

The Role of Robotics Teams' Collaboration Quality on Team Performance in a Robotics Tournament

Muhsin Menekse,^a Ross Higashi,^b Christian D. Schunn,^b
and Emily Baehr^b

^aPurdue University, ^bUniversity of Pittsburgh

Abstract

Background Working effectively in teams is an important 21st century skill as well as a fundamental component of the ABET professional competencies. However, successful teamwork is challenging, and empirical studies with adolescents concerning how the collaboration quality of team members is related to team performance are limited.

Purpose/Hypothesis This study investigated the relationship between team collaboration quality and team performance in a robotics competition using multiple measures of team performance, including both objective task performance and expert judge evaluations, on a diverse set of supporting performance dimensions.

Design/Method Data included Table Score, Robot Design, Research Project, Core Values, and Collaboration Quality scores for 366 youths on 61 K-8 robotics teams that participated in a FIRST LEGO League Championship. Regression and mediation analyses were conducted to explore the relation between effective team collaboration and team performance. Furthermore, analysis of variance was conducted to explore the relationship between Collaboration Quality and team experience.

Results Collaboration Quality was a good predictor of robotics team performance across all measures (with $R^2 = .50$ and $p < .001$). Mediation analysis revealed that the Robot Design acted as a full mediator for the predictive effect of Collaboration Quality on the Table Score. In addition, the cumulative amount of team experience was significantly related to Collaboration Quality.

Conclusions Overall, this study using collaboration performance assessments and actual competition data with a large number of teams confirms the importance of high-quality teamwork in producing superior products with students engaged in authentic engineering tasks.

Keywords teamwork; collaborative learning; educational robotics; informal learning; K-12

Introduction

K-12 robotics competitions have emerged as popular educational activities in recent years. By 2015, more than 230,000 students were participating in 29,000 First Lego League robotics teams across 80 countries (Close, 2015). Past research has shown that being part of such robotics teams has the potential to significantly influence students' academic and social skills by allowing them to actively engage in critical thinking and problem solving through

designing, assembling, coding, operating, and modifying robots for specific goals (Bascou & Menekse, 2016; Benitti, 2012).

Robotics teams design solutions for a wide variety of sometimes ill-defined problems. In doing so, they negotiate an open-ended problem space, taking it upon themselves to identify, investigate, and implement solutions from among a large number of possible directions. Problem-based learning (PBL) activities such as these have the potential to develop not just technical skills but also the professional skills that enable learners to effectively apply their technical knowledge in authentic conditions (Hmelo-Silver, 2004; Hmelo-Silver, Duncan, & Chinn, 2007).

Working effectively in teams is a critical 21st century skill (Borrego, Karlin, McNair, & Beddoes, 2013; Koenig, 2011; Shuman, Besterfield-Sacre, & McGourty, 2005; Tonso, 2006), and teamwork settings are good venues for studying factors in effective team performance. One factor or skill of particular interest is collaboration, or a student's ability to work with peers interdependently through social cohesion and interaction. This skill is educationally relevant in two distinct ways: (a) as a facilitator of interactions that increase content learning, and (b) as an important skill itself (Chi & Menekse, 2015; Clark et al., 2010). While solving challenges as part of a robotics team appears on the surface to provide students with an opportunity to build collaboration skills, past research on instructional design suggests that the structure of the task affects whether such skills can be fostered (Stahl, 2005), and the development of effective PBL tasks is particularly critical (Mehalik, Doppelt, & Schunn, 2008).

To explore this issue, this study investigated the relationship between team Collaboration Quality and team performance in a robotics competition using multiple measures of team performance. In addition, we explored whether participation in robotics competitions tends to develop collaboration skills by examining the relationship between the teams' cumulative competition experience and the quality of the members' collaboration. Our overall goal was to gain insight into how robotics competitions involve and enhance collaboration in an authentic way, and to what degree participation in such a competition is likely to develop collaboration skills. The relationships between these factors reflect on both the role of collaboration and the authenticity of the competition tasks in relation to engineering tasks where effective collaboration is widely accepted as being important to success. Our specific research questions and hypotheses were as follows:

RQ 1. What is the relationship between effective team collaboration and team performance outcomes in a robotics competition?

Hypothesis 1. Teams with higher Collaboration Quality scores will produce superior robots and score higher in the overall competition.

RQ 2. Does participating in robotics competitions build competency in collaboration skills?

Hypothesis 2. Teams that have participated in more competitions will demonstrate higher Collaboration Quality scores.

The structure of the paper is as follows: the Background Section reviews robotics competitions and problem-based learning (PBL), educational robotics, and collaboration literature in K-12 formal and informal settings; the Methods Section presents the data sources and team scores used in this study; the Results Section discusses the analyses conducted and the results;

and General Discussion summarizes the findings, concluding by discussing some of the limitations of the study and the future work needed.

Background

Learning in Robotics Competitions

Robotics competitions, which have contributed to a broad recognition of educational robotics, provide unique opportunities for students and teams to work toward a shared goal in a certain timeframe (Bascou & Menekse, 2016; Danahy et al., 2014; Eguchi, 2014). Common goals for many competitions include the development of academic skills and interest in and awareness of science, technology, engineering, and mathematics (STEM); a focus on the ability to work effectively in teams; and the development of cooperation and respect toward the other teams participating in the competitions.

The robotics competition with the largest number of teams (not only in the United States but also worldwide) is the FIRST LEGO League (FLL), which began in 1998 as a joint effort between the FIRST (For Inspiration and Recognition of Science and Technology) Organization and the LEGO Group to introduce robotics to 9- to 14-year-old students. Participation is typically voluntary, either as part of an elective class or as an afterschool activity. To compete in the FLL, teams are required to use LEGO kits (Mindstorms robot sets and software) to work on an authentic scientific-themed challenge, with past themes including climate change, senior solutions, food safety, and medicine. FLL organizers release a challenge each year, after which competing teams spend the next several months preparing for a local Grand Championship competition. At these competitions, team performances are evaluated based on scores in four areas: the robot game, a research project, the quality of Robot Design and programming, and demonstration of FLL Core Values (see the Methods Section for more details). The data used in this study come from the 2015 FLL Western PA Championship.

FLL tasks follow a similar value structure as robotics engineering itself (Jordan & McDaniel, 2014). Specifically, the robot design component of the FLL competitions, which focuses on mechanical design, programming, and strategy and innovation, is comparable to robotics engineering competencies. Robotics engineering draws on the expertise of many engineering disciplines including mechanical, industrial, electrical, and computer engineering. Similar to robotics engineers, students on FLL teams are expected to design, build, program, test, and redesign as needed robots and other robotics devices to meet the challenge of solving authentic problems such as using robots for disaster response. These authentic tasks in FLL competitions achieve two primary goals: first, they allow students to connect what they learn about robotics to what they could do in the face of real-world challenges; and second, authentic tasks and plausible scenarios are structured to motivate students to overcome potential challenges in learning robotics.

Past research has shown that such use of educational robotics can increase interest and engagement in STEM (Kim et al., 2015; Mohr-Schroeder et al., 2014), as well as increase critical thinking and problem-solving skills (Okita, 2014), computational thinking (Grover & Pea, 2013; Menekse, 2015), mathematics (Alfieri, Higashi, Shoop & Schunn, 2015; Martinez Ortiz, 2011), physics (Williams, Ma, Prejean, Ford & Lai, 2007), and science literacy (Sullivan, 2008). For example, Verner and Ahlgren (2004) showed students learned key engineering skills such as systems-thinking, problem-solving, and teamwork skills by designing, building, and operating educational robots. Petre and Price (2004) explored the role of

participation in robotics competitions on student development of such engineering design principles as determining possible solutions and communicating these solutions to others.

In contrast, in a meta-study of robotics in education, Benitti (2012) found that many educational robotics studies, in general, remain inconclusive with respect to student learning outcomes. Specifically, Benitti (2012) found three of six experimental or quasi-experimental robotics studies exploring student learning outcomes in various subjects (e.g., mathematics, computation, among others) did not find a significant difference between experimental and control conditions. In addition, at the college level, Fagin and Merkle (2002) conducted a large-scale experimental study exploring the effects of robots on learning introductory computer programming using the Ada/Mindstorms programming environment, finding that the college students in the robotics condition performed worse than their peers in regular computer science courses that included no robotics instruction.

A limited number of studies have investigated robotics competitions in particular, although these tend to be survey-based rather than performance-based. In 2005, an extensive evaluation report of the FIRST Robotics Competitions (including the FLL for 9- to 14 year-olds and other competitions targeting older and younger populations) was published based on survey data from student participants (Melchior et al., 2005). According to this report, 55% of the participants in the FIRST Robotics Competitions were from under-represented minority groups, with 41% being female. In addition, a majority of participants reported that being part of their FIRST Robotics teams provided them with the opportunity to form positive relationships with their peers, a chance to play a leadership role and to assume real responsibilities, and to participate in decision-making processes. Moreover, students expressed an increased value for teamwork, interest in science and technology, and self-esteem. A second study based on data from FIRST Robotics Competitions found that the students who participated in them had a more positive attitude toward the social implications of science, normality of scientists, attitude toward scientific inquiry, and adoption of scientific attitudes (Welch & Huffman, 2011).

Robotics Competitions as PBL

PBL is an instructional design in which students working in small groups pursue solutions to realistic open-ended problems in specific domains, independently seeking information and resources as necessary, while teachers provide indirect guidance (Barrows, 1996, 2002; Hmelo-Silver, 2004). This style of instruction originated in medical schools, where it has been found to enhance a disposition toward lifelong learning (Shin, Haynes, & Johnston, 1993), self-regulated learning (White, 2007), and critical thinking (Tiwari, Lai, So, & Yuen, 2006).

The FLL challenge meets four criteria commonly used to distinguish PBL activities (cf. Barrows, 2002): (a) ill-structured problems, (b) a learner-centered method, (c) teachers as facilitators, and (d) authentic tasks. Each year's FLL challenge includes an ill-structured "game board" problem based on an 8' × 4' tabletop setup of props with which the robot is expected to interact in various ways to earn points, some mutually exclusive. Students are thus forced to select tasks by deliberately considering the interrelated set of physical and strategic design constraints. To do so, team members identify the knowledge and skills they are missing, then seek out experts, lessons, online videos, or online forums to acquire what they need to build and program their designs. This learner-centric method is augmented by a focus on teachers as facilitators as emphasized in the FLL Core Values: "We [the students] do the work to find solutions with guidance from our Coaches and Mentors" (Core Values, 2016).

Finally, the task is authentic: robot locomotion and manipulation must address appropriate problems, and programming and mechanical designs are aligned with valuable skills in engineering.

Learning Through Collaboration

Professional skills such as collaboration and self-directed learning are both an outcome of PBL experiences (Hmelo-Silver et al., 2007) and predictors of success within them (Barron, 2000; Vye, Goldman, Voss, Hmelo, & Williams, 1997). We might, therefore, surmise that PBL both builds and depends upon collaborative work skills. The effects of collaboration on learning outcomes in PBL are theorized to originate from social-motivational factors, such as shared goals and interests, team spirit, and peer pressure; and cognitive factors, such as opportunities for activation of knowledge, argumentation, and elaboration (Dolmans, De Grave, Wolfhagen, & Van Der Vleuten, 2005; Schmidt, Rotgans, & Yew, 2011). However, while individual and collaborative aspects of PBL appear to mutually reinforce one another (Yew, Chng, & Schmidt, 2011), Sweller, Kirschner, and Clark (2007) found that the relative benefits of collaboration in PBL may actually result from students backfilling a vacuum in guidance that PBL created in the first place.

Collaborative effects in learning are not unique to PBL scenarios, of course. Many studies have revealed the significant role of peer interactions and verbal communication for knowledge construction (Chi, 2009; Hogan, Nastasi, & Pressley, 1999; Jeong, 2013; Jeong & Chi, 2007; Menekse, Stump, Krause, & Chi, 2013; Webb, 1989). However, some studies in learning sciences and educational psychology have shown that achieving successful collaboration is challenging and that working in small groups is not always beneficial in terms of group performance and individual learning (e.g., Barron, 2003; Chi & Menekse, 2015; Nokes-Malach, Richey, & Gadgil, 2015; Purzer, 2011; Stump, Hilpert, Husman, Chung, & Kim, 2011). Furthermore, past research indicates that peer interaction is important not only for the improvement of academic achievement but also for social skills development, including helping others, sharing, taking turns, showing respect, and working collaboratively.

Most of these collaborative learning studies have been conducted in formal educational settings such as schools and other academic environments (e.g., Denis & Hubert, 2001), with relatively few focusing on collaborative learning in informal settings such as robotics competitions (e.g., Verma, Puvirajah, & Webb, 2015). Such informal learning settings provide a unique environment for young people to interact with peers from different age groups, to learn from student mentors, and to engage with apprenticeship experiences. Learning practices in these informal environments are quite different from those associated with traditional schools, which are typically authoritative, with the teachers determining the formation, organization, and presentation of the content. In addition, instruction is primarily direct, student attendance is mandatory, and the primary motivation is the transmission of knowledge rather than encouraging curiosity or promoting creativity. On the other hand, learning practices in informal environments are usually nondirect, require voluntary participation of students, and provide learning nourished through curiosity, observation, and interactive activities (Falk & Dierking, 2000).

Robotics teams, specifically, ask students to work collaboratively in order to design, build, and program robots for various tasks. Thus, student social and discursive practices during these collaborations could play a substantial role in team harmony and success. Attending competitions as part of a robotics team has the potential to provide diverse collaborative learning experiences that enrich the knowledge and interest of students (Johnson & Londt,

2010) via authentic verbal and social interactions with peers (Puvirajah, Verma, & Webb, 2012). However, Benitti's (2012) review study indicated inconclusive results regarding the effectiveness of educational robotics on teamwork skills. For example, the teamwork study of robotics summer camps conducted by Nugent and colleagues (Nugent, Barker, Grandgenett, & Adamchuk, 2009) found no difference between the control and robotics conditions in terms of student attitudes toward teamwork.

Two more recent studies of collaborative learning in robotics teams primarily focused on the nature of collaboration through in-depth discourse analysis (Jordan & McDaniel, 2014; Verma et al., 2015). However, they did not explore the role of collaboration on team performance. Furthermore, both of these studies involved small samples: Verma et al. (2015) studied one team with nine individuals, and Jordan and McDaniel's study (2014) included one classroom with 24 students working in groups of three to four members.

To address the limitations of previous studies, we used actual performance data from a regional championship of a robotics competition, collecting data from a sufficiently large number of teams to support analyses at the team level in this study. To explore our research questions, we examined the ways in which robotics competition scores correlate with successful collaboration and the evidence that participation in such competitions builds team-level competency in collaboration skills. The quantitative investigation of Collaboration Quality as a predictor of task performance also appears to be a unique contribution to research on PBL.

In this study, we employed the Enyedy and Stevens' (2014) collaboration-as-learning methodological approach, which uses four dimensions to differentiate and explain various goals for studying collaboration (e.g., as the outcome vs. as a method for learning). The first dimension distinguishes between individual versus collective processes, while the second describes the outcomes as individual or collective and the third is the degree to which the outcomes are within collaboration (as proximal) versus outside the collaboration (as distal). Finally, the fourth dimension refers to taking a normative versus endogenous stance on collaboration. Based on these four dimensions, the collaboration-as-learning approach operates on the *collective* unit for processes and outcomes, the *proximal* degree, and on an *endogenous* stance on collaboration. In other words, the collaboration-as-learning approach focuses on the collective and distributed units of cognition and learning rather than focusing on individual performance. Furthermore, this approach sees the collective unit (group, team) as durable, meaning that the unit continues to operate when new members join and old members leave. This assumption is corroborated by the literature from the field of organizational theory (e.g., Brown & Duguid, 1998), which posits that organization-level knowledge may be encoded into practices such as organizational routines, norms, and networks of relationships rather than being held by individuals. Accordingly, in our study, the unit of analysis both for processes and outcomes were robotics teams rather than individual students within the teams. In addition, we took an endogenous approach by focusing on the collaborative effort rather than the normative outcomes of individual success.

Methods

Participants

Data were collected from all teams (366 adolescences on 61 teams) that participated in the FIRST LEGO League Western PA Championship. Approximately 57% of the participants were identified as female, based on the data from the 148 participants who completed an optional background survey. In addition, the average age of participants was 11.7 years with a

standard deviation of 1.3. Teams attending the championship event had previously competed in at least one local qualifying event in which the Table Score (see below) was used as a qualifying metric.

Measures

The four FLL scores obtained for each team during the competition were Table Scores, Core Values, Project Score, and Robot Design. Table Scores reflected the performance of the team's robot on the competition task itself. All 61 teams were judged on the other three categories (Core Values, Project Score, and Robot Design) on a 1 to 4 scale, with 1 indicating a beginning level and 4, an exemplary level. Additionally, each team was assessed for Collaboration Quality on a brief performance task in an additional interview based on a 1 to 3 scale, with 1 indicating minimal and 3, substantial. Our research team developed this Collaboration Quality score, which was not part of the original four FLL scores. All measures are discussed in detail in the following section, with the judging processes being described in the Procedure Section.

Table Scores Table Scores indicated the actual robot performance on the specific challenges for the 2015 FLL game called "World Class." Teams earned points when their autonomously controlled LEGO robots successfully transported or manipulated small set-piece mechanisms on the tabletop game board. All teams were provided with the layout in advance. Due to the number and difficulty of the challenges, a perfect score was almost unattainable as students were required to make strategic decisions regarding which challenges to attempt. Each team's robot performed for three rounds, with the best score being used for the tournament ranking. For this research, however, we used the mean score across the three rounds instead of the best scores as a more reliable estimate of team performance.

Core Values FLL defines, promotes, and emphasizes a particular set of Core Values that the students and coaches must exemplify throughout the competition season. A team's final Core Values Score is based on a short presentation and follow-up interview with volunteer judges working in pairs, who evaluate teams on the three subscores of inspiration, teamwork, and gracious professionalism using detailed rubrics developed by the FLL. Inspiration is judged based on evidence of a team's ability to integrate the FLL values into their daily lives, their team spirit, and their balanced emphasis on all aspects of the FLL (i.e., friendly competition and learning). The teamwork subscore is based on the judges' assessment of each team's efficiency and effectiveness in problem solving, time management, distribution of roles and responsibilities, and team independence with minimal involvement of the team coach. For gracious professionalism, teams are judged on their attitude and respect toward their own team members (especially younger ones) as well as their display of friendly competition (e.g., being willing to assist other teams). Relevant self-reported actions of the teams outside of the competition, such as mentoring a younger FLL team, are taken into account in this score.

Project Scores In addition to designing, building, and programming robots, each FLL team was also responsible for a research project, devising an innovative solution for a problem that they identified corresponding to the theme of the competition. The project topic for this competition was technology-enabled distance learning. The teams typically brought posters and/or showed videos to complement the project presentations they gave to the judges. The Project Scores were evaluated based on the societal value and clarity of the target issue, the innovation and creativity of the proposed solution, and the quality of the presentation, again using a detailed rubric provided by the FLL. Specifically, the research subscore was based on the quality of the problem identified, the quality of the sources of information used to solve the problem, the depth of analysis of the issue, and an extensive review of existing solutions. The

solution subscore was based on the value of the proposed solution, the originality of the application, and the comprehensiveness of the evaluation of an implementation for the solution.

Robot Design Robot Design scores involved mechanical design, programming, and strategy and innovation subscores based on an FLL-provided rubric. Mechanical design was evaluated based on the durability, efficiency, and mechanization of the physical robots. By contrast, the programming subscore was assessed in terms of three characteristics of the team's computer programs developed for controlling the robots: quality, efficiency, and autonomy (i.e., the robots require minimal to no driver intervention). Based on their performance in the interview, judges also gave a strategy and innovation subscore assessing the team's use of a good design process, the quality of the team's game strategy, and the innovative nature of the team's hardware and software solutions. The design process evaluation was based on how clearly the teams explained their design progression, obstacles, and solutions while developing their robots. Judges specifically evaluated the teams' ability to develop and explain the advancement of their design process in which possible solutions were developed, alternatives considered and reduced, selections tested, and designs improved as the teams engaged in multiple cycles of redesign process.

Collaboration Quality Score We developed a 10-min-long performance task and an accompanying rubric to evaluate the teams on a brief challenge task for the Collaboration Quality Score. The task required teams to outline on paper the structure of a computer program that would produce a reliable path for a robot to move from a start to a finish point, while successfully avoiding all obstacles on a given complex map (see Figure 1). This additional performance task was not explained to any of the teams beforehand. Program design tasks such as this are typical in the FLL, so all students would understand it in context. However, the specific challenge was novel to all teams, and, therefore, their performance on this task did not reflect rehearsed or highly coached behaviors. By contrast, for example, most teams had highly rehearsed research project presentations. Furthermore, the task did not involve the actual writing of code and thus was not biased by a particular programming language a given team was using.

Developed based on prior research on collaborative learning (e.g., Chi & Menekse, 2015; Kuhn, 2015), the rubric for judging Collaboration Quality involved a holistic judgment combining the amount of discussion, the depth of shared contributions building on one another's ideas, the elaboration of one another's ideas, the use of how and why questions in exploring one another's ideas, and the joint nature of the decisions (see Table 1).

Procedure

Data were collected at the FIRST LEGO League Western PA Championship, which included 61 teams, each provided with a strict schedule for their judging, thus ensuring approximately equal amounts of time for each team. Judging for Table, Project, and Robot Design scores was conducted by separate pools of judges based on FLL procedures. Our research team was responsible for evaluating the Collaboration Quality scores and served as volunteer judges for the Core Values scores upon request from the FLL tournament organizers.

The Collaboration Quality scores were evaluated by 12 volunteers, randomly divided in 6 pairs by our research team before the competition day; each pair evaluated approximately 10 robotics teams during the competition. Most of the 12 volunteers were graduate students who volunteered for the event, 10 of whom were blind to the study hypotheses. These volunteer judges received training on how to use the Collaboration Quality rubric (Table 1) before the competition, and they had no access to the other scores (Table Score, Robot Design, Project Score) that the robotics teams received.

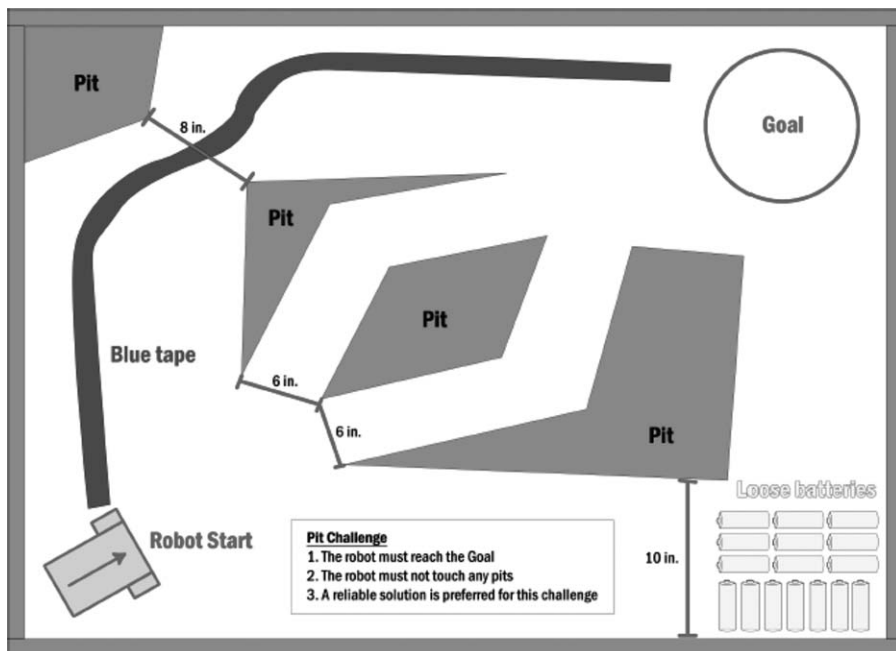


Figure 1 The map from the brief challenge task used to evaluate Collaboration Quality.

The Collaboration Quality assessment was administered as an auxiliary task after the Core Values judging. That is, each team arrived expecting to be judged only on Core Values. The team gave a presentation on the Core Values topic and then answered questions from the judges, a process that took approximately 10 min, and the judges provided Core Values scores based solely on this part. Then teams were given the collaboration task (Figure 1), which they worked on for approximately 10 min, while the volunteer judges observed their collaboration process using the Collaboration Quality rubric (Table 1) and assigning scores for each team independently. An overall Collaboration Quality Score for each team was obtained by averaging the two scores given by two judges in all cases. As no recording of students was permitted per IRB-approved protocol, the judges recorded their scores as the students worked. The average inter-rater reliability for the Collaboration Quality Score across the six pairs of judges was .84 (Cronbach's alpha), indicating a good consistency across judges when applying the scoring rubric (Stemler, 2004). The Cronbach's alpha values across the six pairs of judges ranged from .75 to .92, and in 69% of the ratings, the judging pairs exhibited perfect agreement.

All other judged events were scored in private rooms with only the judges, the team members, and a single adult observer from each team present. Most FLL judges were volunteers recruited from the local community who had received training in the events they were judging prior to attending the championship. Many had judged at previous qualifying events or in previous years. Judges were divided into teams of two to three, using a detailed rubric to evaluate each category. (These rubrics are publicly available at FLL websites.) Multiple teams judged each category, and the judges within a category (e.g., Robot Design) met after judging to roughly calibrate their scoring and to make any adjustments to their scores they felt were needed.

Table 1 Rubric for Judging Collaboration Quality in the Brief Challenge Task

Score 1	Score 2	Score 3
A <i>minimal</i> level of discussion. None or only one student generates detailed statements.	A <i>moderate</i> level of discussion. One student's statements are mostly substantive and the others' vary between detailed and shallow.	A <i>substantial</i> level of discussion. Substantive statements of each student build upon those of others, indicating a shared line of reasoning.
Students do not clarify or complete their partners' statements, instead voicing generic responses of agreement.	Statements are discontinuous as each student makes assertions independent from those of others.	Students clarify or complete their peers' statements through expanding, elaborating, restatement, or rebuttal.
One student decides what to write while the others agree but contribute very little.	One student contributes most to what will be written while the others take a smaller, though substantive, role.	Conclusions are jointly constructed with two or more students involved fairly equally in determining what to write.
None of the students ask why/how type questions, discuss each others' claims, or elaborate in response to questions.	Some students effectively engage in the collaboration process. A few why/how type questions are asked and discussed.	Most students effectively engage in the collaboration process. More than one type of why/how questions is asked and discussed.

The Judge Advisor, a key volunteer, oversaw the judging process, leading the judging team and working with the tournament organizers to ensure that the event met judging standards for a sanctioned FLL event. The decisions and final scores for the three judged categories (Project, Robot Design, Core Values) were awarded by a consensus among the judges and the Judge Advisor, meaning there was only one final score for each of these three categories. Therefore, it was not possible to calculate inter-rater reliability values for these categories.

Table Scores were determined by the performance of the robots in a public setting and recorded and verified by volunteer referees who were trained and supervised similarly to the other judges. Teams were able to dispute referee scoring decisions at the time of marking, prior to the board being reset for the next round. Every team competed in three scored rounds of competition; data from all three rounds were used in the analysis here. Final scoring results for judged and tabletop events were obtained from the tournament organizers following the conclusion of the competition.

Analysis

To address the first research question, we began by calculating the Pearson product-moment correlation coefficient (Pearson correlation) to assess the degree to which the variables were linearly related. Next, we conducted multiple linear regression analyses to evaluate how well the Collaboration Quality, Core Values, Project, and Robot Design scores predicted the Table Scores. We calculated both partial and semipartial correlation coefficients, which provide information on the relative importance of independent variables on a dependent variable. Partial correlation is a measure of the strength of a linear relationship between two variables while controlling for the effect of one or more other variables, and the semipartial indicates the unique contribution of an independent variable. Specifically, the squared semipartial correlation indicates how much R^2 would change if that variable were removed from the

regression equation (Cohen, 1988). We used IBM SPSS Statistical software (version 24) for all statistical analysis.

As a follow-up to the regression analyses, we conducted mediation analyses using the Hayes statistical mediation analysis approach (Hayes, 2009) and his PROCESS macro for SPSS (Hayes, 2013). The goal was to explore whether Robot Design was a mediator for the relation between Collaboration Quality and Table scores. Next, we conducted additional linear regressions to further delineate the relationship between Collaboration Quality and the subcomponents of Robot Design, Mechanical Design, Programming, and Strategy and Innovation scores.

To address the second research question, we conducted a one-way analysis of variance (ANOVA) to explore the relationship between Collaboration Quality and team experience by using team-level data, specifically the team numbers indicating the team experience and Collaboration Quality scores for each team. In addition, we conducted a second one-way ANOVA using individual-level data, specifically the participants' ages, to explore the difference, if any, across team members in terms of their ages as a possible alternative explanation for the effects of team experience.

Results

Collaboration Quality and Overall Team Performance

Because the primary goal of this study was to examine the effect of Collaboration Quality on team performance, our first set of analyses focused on investigating the relationship between each team's Collaboration Quality scores and the other performance variables. Pearson correlation coefficients were statistically significant for all performance variables (see Table 2), ranging from moderate to large correlations (Hemphill, 2003), establishing that all measures had adequate reliability. These results indicate that strong collaboration skills on the short performance task predicted the team would do well on other long-term performance variables. From statistics and measurement theory, the expected strength of an observed correlation is equal to the true correlation between the constructs multiplied by the reliability of each measure. If a measure has close to zero reliability, no significant correlation would be observed. Further, since in education and social sciences, essentially all outcomes are determined in multiple ways, strong correlations are pragmatic evidence of the reliability of the measures. (Note that this evidence does not establish validity of the measure.)

The strongest correlation observed was between the Collaboration Quality and Project scores. Interestingly, we expected to see the strongest correlation between the Core Values and the Collaboration Quality scores, for two reasons. First, Core Values can be considered as arguably conceptually the closest match to collaboration skill. In the context of our study, the Core Values category is the most closely related to our Collaboration Quality scores because both are indicators of teamwork behaviors to some degree and were evaluated by observing teams during real-time actions. Second, Core Values is potentially prone to "halo effects" in judging. The halo effect refers to the cognitive bias resulting when the overall impression of certain people influences how evaluators feel and characterize most of the behaviors of those people. Based on the literature, the halo effect was expected to have a certain effect on the judges' ratings of the teams' behaviors. Table 3 shows the means and standard deviations for all scores.

Next, we focused on the robotics performance task as measured by the Table Score. More specifically, we conducted multiple linear regression analyses to evaluate how well the

Table 2 Pearson Correlations Between the Collaboration Quality Score and Other Measures

	Table Score	Core Values	Project Score	Robot Design
Collaboration Quality	.39**	.49***	.55***	.46***

** $p < .01$, *** $p < .001$.

Collaboration Quality, Core Values, Project, and Robot Design scores predicted the Table scores of the teams. In principle, Robot Design reflected the quality of the robots used for competition, whereas Core Values and Collaboration Quality indicated the process constructs that would lead to teams developing more effective robots in terms of software and hardware for the competition, and Project scores represented general academic skills and supports.

The linear combination was significantly related to the average Table Score, $F(4,56) = 13.91, p < .001, R^2 = .50$, with all bivariate correlations between the average Table Score and other predictors being statistically significant. However, the only significant partial correlation was between the Robot Design Score and the average Table Score, indicating that the Robot Design Score for each team is the best predictor for the average Table Score, as expected (Table 4).

We followed this analysis by examining whether the Pearson correlation between Collaboration Quality and Table Scores is indeed explained by the Robot Design scores (i.e., whether Robot Design was a mediator). Mediation occurs when one factor predicts the value of a second factor, the second subsequently predicting the value of a third, with this indirect pathway significantly reducing the relationship between the first and third variables. In such cases, the first factor does, in fact, predict the third, but primarily indirectly as the second variable serves as a “mediator” joining the other two. We used the Hayes statistical mediation analysis approach, which employs an ordinary least square (or logistic) regression path analysis to estimate direct and indirect relationships in mediation models through bootstrap confidence intervals. In our analysis, we used 1,000 bootstrap samples to explore the indirect relationship in our mediation model. This analysis found that Robot Design acted as a full mediator for the predictive effect of Collaboration Quality on the Table Score (see Figure 2). More specifically, we found a significant indirect relationship between Collaboration Quality through Robot Design quality to average Table Score performance, $(.46) * (.66) = .30$, 95% Confidence Interval = [.15, .52]. This mediation pathway accounts for 78% (.30/.39) of the total relationship between collaboration and the Table Score, with the remaining direct relationship (.09) not being statistically significant. Overall, these results show Collaboration Quality is predictive of Table scores only to the extent to which it predicts the Robot Design Score.

Table 3 Means and Standard Deviations of Each Performance Score Across All Teams ($N = 61$)

Measure	Mean	Standard deviation	Possible maximum score
Table Score (aka Robot Performance)	99.9	63.2	858 (theoretical) 433 (observed)
Core Values	3.00	0.64	4.00
Project Score	2.50	0.98	4.00
Robot Design	2.49	0.84	4.00
Collaboration Quality	2.16	0.60	3.00

Table 4 Bivariate, Partial, and Semipartial Correlations of the Predictors with Table Scores

Predictors	Bivariate correlation	Partial correlation	Semipartial correlation
Collaboration Quality	.39**	.14	.10
Core Values	.30**	.01	.01
Project	.24*	-.12	-.09
Robot Design	.70***	.63***	.57***

* $p < .05$, ** $p < .01$, *** $p < .001$.

Finally, we investigated whether Collaboration Quality predicted the individual subcomponents of Robot Design: Mechanical Design, Programming, and Strategy and Innovation. If the FLL task follows a similar value structure as robotics engineering itself, we would expect Collaboration Quality to be predictive of all three. A series of linear regressions show that, as expected, Collaboration Quality was a significant predictor of all three of the Robot Design subdimensions individually, although to different degrees. A one-point increase in Collaboration Quality predicted a 0.25 point increase in the Mechanical subscore [$F(1,59) = 8.638, p = .005$], a 0.39 point increase in the Programming subscore [$F(1,59) = 15.459, p < .001$], and a 0.39 point increase in the Strategy and Innovation subscore [$F(1,59) = 15.100, p < .001$]. In other words, Collaboration Quality appeared to predict success in all main aspects of designing the robots, but more strongly with software design and creativity than with hardware design.

Collaboration Quality and Team Programming Performance

Since robotics competitions are considered an important opportunity for learning programming and Collaboration Quality is more strongly related to programming than physical design, we further explored the relationship between Collaboration Quality and team programming performance. It is possible that collaboration may simply produce successful but not elegant code, or alternately, it may help lead to higher quality code (e.g., more comments, more modular code). The quality of code was evaluated based on the Programming Score in the FLL context, which included three components: (a) programming efficiency (Modular, streamlined, understandable code containing no unnecessary commands); (b) programming quality (Code matches design intent and behaves consistently in a reliable way); and (c)

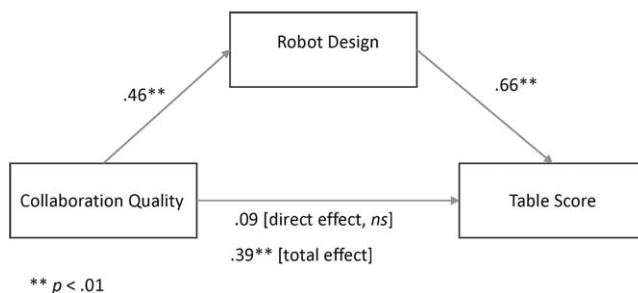


Figure 2 The relationship between Collaboration Quality and Table Score mediated by Robot Design Score. Standardized regression coefficients are shown.

autonomous features of the robot's movement (Program uses sensors and error-correction algorithms to reduce reliance on human interaction.). We conducted linear regressions to evaluate the strength of the relationship between the Collaboration Quality Score and each programming subscore.

Collaboration Quality predicted each of the three programming subdimensions at approximately similar levels. Specifically, a one-point increase in Collaboration Quality predicted a 0.34 point increase in Programming Efficiency [$F(1,57) = 9.822, p = .003$], a 0.37 point increase in Programming Quality [$F(1,57) = 13.803, p < .001$], and a 0.36 point increase in Autonomous Features of Movement [$F(1,57) = 12.382, p = .001$].

Collaboration Quality and Team Experience

The dataset used in this study also included indirect information about the number of years of participation in competitions by each attending robotics team. The team identification number in the FLL competitions is based on when the team was first created, with smaller numbers reflecting more years in existence. This number is assigned by the FLL Team Information Management System (TIMS) when teams registered for their first FLL event, with the same number being used in different FLL tournaments and competitions throughout the teams' existence. Although the team membership changes annually, with students joining and leaving teams, there is often a high enough percentage of returning members from year-to-year such that each team develops successful routines and knowledge that is carried over from one year to the next. It is possible that Collaboration Quality is influenced by this transferred knowledge. In our dataset, team identification numbers ranged from two digit numbers (e.g., 30) to five digit numbers (e.g., 13,000), and based on these identification numbers, we divided the 61 participating teams into three categories: most experienced, experienced, and least experienced teams. More fine-grained distinctions are less likely to be meaningful as they simply reflect the order in which a team registered within a competition year.

Our goal was to explore whether the teams' cumulative amount of competition experience was associated with team Collaboration Quality; Table 5 indicates the means and standard deviations of Collaboration Quality scores for the three experience levels. Using a one-way ANOVA, the relationship between experience and Collaboration Quality was found to be significant [$F(2,58) = 4.48, p = .01, R^2 = .13$]. Follow-up post hoc Tukey HSD (honest significant difference) tests revealed that the least experienced teams had significantly lower Collaboration Quality scores than either the experienced [95% CI = 0.11 to 1.75*] or most experienced teams [95% CI = 0.02 to 1.86*]. The difference between most experienced and experienced teams was very small and not statistically significant [95% CI = -0.85 to 0.87].

Furthermore, we also investigated whether there was a difference in the participants' ages across most experienced, experienced, and least experienced teams to eliminate a possible confound that participants on more experienced teams were older than their peers on other teams. While we had Collaboration Quality and team performance data for all 61 teams (with 366 individual participants), we had participant age data for 148 individuals since the individual background survey was optional. Among these 148 participants, 61 participants were from the most experienced, 41 participants were from the experienced, and 46 participants were from the least experienced teams. We conducted a one-way ANOVA to further explore the difference, if any, in terms of age, with the results showing no significant difference across experience levels, $F(2,145) = 1.95, p = .15$. Thus, the relationship between team experience and Collaboration Quality is not likely the result of more experienced teams having older team members.

Table 5 Means and Standard Errors of Collaboration Quality Scores Across Three Groups Based on Cumulative Amount of Team Competition Experience

Cumulative competition experience	<i>N</i>	Mean	Standard error
Most experienced	16	2.32	0.11
Experienced	26	2.30	0.09
Least experienced	19	1.84	0.17

General Discussion

This study explored the relationship between Collaboration Quality among robotics competition team members and various measures of team performance in the competition, including both objective task performance and expert judge evaluations for a diverse set of supporting performance dimensions. Here, collaboration was independently assessed using an unexpected new performance task that, thus, could not be biased by coaching nor direct preparation. The regression analyses revealed that the performance assessment of Collaboration Quality was a good indicator of a robotics team's overall performance across all measures, that is, the objective performance on the competition task as well as expert ratings of success along the other main dimensions (Core Values, Project Quality, and Robot Design).

Focusing on the objective competition task performance, further analysis showed that the relationship between Collaboration Quality and competition performance appeared to be mediated by the relationship with Robot Quality. In other words, these results suggest that students taking part in the FLL problem-based learning activity were engaged in an authentic engineering process in which good teamwork produced superior products, which then scored high in the competition. This pattern of results goes beyond the narrower possible explanation that good collaboration dynamics influenced only how well the teams were able to run their robots on the day of the competition. Collaboration Quality also predicted the individual dimensions of Robot Quality (Mechanical, Programming, and Strategy and Innovation) and the individual subdimensions scores of Programming (Programming Efficiency, Programming Quality, and Autonomous Movement Features). This broad predictability suggests that the relationships between performance and collaboration are not localized with respect to a single area of design, but are present throughout. In conjunction with the earlier result that Collaboration Quality was also associated with Core Values and Research Project scores, these findings suggest that the benefits of collaboration may be both broad and deep with respect to robotics competition tasks. This mediation and the simultaneous support for Hypothesis 1 (Collaboration Quality predicts Robot Quality and Robot Quality predicts Table Score) also suggest that the competition task authentically captures an underlying value that collaboration is argued to play in engineering.

Finally, the idea that team experience was significantly related to Collaboration Quality (Hypothesis 2) is consistent with the idea that collaboration is indeed a skill that can be developed and may be partly constituted at the organizational level as more experienced teams engaged in more substantial levels of discussion compared to the students on less experienced teams, regardless of students' ages or individual experience.

Limitations

The findings of this study are limited by a number of factors. First, since the data and analyses are fundamentally correlational in nature, causal conclusions cannot be drawn about either the relationship between Collaboration Quality and robotics team performance, or between

team experience and Collaboration Quality. However, the multiple regression analyses rule out a third variable explanation for the relationships observed. Furthermore, that Robot Design, rather than research Project Score, was the apparent mediator rules out a general intelligence/broad motivational confound in which gifted or highly motivated teams score higher on all performance dimensions. Nonetheless, future research is needed to examine Collaboration Quality at multiple time points or to consider interventions aimed at Collaboration Quality to more directly assess questions of causality.

Second, although our sample size was large by comparison to past research, it is still relatively modest for multiple regression since all of our analysis is at the team level (61 teams and 366 individuals). Resulting reductions in power could have prevented our finding a significant mediation of Collaboration Quality to Tables scores by Core Values scores (which included self-reported teamwork) in addition to mediation by Robot Design scores. Theoretically, both factors should have contributed to this mediation. It is worth emphasizing that the power of the current study of teams was relatively strong compared to past studies as it is unusual to find datasets of large numbers of teams working on a shared complex performance task.

Third, the validity of some of the expert scores is limited since reliability measurements were not obtainable and scores could have been biased by outside factors. In addition, since the judges were volunteers and different groups of judges evaluated different teams, the expert scores might be subject to reduced reliability. This reduced reliability could weaken the effects observed; however, our analyses found robust effects, minimizing the possible effects of bias on the reliability of the measures of performance and quality. Further, the FLL training, rubrics, and separate scoring procedures by domain provide face validity to the measures. Finally, the context of an informal setting (vs. a formal classroom setting where students are required to attend) could introduce bias into the study as students self-select to participate.

Implications for the Design of Learning Experiences and Future Research

The mediation pattern found here indicating that collaboration skill predicts Robot Design quality, which, in turn, predicts the competition game board scores, suggests that the task design used in the FLL has an underlying value structure that favors team collaboration: better collaboration is rewarded with better competition outcomes. It is also consistent with a constructionist design pattern in which the necessity for collaboration provides an opportunity to learn to collaborate better (Rummel and Spada, 2005; Tsai, 2010). In addition, the fact that the predictive reach of Collaboration Quality extended down to the level of detail like programming efficiency suggests that success in FLL competition is thoroughly and robustly tied to collaboration at many levels rather than through a single, brittle link. These findings support the collaborative learning literature, which has also found that the quality of interaction, such as asking in-depth questions, requesting explanations, and co-constructing knowledge, facilitates learning (Chi, 2009; Jeong & Chi, 2007; Volet, Summers, & Thurman, 2009).

However, it remains unclear exactly what specific features of this competition lead to the outcomes observed: Was it the design of the game board tasks, or perhaps the cultural emphasis on teamwork as a core value among participants? Future research could examine the role of each feature of the robotics competitions in necessitating, enabling, and motivating collaborative activity among participants to inform the development of future productive collaborative design activities. Similarly, our study examined effects within one especially popular middle school-level competition. Further research is needed to examine whether similar effect structures are observed in other contexts with different competition structures involving different age groups and populations. In addition, since the students self-select to participate in

robotics competitions, it is important to explore how the context of an informal setting and student motivation to participate on robotics teams influence their collaboration behaviors.

The fact that the level of Collaboration Quality was higher among more experienced teams suggests that taking part in the FLL may indeed be building collaboration skill rather than simply rewarding it. Since those older teams have been in existence longer than any individual could have been a member (due to competition age restrictions), at least some of the knowledge and skills pertaining to collaboration are being stored or passed down within the team units. Robotics competitions may, therefore, be generating smaller, stable communities of practice through their operation (Lave & Wenger, 1991). Future research could further elaborate on this possibility by examining the means by which this persistence of skill occurs and how that expertise is stored and passed down within the team—codified as rules, embedded in norms, written in documents, implicit in routines, or through dispositional change.

Finally, the kind of collaboration skill we measured is a transferrable type. Our measure for collaboration skill was a performance assessment which inherently required that students transfer their learning about collaboration from the competition task to our task. Since the competition tasks and context are nominally similar to engineering, it is also likely that the problem-solving and collaboration skills demonstrated may exhibit transfer to other, more distal tasks as well, although future research should examine this matter empirically.

Acknowledgments

We are grateful to the FLL teams, participants, coaches, volunteer judges, and organizers who made this research possible. We also thank the *Journal of Engineering Education* reviewers and editors for their helpful comments and questions on previous drafts. This research is based on work supported by the National Science Foundation (NSF) under Grant DRL-1416984. Any opinions, findings, and conclusions expressed in this material are those of the authors and do not necessarily reflect those of NSF.

References

- Alfieri, L., Higashi, R., Shoop, R., & Schunn, C. D. (2015). Case studies of a robot-based game to shape interests and hone proportional reasoning skills. *International Journal of STEM Education*, 2(1), 1–13. doi: 10.1186/s40594-015-0017-9
- Barron, B. (2000). Achieving coordination in collaborative problem-solving groups. *Journal of the Learning Sciences*, 9(4), 403–436. doi: 10.1207/S15327809JLS0904_2
- Barron, B. (2003). When smart groups fail. *Journal of the Learning Sciences*, 12, 307–359. doi: 10.1207/S15327809JLS1203_1
- Barrows, H. S. (1996). Problem-based learning in medicine and beyond: A brief overview. *New Directions for Teaching and Learning*, 68, 3–12. doi: 10.1002/tl.37219966804
- Barrows, H. (2002). Is it truly possible to have such a thing as dPBL? *Distance Education*, 23(1), 119–122. doi: 10.1080/01587910220124026
- Bascou, N. A., & Menekse, M. (2016). Robotics in K-12 formal and informal learning environments: A review of literature. *Proceedings of 2016 American Society for Engineering Education Annual Conference*, New Orleans, Louisiana. doi: 10.18260/p.26119
- Benitti, F. B. V. (2012). Exploring the educational potential of robotics in schools: A systematic review. *Computers & Education*, 58(3), 978–988. doi: 10.1016/j.compedu.2011.10.006
- Borrego, M., Karlin, J., McNair, L. D., & Beddoes, K. (2013). Team effectiveness theory from industrial and organizational psychology applied to engineering student project

- teams: A research review. *Journal of Engineering Education*, 102(4), 472–512. doi: 10.1002/jee.20023
- Brown, J. S., & Duguid, P. (1998). Organizing knowledge. *California Management Review*, 40(3), 90–111. doi: 10.2307/41165945
- Chi, M. T. (2009). Active-constructive-interactive: A conceptual framework for differentiating learning activities. *Topics in Cognitive Science*, 1(1), 73–105. doi: 10.1111/j.1756-8765.2008.01005.x
- Chi, M. T. H., & Menekse, M. (2015). Dialogue patterns in peer collaboration that promote learning. In L. B. Resnick, C. Asterhan, & S. N. Clarke (Eds.), *Socializing intelligence through academic talk and dialogue* (pp. 263–274). Washington, DC: AERA.
- Clark, D., Sampson, V., Stegmann, K., Marttunen, M., Kollar, I., Janssen, J., Weinberger, A., Menekse, M., Erkens, G., & Laurinen, L. (2010). Online learning environments, scientific argumentation, and 21st century skills. In B. Ertl (Ed.), *E-Collaborative knowledge construction: Learning from computer-supported and virtual environments* (Ch.1, pp. 1–39). IGI Global: Hershey.
- Close, D. (2015). VS students play well, learn well with Legos while advancing to state FLL event. Retrieved from <http://www.vintonyesterday.org/articles/News/article1016283.html>
- Cohen J. (1988). *Statistical power analysis for the behavioral sciences*. Hillsdale, NJ: Erlbaum.
- Core Values. (2016, November 5). Retrieved from <https://www.firstinspires.org/robotics/fll/core-values>
- Danahy, E., Wang, E., Brockman, J., Carberry, A., Shapiro, B., & Rogers, C. B. (2014). Lego-based robotics in higher education: 15 years of student creativity. *International Journal of Advanced Robotic Systems*, 11, 27. doi: 10.5772/58249
- Denis, B., & Hubert, S. (2001). Collaborative learning in an educational robotics environment. *Computers in Human Behavior*, 17, 465–480. doi: 10.1016/S0747-5632(01)00018-8
- Dolmans, D. H., De Grave, W., Wolfhagen, I. H., & Van Der Vleuten, C. P. (2005). Problem-based learning: Future challenges for educational practice and research. *Medical education*, 39(7), 732–741.
- Eguchi, A. (2014). Educational robotics for promoting 21st century skills. *Journal of Automation Mobile Robotics and Intelligent Systems*, 8(1), 5–11. http://dx.doi.org/10.1207/S1532690XCI2004_1
- Enyedy, N., & Stevens, R. (2014). Analyzing collaboration. In R. Keith Sawyer (Ed.), *The Cambridge handbook of the learning sciences* (pp. 191–212). Cambridge handbooks in psychology (2nd ed.). Cambridge: Cambridge University Press.
- Fagin, B. S., & Merkle, L. (2002). Quantitative analysis of the effects of robots on introductory computer science education. *ACM Journal on Educational Resources in Computing*, 2(4), 1–18. doi: 10.1145/949257.949259
- Falk, J. H., & Dierking, L. D. (2000). *Learning from museums: Visitor experiences and the making of meaning*. Walnut Creek, CA: Alta Mira Press.
- Grover, S., & Pea, R. (2013). Computational thinking in K-12: A review of the state of the field. *Educational Researcher*, 42(1), 38–43. doi: 10.3102/0013189X12463051
- Hayes, A. F. (2009). Beyond Baron and Kenny: Statistical mediation analysis in the new millennium. *Communication Monographs*, 76, 408–420. doi: 10.1080/03637750903310360
- Hayes, A.F. (2013). *Introduction to mediation, moderation, and conditional process analysis: A regression-based approach*. New York, NY: The Guilford Press.
- Hemphill, J. F. (2003). Interpreting the magnitude of correlation coefficients. *American Psychologist*, 58(1), 78–79. doi: 10.1037/0003-066X.58.1.78

- Hmelo-Silver, C. E. (2004). Problem-based learning: What and how do students learn? *Educational Psychology Review*, 16(3), 235–266. doi: 10.1023/B:EDPR.0000034022.16470.f3
- Hmelo-Silver, C. E., Duncan, R. G., & Chinn, C. A. (2007). Scaffolding and achievement in problem-based and inquiry learning: A response to Kirschner, Sweller, and Clark (2006). *Educational Psychologist*, 42(2), 99–107. doi: 10.1080/00461520701263368
- Hogan, K., Nastasi, B. K., & Pressley, M. (1999). Discourse patterns and collaborative scientific reasoning in peer and teacher-guided discussions. *Cognition and Instruction*, 17(4), 379–432. doi: 10.1207/S1532690XCI1704_2
- Jeong, H. (2013). Verbal data analysis for understanding interactions. In C. Hmelo-Silver, A. M. O'Donnell, C. Chan, & C. Chinn (Eds.), *The international handbook of collaborative learning* (pp. 168–183). London: Taylor and Francis.
- Jeong, H., & Chi, M. T. H. (2007). Knowledge convergence and collaborative learning. *Instructional Science*, 35, 287–315. doi: 10.1007/s11251-006-9008-z
- Johnson, R. T., & Londt, S. E. (2010). Robotics competitions: The choice is up to you! *Tech Directions*, 69(6), 16–20.
- Jordan, M. E., & McDaniel, R. R., Jr. (2014). Managing uncertainty during collaborative problem solving in elementary school teams: The role of peer influence in robotics engineering activity. *Journal of the Learning Sciences*, 23, 490–536. doi: 10.1080/10508406.2014.896254
- Kim, C., Kim, D., Yuan, J., Hill, R. B., Doshi, P., & Thai, C. N. (2015). Robotics to promote elementary education pre-service teachers' STEM engagement, learning, and teaching. *Computers & Education*, 91, 14–31. doi: 10.1016/j.compedu.2015.08.005
- Koenig, J. A. (Ed.). (2011). *Assessing 21st century skills: Summary of a workshop*. Washington, DC: National Academies Press.
- Kuhn, D. (2015). Thinking together and alone. *Educational Researcher*, 44, 146–153. doi: 10.3102/0013189X15569530
- Lave, J., & Wenger, E. (1991). *Situated learning: Legitimate peripheral participation*. Cambridge: Cambridge University Press.
- Martinez Ortiz, A. (2011). Fifth grade students' understanding of ratio and proportion in an engineering robotics program. Paper presented at the 2011 American Society for Engineering Education Conference, Vancouver, British Columbia, Canada. Retrieved from <http://www.asee.org/public/conferences/1/papers/2649/view>
- Mehalik, M. M., Doppelt, Y., & Schunn, C. D. (2008). Middle-school science through design-based learning versus scripted inquiry: Better overall science concept learning and equity gap reduction. *Journal of Engineering Education*, 97(1), 71–85. doi: 10.1002/j.2168-9830.2008.tb00955.x
- Melchior, A., Cohen, F., Cutter, T., & Leavitt T. (2005). *More than robots: An evaluation of the FIRST robotics competition participant and institutional impacts*. Waltham, MS: Brandeis University Center for Youth and Communities.
- Menekse, M., Stump, G., Krause, S., & Chi, M. T. H. (2013). Differentiated overt learning activities for effective instruction in engineering classrooms. *Journal of Engineering Education*, 102(3), 346–374. doi: 10.1002/jee.20021
- Menekse, M. (2015). Computer science teacher professional development in the United States: A review of studies published between 2004 and 2014. *Computer Science Education*. doi: 10.1080/08993408.2015.1111645
- Mohr-Schroeder, M. M., Jackson, C. J., Miller, M., Walcott, B. W., Little, D. L., Speler, L. S., & Schooler, W. S. (2014). Developing middle school students' interests in STEM via summer

- learning experiences: See Blue STEM Camp. *School Science and Mathematics*, 114(6), 291–301. doi: 10.1111/ssm.12079
- Nokes-Malach, T. J., Richey, J. E., & Gadgil, S. (2015). When is it better to learn together? Insights from research on collaborative learning. *Educational Psychology Review*, 27(4), 645–656. doi: 10.1007/s10648-015-9312-8
- Nugent, G., Barker, B., Grandgenett, N., & Adamchuk, V. (2009). The use of digital manipulatives in K-12: Robotics, GPS/GIS and programming. In *Proceedings of the 39th IEEE Frontiers in Education Conference 2009 (FIE'09)*. 1–6. San Antonio, TX. doi: 10.1109/FIE.2009.5350828
- Okita, S.O. (2014). The relative merits of transparency: Investigating situations that support the use of robotics in developing student learning adaptability across virtual and physical computing platforms. *British Journal of Educational Technology*, 45(5), 844–862. doi: 10.1111/bjet.12101
- Petre, M., & Price, B. (2004). Using robotics to motivate 'back door' learning. *Education and information technologies*, 9(2), 147–158.
- Purzer, Ş. (2011). The relationship between team discourse, self-efficacy, and individual achievement: A sequential mixed-methods study. *Journal of Engineering Education*, 100(4), 655–679. doi: 10.1002/j.2168-9830.2011.tb00031.x
- Puvirajah, A., Verma, G., & Webb, H. (2012). Examining the mediation of power in a collaborative community: Engaging in informal science as authentic practice. *Cultural Studies of Science Education*, 7(2), 375–408. doi: 10.1007/s11422-012-9394-2
- Rummel, N., & Spada, H. (2005). Learning to collaborate: An instructional approach to promoting collaborative problem solving in computer-mediated settings. *Journal of the Learning Sciences*, 14(2), 201–241. doi: 10.1207/s15327809jls1402_2
- Schmidt, H. G., Rotgans, J. I., & Yew, E. H. (2011). The process of problem-based learning: What works and why. *Medical Education*, 45(8), 792–806.
- Shin, J. H., Haynes, R. B., & Johnston, M. E. (1993). Effect of problem-based, self-directed undergraduate education on life-long learning. *CMAJ: Canadian Medical Association Journal*, 148(6), 969–976.
- Shuman, L. J., Besterfield-Sacre, M., & McGourty, J. (2005). The ABET "professional skills"—Can they be taught? Can they be assessed? *Journal of Engineering Education*, 94(1), 41–55. doi: 10.1002/j.2168-9830.2005.tb00828.x
- Stahl, G. (2005). Group cognition in computer-assisted collaborative learning. *Journal of Computer Assisted Learning*, 21(2), 79–90. doi: 10.1111/j.1365-2729.2005.00115.x
- Stemler, S. E. (2004). A comparison of consensus, consistency, and measurement approaches to estimating interrater reliability. *Practical Assessment, Research & Evaluation*, 9(4), 1–19.
- Stump, G. S., Hilpert, J. C., Husman, J., Chung, W. T., & Kim, W. (2011). Collaborative learning in engineering students: Gender and achievement. *Journal of Engineering Education*, 100(3), 475–497. doi: 10.1002/j.2168-9830.2011.tb00023.x
- Sullivan, F. R. (2008). Robotics and science literacy: Thinking skills, science process skills, and systems understanding. *Journal of Research in Science Teaching*, 45(3), 373–394. doi: 10.1002/tea.20238
- Sweller, J., Kirschner, P. A., & Clark, R. E. (2007). Why minimally guided teaching techniques do not work: A reply to commentaries. *Educational Psychologist*, 42(2), 115–121. doi: 10.1080/00461520701263426
- Tiwari, A., Lai, P., So, M., & Yuen, K. (2006). A comparison of the effects of problem-based learning and lecturing on the development of students' critical thinking. *Medical Education*, 40(6), 547–554. doi: 10.1111/j.1365-2929.2006.02481.x

- Tonso, K. L. (2006). Teams that work: Campus culture, engineer identity, and social interactions. *Journal of Engineering Education*, 95(1), 25–37.
- Tsai, C. W. (2010). Do students need teacher's initiation in online collaborative learning? *Computers & Education*, 54(4), 1137–1144. doi: 10.1016/j.compedu.2009.10.021
- Verma, G., Puvirajah, A., & Webb, H. (2015). Enacting acts of authentication in a robotics competition: An interpretivist study. *Journal of Research in Science Teaching*, 52, 268–295. doi: 10.1002/tea.21195
- Verner, I. M., & Ahlgren, D. J. (2004). Robot contest as a laboratory for experiential engineering education. *Journal on Educational Resources in Computing (JERIC)*, 4(2), 1–15. doi: 10.1145/1071620.1071622
- Volet, S., Summers, M., & Thurman, J. (2009). High-level co-regulation in collaborative learning: How does it emerge and how is it sustained? *Learning and Instruction*, 19(2), 128–143. doi: 10.1016/j.learninstruc.2008.03.001
- Vye, N. J., Goldman, S. R., Voss, J. F., Hmelo, C., & Williams, S. (1997). Complex mathematical problem solving by individuals and dyads. *Cognition and Instruction*, 15(4), 435–484. doi: 10.1207/s1532690xci1504_1
- Webb, N. M. (1989). Peer interaction and learning in small groups. *International Journal of Educational Research*, 13, 21–40. doi: 10.1016/0883-0355(89)90014-1
- Welch, A., & Huffman, D. (2011). The effect of robotics competitions on high school students' attitudes toward science. *School Science and Mathematics*, 111, 416–424. doi: 10.1111/j.1949-8594.2011.00107.x
- White, C. B. (2007). Smoothing out transitions: How pedagogy influences medical students' achievement of self-regulated learning goals. *Advances in Health Sciences Education*, 12(3), 279–297.
- Williams, D., Ma, Y., Prejean, L., Ford, M. J., & Lai, G. (2007). Acquisition of physics content knowledge and scientific inquiry skills in a robotics summer camp. *Journal of Research on Technology in Education*, 40, 201–216. doi: 10.1080/15391523.2007.10782505
- Yew, E. H., Chng, E., & Schmidt, H. G. (2011). Is learning in problem-based learning cumulative? *Advances in Health Sciences Education*, 16(4), 449–464.

Authors

Muhsin Menekse is an Assistant Professor at Purdue University, with a joint appointment in the School of Engineering Education and the Department of Curriculum and Instruction, Neil Armstrong Hall of Engineering, 701 W. Stadium Avenue, West Lafayette, IN, 47907; menekse@purdue.edu.

Ross Higashi is a Graduate Student Researcher in the Learning Sciences and Policy Program at the Learning Research and Development Center at the University of Pittsburgh, 3939 O'Hara Street, Pittsburgh, PA 15260; rmh57@pitt.edu.

Christian Schunn is a Senior Scientist at the Learning Research and Development Center and a Professor of Psychology at the University of Pittsburgh, 821 LRDC, 3939 O'Hara Street, Pittsburgh, PA 15260; schunn@pitt.edu.

Emily Baehr is a Clinical Research Coordinator for the Department of Medicine and Gastroenterology at the University of Pittsburgh, 200 Lothrop Street, Pittsburgh, PA 15213; ecb42@pitt.edu.