizational Psychology and Behavior*, *19*(3), 235-258.

Andrews-Todd, J., & Forsyth, C. M. (2018). Exploring social and cognitive dimensions of collaborative problem solving in an open online simulation-based task. *Computers in Human Behavior*. https://doi.org/10.1016/j.chb.2018.10.025

Andrews-Todd, J., Forsyth, C. M., Steinberg, J., & Rupp, A. A. (2018). Identifying profiles of collaborative problem solvers in an online electronics environment. In K. E. Boyer & M. Yudelson (Eds.), *Proceedings of the 11th International Conference on Educational Data Mining* (pp. 239–245). Buffalo, NY: International Educational Data Mining Society.

Andrews-Todd, J. & Kerr, D. (in press). Application of ontologies for assessing collaborative problem solving skills. *International Journal of Testing.*

Bergner, Y. (2018). *CPSX: A tool for online collaborative problem-solving in open edX*. (Research Memorandum 18-03). Princeton, NJ: Educational Testing Service.

Brace, S., & Roberts, G. (1997). When payup becomes payback: A university’s return on instructional technology investment, *Syllabus*, *11*(1), 30-37.

Burrus, J., Jackson, J., Xi, N., & Steinberg, J. (2013). *Identifying the most important 21st century workforce competencies: An analysis of the Occupational Information Network (O*NET)*. (Research Report 13-21). Princeton, NJ: Educational Testing Service.

Cegala, D. J., Savage, G. T., Brunner, C. C., & Conrad, A. B. (1982). An elaboration of the meaning of interaction involvement: Toward the development of a theoretical concept. *Communications Monographs, 49*(4), 229-248.

Corbett, C., & Hill, C. (2015). *Solving the equation: The variables for women's success in engineering and computing* (pp. 1-141, Rep.). Washington, DC: American Association of University Women.

Davey, T., Ferrara, S., Shavelson, R., Holland, P., Webb, N., & Wise, L. (2015). *Psychometric considerations for the next generation of performance assessment*. Washington, DC: Center for K-12 Assessment & Performance Management, Educational Testing Service.

Dougherty, R. C., Bowen, C. W., Berger, T., Rees, W., Mellon, E. K., & Pulliam, E. (1995). Cooperative learning and enhanced communication: Effects on student performance, retention, and attitudes in general chemistry. *Journal of Chemical Education, 72*(9), 793-797.

Forsyth, C. M., Graesser, A. C., Pavlik Jr, P., Cai, Z., Butler, H., Halpern, D., & Millis, K. (2013). Operation aries!: Methods, mystery, and mixed models: Discourse features predict affect in a serious game. *Journal of Educational Data Mining*, *5*(1), 147-189.

John, O. P., & Srivastava, S. (1999). The Big Five trait taxonomy: History, measurement, and theoretical perspectives. In L. A. Pervin & O. P. John (Eds.), *Handbook of personality theory and research* (pp. 102–138). New York, NY: Guilford Press.

Kerr, D., Andrews, J. J., & Mislevy, R. J. (2016). The in-task assessment framework for behavioral data. In A. A. Rupp & J. P. Leighton (Eds.), *The Wiley handbook of cognition and assessment: Frameworks, methodologies, and applications* (pp. 472-507). Hoboken, NJ: Wiley-Blackwell.

Kerr, N., & Bruun, S. (1983). The dispensability of member effort and group motivation losses: free-rider effects. *Journal of Educational Computing Research*, *5*, 1–15.

Kreijns, K., Kirschner, P. A., & Jochems, W. (2003). Identifying the pitfalls for social interaction in computer-supported collaborative learning environments: A review of the research. *Computers in Human Behavior*, *19*(3), 335-353.

Lau, L. K. (2003). Institutional factors affecting student retention. *Education*, *124*(1), 126-136.

Milgram, D. (2011). How to recruit women and girls to the science, technology, engineering, and math (STEM) classroom. *Technology and Engineering Teacher*, *71*(3), 4.

National Science Foundation, National Center for Science and Engineering Statistics (2017). *Women, minorities, and persons with disabilities in science and engineering: 2017*. Special Report NSF 17-310. Arlington, VA.

Noonan, R. (2017). *Women in STEM: 2017 update* (ESA Issue Brief #06-17). Washington, DC: Economics and Statistics Administration, U.S. Department of Commerce.

Rice, R. E. (1992). Contexts of research on organizational computer-mediated communication. In M. Lea (Ed.), Contexts of computer-mediated communication (pp. 113–144). London, UK: Harvester Wheatsheaf.

Soller, A. L. (2001). Supporting social interaction in an intelligent collaborative learning system. *International Journal of Artificial Intelligence in Education*, *12*, 40-62.

von Davier, A., & Halpin, P. F. (2013). *Collaborative problem-solving and the assessment of cognitive skills: Psychometric considerations* (Research Report 13-41). Princeton, NJ: Educational Testing Service.

Walther, J. B. (1996). Computer-mediated communication: Impersonal, interpersonal, and hyperpersonal interaction. *Communication Research*, *23*(1), 3–43.

Zhu, M., & Zhang, M. (2017). *Network analysis of conversation data for engineering professional skills assessment*. (Research Report 17-59). Princeton, NJ: Educational Testing Service,.

Appendix: Baseline Linear Regression Results[3]

Tests of Between-Subjects Effects on Dependent Variable: Team Levels Completed

Source	Type III Sum of Squares[4]	*df*	Mean Square	*F*	*p*-value	Partial η^2
Corrected Model	191.509	12	15.959	25.423	< 0.001	0.730
Intercept	29.784	1	29.784	47.447	< 0.001	0.296
Gender	0.336	1	0.336	0.536	0.466	0.005
Race/Ethnicity	1.774	2	0.887	1.413	0.248	0.024
Class	**167.345**	**7**	**23.906**	**38.083**	**< 0.001**	**0.702**
WorkStyle	0.348	1	0.348	0.554	0.458	0.005
Collaboration	0.105	1	0.105	0.167	0.684	0.001
Error	70.935	113	0.628			
Total	824.000	126				
Corrected Total	262.444	125				

Note: $R^2 = 0.730$ (Adjusted $R^2 = 0.701$); *df* = degrees of freedom.

Parameter Estimates from Regression with Dependent Variable: Team Levels Completed

Variable	Coefficient	*SE*	*t*	*p*-value	Partial η^2
Intercept	1.712	0.415	4.120	< 0.001	0.131
Female	-0.150	0.204	-0.732	0.466	0.005
Male	0[5]				
African American	-0.281	0.192	-1.464	0.146	0.019
Hispanic	0.055	0.173	0.317	0.752	0.001
White	0[4]				
Class A	**1.141**	**0.302**	**3.774**	**< 0.001**	**0.112**
Class B	**2.249**	**0.338**	**6.660**	**< 0.001**	**0.282**
Class C	**2.361**	**0.282**	**8.375**	**< 0.001**	**0.383**
Class D	0.467	0.299	1.562	0.121	0.021
Class E	**1.819**	**0.258**	**7.053**	**< 0.001**	**0.306**
Class F	0.457	0.284	1.608	0.111	0.022
Class G	**-1.346**	**0.256**	**-5.261**	**< 0.001**	**0.197**
Class H	0[4]				
WorkStyle	-0.060	0.080	-0.744	0.458	0.005
Collaboration	-0.024	0.059	-0.408	0.684	0.001

Note: *SE* = standard error; *t* = Student's *t* statistic

[3] Significant results ($p < 0.05$) for key variables are in bold text.
[4] Default option in SPSS 23 and there is no nested design here when Type I or Type II sums of squares would be used.
[5] Parameter set to zero because it is redundant.