The role of collaborative interactions versus individual construction on students’ learning of engineering concepts

Muhsin Menekse & Michelene T. H. Chi

To cite this article: Muhsin Menekse & Michelene T. H. Chi (2018): The role of collaborative interactions versus individual construction on students’ learning of engineering concepts, European Journal of Engineering Education

To link to this article: https://doi.org/10.1080/03043797.2018.1538324

Published online: 06 Nov 2018.

Submit your article to this journal

View Crossmark data
The role of collaborative interactions versus individual construction on students’ learning of engineering concepts

Muhsin Menekse and Michelene T. H. Chi

ABSTRACT
This study primarily investigated the role of interactional factors in an unstructured face-to-face collaborative learning environment with challenging engineering activities. We explored dialogue patterns in terms of quality of interaction, students’ scaffolding instances, and discourse moves for productive interactions of collaborative dyads in the context of the Interactive-Constructive-Active-Passive (ICAP) framework. The sample included 72 engineering students for the interactive and constructive conditions. Students’ understanding of material science and engineering concepts were measured using pre and posttest design. Results showed students in the interactive condition performed significantly better than students in the constructive condition. Verbal analysis of approximately 12 hours video recordings and 210 pages of transcriptions for students’ dialogue in the interactive condition indicated a strong relation between the quality of interaction, scaffolding instances, and individual learning gains. In addition, a verbal analysis examining each utterance based on the discourse moves revealed that the certain moves are significantly linked with learning outcomes.

ARTICLE HISTORY
Received 19 May 2017
Accepted 15 October 2018

KEYWORDS
Engineering education research; teamwork; assessment of learning outcomes; active learning; peer learning/teaching

1. Introduction
As the growing body of research studies has explored the benefits of collaborative learning in different settings, peer collaboration has become a central part of educational practices across domains and age groups. Some of the cognitive benefits that were illustrated in the collaborative learning studies included comprehension of ideas, retention of knowledge, integration of new and old knowledge, enhanced reasoning and higher order thinking, and improvement in transfer of knowledge (Dillenbourg 1999; Gijlers and de Jong 2013; King 2002; Kuhn 2015; O’Donnell and Hmelo-Silver 2013; Puntambekar 2006; Stahl and Hesse 2009; Volet, Summers, and Thurman 2009). Many studies have hypothesised that acquiring these cognitive benefits is one of the key reasons for academic achievement, especially for complex tasks (e.g. fundamental engineering concepts) that require higher level reasoning (e.g. Kaartinen and Kumpulainen 2002; Menekse et al. 2017; O’Donnell 2006).

The value of collaborative learning has been well documented for both cognitive and interpersonal competencies across domains including engineering education. However, positive achievement results were mainly found in structured learning environments in which additional feedback, support, and scaffolding are provided (Asterhan, Schwarz, and Cohen-Eliyahu 2014; Gijlers and de Jong 2013; Hmelo-Silver and Barrows 2008; Kirschner, Paas, and Kirschner 2009; King 1999; Rummel and Spada...
Studies in the context of computer-supported collaborative learning (CSCL) often illustrated the effectiveness of feedback, scaffolding, scripts, and awareness tools to enable better coordination and communication among group members in order to improve individual learning (e.g. Kimmerle and Cress 2008; Stahl, Koschmann, and Suthers 2014). For example, Kollar, Fischer, and Slotta’s (2007) study explored the role of interaction between structured external scripts in a CSCL environment and students’ internal scripts with respect to individual students’ science learning. The internal scripts in this study indicated the procedural knowledge on argumentation that students possess and often use in different situations. Results indicated that providing high structured external scripts supported the acquisition of domain-general knowledge for all students regardless of their internal scripts. Likewise, Rummel and Spada (2005) conducted an experimental study to explore the effects of the external support on successful collaborations. They found the dyads that observed an example of model collaboration, and the dyads that were provided with precise scripts for successful collaboration outperformed the dyads without scripts or support during collaboration.

While the studies have shown the significance of external support for successful collaboration, students in classroom settings are often asked to work in small groups or teams for various tasks without resources to receive systematic external support. And the results of student-led face-to-face collaborative learning are not consistent when collaborative learning occurs in relatively unstructured environments without scaffolding or feedback (Gadgil and Nokes-Malach 2012; Summers and Volet 2010; Yetter et al. 2006; Wentzel and Watkins 2011). For example, Yetter and colleagues (2006) found individual college students outperformed unstructured collaborative groups in solving mathematics problems. In addition, review studies have shown a small to moderate effect size for learning outcomes in collaborative learning tasks when compared with individuals studying alone (Lou, Abrami, and d’Apollonia 2001; Lou et al. 1996; Slavin 1995). For example, Lou and colleagues (2001), in their meta-analysis study, found a small effect size of 0.15 (Cohen’s d) comparing individual learning outcomes in small groups versus alone. Similarly, Slavin (1995) found an effect size of 0.21 (Cohen’s d), and Springer, Stanne, and Donovan (1999) found an effect size of 0.33 (Cohen’s d) based on standardised measures.

Prior engineering education studies have examined small groups or team-based learning of engineering students (e.g. Hsiung 2012; Huang, Shih, and Lai 2011; Terenzini et al. 2001). However, majority of these studies relied on students’ survey responses and/or other outcomes measures without analyzing the verbal data that reveals how groups actually collaborate (Fila and Loui 2014; Hsiung 2010; 2012; Huang, Shih, and Lai 2011; Mishra, Ostrovska, and Hacaloglu 2015; Pasha-Zaidi et al. 2015; Terenzini et al. 2001; Zafft, Adams, and Matkin 2009). There are a limited number of studies have examined verbal data in engineering student teams (e.g. Eggert et al. 2014; Purzer 2011; Tonso 2006). Some studies have shown that group interactions and discourse processes can facilitate learning (e.g. Purzer 2011). However, these verbal interactions may also prevent successful collaboration and lead to unproductive learning (Kuhn 2015). Consequently, it is important to investigate how and what types of verbal interactions and discourse patterns within small groups influence the individual achievement of engineering students.

In summary, studies have shown that collaborative learning is often successful in structured collaborative learning environments that contain external support in the form of feedback, scripts, or scaffolding. However, the learning results are often inconsistent or the effect sizes are small for studies that examined collaborative learning in unstructured learning environments in which there is no external support. This suggests that external support is a significant component of successful collaboration. However, external support is not often available for collaboration in classroom settings. Therefore, it is important to understand how students’ dialogue patterns are related to students’ learning outcomes when there is no external support. Understanding the kind of dialogue patterns that engender more learning will also enable researchers to design more appropriate scaffolding. In addition, although some engineering education researchers have studied discourse in small groups or teams, most did not explore the verbal (or process) data to understand the relations among the quality of group interactions and individual performance except Purzer’s (2011) study. Thus, more
studies are needed to systematically explore verbal data in small groups of engineering students. Accordingly, the primary goal of this study is to investigate the relationship between engineering students’ dialogue and their learning outcomes in an unstructured face-to-face collaborative learning environment by conducting a detailed analysis of student discussion at different grain sizes.

In this study, we compared students’ learning of materials science and engineering concepts across collaborative dyads and individual conditions by using an experimental design. We specifically investigated dialogue patterns and the quality of interaction between collaborative dyads in the context of Chi’s (2009) Interactive-Constructive-Active–Passive (ICAP) framework. The ICAP framework differentiates learning activities into four modes as interactive, constructive, active, or passive. The four modes of activities correspond to four sets of differentiable underlying cognitive processes. Based on the cognitive processes corresponding to each mode, the ICAP framework asserts that different modes of activities have differential learning effectiveness because they have different attributes and involve different cognitive processes. Chi (2009) hypothesises that, in terms of learning outcomes, interactive activities are likely to be superior to constructive activities, which in turn are almost always better than active activities, and all three ‘active’ modes are better than a ‘passive’ mode such as receiving instruction only (Chi and Wylie 2014; Menekse 2012; Menekse et al. 2013; Streveler and Menekse 2017).

In this study, we only included interactive and constructive conditions since these two modes are hypothesised to provide the highest level of learning compared to active and passive modes (Chi 2009; Menekse et al. 2013). Also, based on our review of published experimental studies, we found more studies that compared interactive conditions to active or passive conditions, but there were relatively few studies for the comparison of interactive and constructive conditions. Here, interactive condition was designed as student-led face-to-face collaborative dyads and constructive condition was designed as individual students with no collaboration with a peer, teacher, or system. Next, we will explain the interactive and constructive modes in detail.

1.1. Interactive mode

The interactive mode refers to instructional settings that allow a group of learners to collaboratively develop knowledge and understanding beyond the information contained in the given materials by building upon one another’s understanding. Behaviourally, collaborative groups should provide and receive feedback, ask each other questions, propose arguments and rebuttals, elaborate on each other’s ideas, and so forth. The processes underlying these overt interacting activities can include generating inferences to create new knowledge, revising one’s own mental model based on the partner’s inferences, knowledge, and point of view. Incorporating the partner’s contributions leads to the potential of creating new knowledge that neither partner could have generated by constructing alone.

Being in a dyad or small group, however, does not directly translate into effective learning (e.g. Chi and Menekse 2015; Kuhn 2015). In order to operationalise the interactive behaviours, Chi and Wylie (2014) proposed two criteria for dyad’s or small group’s learning: all group members’ content relevant contributions to discussion must be primarily constructive, and there must be sufficient turn taking between group members. Here, constructive utterances imply contributions not only to the relevant topic under discussion, but that the contributions must add some new information to the original presented information. For example, asking evaluative questions and receiving comprehensive responses; proposing an alternative idea, challenging a response are all examples of constructive contributions.

Besides constructive contributions, the distribution and frequency of turn taking among group members are also critical for successful interactions. Chi and Wylie (2014) hypothesised that frequent turn taking makes it easier for learners to incorporate their partners’ understanding and adjust their own mental model due to dynamic revisions of knowledge. Moreover, the frequency of turn taking decreases the probability that one partner will dominate the discussion and provides an equal
playing field for all. Often one partner may dominate the discussion and when this occurs, the dominant partners perform better than the non-dominant or ‘listening’ partners, as evidenced by the learning by teaching literature (e.g. Renkl 1995; Roscoe and Chi 2007). Therefore, ideal case of interactive learning environments requires a joint dialogue pattern in which all partners make substantive contributions to the topic in a co-constructive way.

1.2. Constructive mode

Constructive mode refers to instructional settings that allow individual students to generate additional outputs or products that contain new ideas, which go beyond the given information (Chi 2009). The cognitive processes hypothesised to underlie being constructive are those that can generate new ideas, insights, and conclusions in a way that allows learners not only to infer new knowledge, but also to repair or improve their existing knowledge.

Some of the constructive activities include self-explaining; drawing a concept map; generating questions; comparing and contrasting cases; integrating text and diagrams; drawing analogies; and generating predictions. For example, both drawing concept maps, and comparing and contrasting cases have been shown to be effective learning strategies in various studies (e.g. Biswas et al. 2005). Concept map generation requires learners to construct the relationships among ideas and concepts, thus the process itself helps learners to organise their mental models. Similarly, contrasting cases is another effective constructive learning strategy that helps learners to detect the differences and similarities between concepts, models, and/or ideas (Alfieri, Nokes-Malach, and Schunn 2013). Overall, these constructive activities facilitate learning by providing opportunities for learners to generate their own knowledge and understanding by incorporating different methods.

2. Present study and research questions

We designed an experimental study with two conditions, as collaborative dyads (interactive condition) and individual students (constructive condition), to investigate student learning across conditions, as well as to explore productive dialogue patterns within the interactive condition. The verbal analyses addressed the role of interactional factors in unstructured face-to-face collaboration and investigated the quality of interactions, the scaffolding durations, and the discourse moves between collaborative dyads while studying materials science and engineering concepts.

Specifically, there were four research questions in this study: (1) Do engineering students in interactive condition perform better on learning measures than the students in constructive condition? (2) How do the collaboration quality scores relate to students’ individual learning outcomes and dyads’ performance? (3) To what degree does the scaffolding duration relate to the dyads’ performance? (4) What kind of discourse moves have a significant effect on students’ knowledge construction while working collaboratively in an unstructured face-to-face learning environment?

3. Methods

3.1. Data

The data comes from a lab study (Menekse et al. 2013) that evaluated the ICAP framework by employing an experimentally designed study using concepts from materials science and engineering. There were all four conditions (i.e. four ICAP categories) in the 2013 study. Student learning data were originally reported in Menekse et al. (2013) paper. However, no verbal analysis or process data of collaboration (i.e. video records and transcriptions) were reported in Menekse et al. (2013) study. So, additional analyses for this study were carried out on the process data of the collaborative condition in order to investigate the effective and ineffective dialogue patterns.
The verbal data included approximately 12 hours of video and audio records resulting in 210 pages of transcriptions.

3.2. Participants and experimental conditions

The sample in two conditions (i.e. interactive and constructive) included 72 first- and second-year engineering students at a large state university in the U.S. Twenty-six participants were female and 46 were male. Forty-eight of the participants were randomly assigned to interactive condition and the remaining 24 were assigned to constructive condition. It was required for participations to have completed a college-level general chemistry class with a grade of ‘B’ or better in order to be familiar with the terminology used in the activities. Based on the United States grading system, the grade B indicates 3.00 on a 4.00 scale GPA (grade point average) and it is typically an equivalent of the ‘Lower Second’ grade in the United Kingdom grading system. Also, each participant received $15 of compensation for their participation. Participants were recruited via announcements emails and postings on bulletin boards. Once a student contacted us to participate, we scheduled with them based on their available times. For interactive condition, we tried to schedule two participants for the same time based on their availability. Due to scheduling challenges, the whole data collection took almost three months. The lab space that we used for data collection was a small room with chairs, desks, desktop computers, and a video camera located in one of the university buildings.

3.3. Materials

Intervention materials for both conditions consisted of an introductory text, a learning activity with graphs, figures and questions, and pre- and posttests. Since these materials were described in detail in Menekse et al. (2013) study, we will briefly explain them here.

3.3.1. Introductory text

An introductory text was created, consisting of definitions and short descriptions for materials science and engineering concepts, taken from two introductory materials science and engineering textbooks that are commonly used in universities and colleges (Callister 2006; Newell 2009). The introductory text was two pages long and contained definitions of terms such as bond energy, bond strength, and tensile properties. The main goal of the introductory text was to review the definitions of terminology used in the tests and the learning activity.

3.3.2. Learning activity

A learning activity with two graphs, two figures, and an activity sheet was created. The graphs and figures illustrated the properties of different metals in terms of elastic modulus, bond energy, thermal expansion, and melting points. (Please see Appendix 1 for the learning activity and activity sheet). We also constructed an activity sheet with five short-answer questions to guide students to interpret specific aspects of the information provided in the graphs and figures. For example, one question asked students to compare values for the different metals in each graph and to justify their findings for each comparison. Students were asked to write their responses on the worksheets. The accompanying activity sheet delivered an inquiry-oriented activity in which the data and relations embedded within the graphs and figures were followed by questions that directed students to generate analyses and conclusions.

3.3.3. Measures

A pre–posttest design was used to measure students’ prior knowledge and their learning outcomes after the intervention. The pretest consisted of 24 questions and the posttest consisted of the same 24 questions along with six additional questions. There were three types of question formats: multiple-choice, true–false, and open-ended. The multiple-choice questions had five responses that
included one correct answer. The true-false questions were two-tiered in that the first part required students to determine the correctness of a given statement, and the second part required students to explain their selection. The open-ended questions required short answers.

3.4. Procedure

All participants read the introductory text individually for 10–15 minutes. Students didn’t have access to this text after this time frame. Then all participants were given 25 minutes to complete the pretest individually. The learning activity was provided for each dyad in the interactive condition and they were instructed to study together. Dyads were told to reach a consensus for each question before writing their answers on the activity sheet, and their discussions were videotaped. No content-related feedback nor scaffolding was provided during the sessions. Students were given up to 30 minutes to complete the learning activity. Finally, all participants took the posttest individually and they were given 30 minutes to complete the posttest. Each session lasted approximately 90-100 minutes with data collection being completed in one session.

The same sequence of procedure was administered to the constructive condition. After reading the introductory text and completing the pretest, the same learning activity was provided for each individual participant. Similarly, no content-related feedback nor scaffolding was provided during sessions. They were given up to 30 minutes to complete the learning activity. All participants were given 30 minutes to complete the posttest individually.

4. Coding and scoring of verbal data

Guided by the technique described in (Chi 1997), three types of coding were carried out on the transcriptions of student dialogues: (1) Interactional quality, (2) scaffolding durations, and (3) discourse moves. Each category is described in detail in the following sections.

4.1. Coding scheme for the interactional quality scores

The entire corpus of transcribed protocols of students’ dialogue was used for this coding. Each dialogue exchange that addressed a question on the worksheet was considered a unit, regardless of how long the exchanges were. Since there were five questions on the activity worksheets, each dyad received a total of five interactional quality scores based on the dialogue patterns for each question. The important metric here was the proportion of substantive statements from each student, with a more equal distribution reflecting a better joint construction. The substantive statements refer to conversations in which students contribute to the dialogue by not only generating new ideas but also clarifying or completing their partners’ statements through expanding, elaborating, restating, or questioning. Table 1 shows the coding scheme for this analysis.

We coded the dialogue units as highly interactive (with a score of 3), medium, or low interactive (with scores of 2 and 1, respectively). A highly interactive score is given when substantive statements of each student build upon those of the other throughout the question segment, whereas a score of 1 is given when the dialogue is mostly one-sided or dominated by one student. The spectrum of the interactional quality ranged from dyads who largely construct their ideas and write them down independently with only minor statements of approval from the other, to collaborative discussions in which both students reach a shared understanding indicated by the proportion of substantive statements and responses of clarifying statements and restatements.

In addition, our coding scheme focused on capturing instances of a shared line of reasoning rather than two distinct lines. In addressing a specific problem, each student may assert a position, but one will then give in and go along with the other without going through the process in which a single line of reasoning is fleshed out over time and fully understood by both partners. As an indication of the shared or separate nature of reasoning, when writing down conclusions we noticed two varieties: the
one who developed the conclusion writes it out after the partner notes a verification. For example, one of the students explains and the partner says, ‘Okay, that sounds good,’ and the first student writes silently. Alternatively, with another dyad, both students engage in constructing the written explanation back and forth; one student verbalises while writing and the other student offers suggestions for wording and verifies/repeats aloud the ideas written. Please see Appendix 2 for examples illustrate how interactional quality scores were assigned. A similar version of this coding scheme was also used in Menekse et al. (2017) study to explore the team collaboration quality of robotics teams.

4.2. Coding scheme for the scaffolding durations

Dyads in this study were collaboratively responsible for answering the questions and writing on the provided activity sheet. Writing typically occurred at the end of the discussion for each question, and a writing episode refers to sections when one or both students decided to write the answers for the questions on the activity sheet. Writing episodes involve students’ discussion regarding what to write on the activity sheet. In some cases of writing episodes, when students were working collaboratively, they behaved very much like tutors and tutees without the explicit assignment of these roles by researchers. In other words, one of the two students would act like a tutor by asking questions and scaffolding the other student (i.e. tutee). So, in this coding scheme, we explored how both students behaved before and during the writing episodes. Dyads typically have one partner taking the role of the ‘writer’ and the other taking the role of the non-writer, even though these roles were not assigned by the activity sheet or the researchers. The ‘writer’ in this context refers to one of the dyads who wrote on the activity sheet and the ‘non-writer’ refers to the partner student. The students decided who would write on the activity sheet and their roles could change multiple times from question to question during their collaboration.

We specifically investigated the ‘scaffolding duration’ (in seconds) within a writing episode, which is the time intervals beginning with the ‘writer’ student being prompted and/or guided and/or verbally supported by the ‘non-writer’ student while writing on the activity sheet. A scaffolding duration occurs when the non-writer partner offered a significant addition to the writer’s thought process by challenging a partner’s claim, asking questions, completing the sentences, or adding new ideas to on what response to write on the activity sheet. On the other hand, the time intervals in which the non-writer student simply accepted the writer’s claim without discussing the ideas extensively were not counted towards these scaffolding durations.

The dialogue segment below shows an example of the scaffolding duration. In this example, the writer (S2) initially had difficulty figuring out the relationship between bond strength and bond energy while answering one of the questions on the activity sheet. Then the non-writer (S1) partner

<table>
<thead>
<tr>
<th>Coding scores</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Score 1</td>
<td>• There is little substantive discussion or only one student’s statements are substantive.</td>
</tr>
<tr>
<td></td>
<td>• Students do not clarify or complete their partners’ statements, instead voicing generic responses of agreement.</td>
</tr>
<tr>
<td></td>
<td>• One student decides what to write while the other agrees but contributes very little or nothing.</td>
</tr>
<tr>
<td>Score 2</td>
<td>• One student’s statements are mostly substantive and the other varies between substantive and shallow statements and responses.</td>
</tr>
<tr>
<td></td>
<td>• Statements and responses are discontinuous as each student makes assertions independent from those of the other.</td>
</tr>
<tr>
<td></td>
<td>• One student contributes most to what will be written while the other takes a smaller, though substantive, role.</td>
</tr>
<tr>
<td>Score 3</td>
<td>• Substantive statements and responses of each student build upon those of the other, indicating a shared line of reasoning.</td>
</tr>
<tr>
<td></td>
<td>• Students clarify or complete their partners’ statements through expanding, elaborating, restatement or rebuttal.</td>
</tr>
<tr>
<td></td>
<td>• Conclusions are co-constructed with both students involved fairly equally in determining what to write.</td>
</tr>
</tbody>
</table>
completed and elaborated the writer’s initial statement by explaining the relationship. After the non-writer’s (S1) contribution, the writer (S2) continued writing on the activity sheet. The beginning of the scaffolding duration started with the non-writer (S1) student’s contribution to thought process of writer (S2) and ended when the writer student figured out the relation between bond energy and bond strength. The time stamps show that this entire dialogue took 51 seconds.

S2: [Starts writing] The higher bond strength is …
S1: [Reads the writing] Higher the bond strength … (Time Stamp 15:41)
S2: I don’t know how to word this.
S1: Neither do I. Well, if you have a higher bond strength then there’s going to be more bond energy?
S2: Bond energy.
S1: Higher bond strength is going to mean the attraction is greater, right? But, when you look at it, when it comes to energy, there’s actually more energy, oh that’s potential energy.
S2: The magnitude …
S1: Right, the magnitude is greater.
S2: The attraction is greater.
S1: When you have a higher bonding energy there’s going to be, well higher bond strength, which means there’s more energy, which makes them strong in the first place. (Time Stamp 16:32)

The average time-length of scaffolding durations was 255 seconds (4.25 minutes) for the five questions per each dyad. In the analysis and result section, we present the relationship between the scaffolding duration and dyads’ learning scores.

4.3. Coding scheme for the discourse moves

The entire corpus of transcribed protocols was re-segmented at a finer grain size, at the ‘statement’ level. A ‘statement’ can be a single word, half sentence, full sentence, or multiple sentences depending on the context. Each statement was coded as claim, accept, oppose, elaborate, expand, question, response, or off-task. Table 2 provides descriptions and examples for each discourse move.

<table>
<thead>
<tr>
<th>Moves</th>
<th>Description</th>
<th>Example</th>
</tr>
</thead>
<tbody>
<tr>
<td>Claim</td>
<td>Proposing the initial idea; first response to questions on the activity sheet.</td>
<td>‘Metal C has the greatest coefficient’ ‘So, elastic modulus of metal A is greater than metal B’</td>
</tr>
<tr>
<td>Accept</td>
<td>(1) Expression of acceptance and/or agreement with peer’s claim; or (2) Repetition of the peer’s comment, claim, explanation without adding anything new.</td>
<td>‘I agree’ ‘Yeah, that sounds right’</td>
</tr>
<tr>
<td>Oppose</td>
<td>(1) Raises an alternative to peer’s claim; or (2) Challenges peer’s claim; or (3) Briefly rejecting or disagreeing with peer’s claim.</td>
<td>‘No, I think it is the difference between both’ ‘I do not think so’ ‘It might be this’</td>
</tr>
<tr>
<td>Elaborate</td>
<td>(1) Completing peer’s claim and/or explanation; or (2) Adding new ideas on a peer’s claim and/or explanation.</td>
<td>‘Like this, it expands a little bit that’s all I can tell. And then this one and this one seems equal’ ‘Yeah, so the max highest is iron and then the one is the second lower actually this one is max highest.’</td>
</tr>
<tr>
<td>Expand</td>
<td>(1) Reflecting on or clarifying own claim; or (2) expanding/elaborating own claim by adding explanations and/or new information.</td>
<td>‘The melting point plus a greatest stretch expand’ ‘We do not know the exact temperature but you can get a comparison’</td>
</tr>
<tr>
<td>Question</td>
<td>Asking for explanation, clarification or approval.</td>
<td>‘That is the one, right?’ ‘Does this make it more elastic?’ ‘Which one?’</td>
</tr>
<tr>
<td>Response</td>
<td>Providing any type of response(s) to peer’s yes/no type or wh type questions.</td>
<td>‘No, relation is between bond strength and elastic modulus’ ‘Yes’ ‘It depends’</td>
</tr>
<tr>
<td>Off-task</td>
<td>Comments that are not are not related to topic/content.</td>
<td>‘I am late’ ‘You said you had chicken scratch’</td>
</tr>
</tbody>
</table>
The motivation for coding discourse moves was to explore the students’ reasoning and their contribution to dialogue in a more detailed manner. Each discourse move was used to highlight the participatory roles of students and to dissect the complexities of the dyadic exchanges. Although the two prior coding categories (i.e. interactional quality and scaffolding durations) were also focused on the student contribution, the coding for discourse moves provides a more comprehensive understanding of the individual student’s role in shaping the discussion since each individual statement received a discourse move coding.

We iteratively developed a coding scheme to document students’ discourse moves. We initially started with the four broad categories of claim, accept, oppose, and discuss. While we started the coding with these four discourse moves, we revised them and added more categories as needed to identify more specific utterances. For example, after reading a couple of transcripts, the ‘discuss’ code was divided into four different codes as elaborate, expand, question, and response because we realised the discuss code alone was not sufficient to examine the students’ interaction at a finer grain size. The final protocol involved eight discourse moves as: Claim, accept, oppose, elaborate, expand, question, response, and off-task.

5. Analysis and results

5.1. Learning results across conditions

A one-way analysis of covariance (ANCOVA) was conducted to investigate students’ learning outcomes across two conditions. The dependent variable was students’ posttest scores and the covariate was the students’ pretest scores. A preliminary analysis for the homogeneity of slopes assumption indicated that the relationship between the dependent variable and the covariate did not differ significantly as a function of the independent variable, $F(1, 68) = 0.001, MSE = 169.65, p = .98$, partial $\eta^2 = 0.00$. The ANCOVA was significant, $F(1, 69) = 8.48, MSE = 167.19, p < .01$, partial $\eta^2 = 0.11$. To calculate the effect size as Cohen’s $d$, the following formula was used (Cohen 1988; Lakens 2013):

$$d = \frac{\bar{X}_i - \bar{X}_k}{\sqrt{MSE}} = 0.73$$

In sum, these results answer our first research question and suggest students in the interactive condition (dyads) performed significantly better than the ones in the constructive condition (individuals) with a moderate to strong effect size of 0.73, controlling for pretest results. Table 3 provides means and standard deviations for pretest and posttest by percentages for individual students in both conditions.

5.2. Relation between interactional quality scores and learning outcomes

Two raters coded 10 of the 24 transcribed protocols individually. The initial percent agreement was 82% for the interactional quality scores. The disagreements between raters were discussed and resolved. The rest of the transcripts were coded by one of the raters. Each dyad received an interactional quality score as an average score of five question segments. Overall, the average interactional quality score across 24 dyads was 1.83.

We investigated the relation between dyads’ interactional quality scores and their normalised learning gain scores (Hake 1998) per dyad by using Pearson product moment correlation coefficient.

### Table 3. Means and standard deviations of students’ scores for pretest and posttest by percentages.

<table>
<thead>
<tr>
<th></th>
<th>Pretest</th>
<th>Posttest</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$M$</td>
<td>SD</td>
</tr>
<tr>
<td>Interactive</td>
<td>52.28</td>
<td>15.16</td>
</tr>
<tr>
<td>Constructive</td>
<td>53.14</td>
<td>15.11</td>
</tr>
</tbody>
</table>
The normalised gain scores were calculated by using the following formula: Normalised Gain Scores = (posttest % – pretest %)/(100 – pretest %). The normalised gain scores for each dyad indicates the proportion of the maximum potential improvement obtained beyond the pretest. The correlation was significant for interactional quality scores and normalised gain scores, $r(22) = 0.47$, $p < .05$. Figure 1 shows the scatterplot for this relation.

5.2.1. Individual student performance within each dyad
In addition to dyads’ average scores, individual students’ performance within each dyad was also investigated. We wanted to compare and see which factor was more important: the interactional quality score or the achievement of one’s partner. We divided the 24 dyads (48 individuals) into four groups based on two factors: their interactional quality scores (divided into two groups as high or low based on a median split), and whether each partner within the dyad is a low or high performer, relative to each other, based on their normalised gain scores. This gives us four groups of participants with 12 individuals within each group: (1) low and (2) high performers compared to their partners within the 12 dyads with lower interactional quality scores, and (3) low and (4) high performers within the 12 dyads with the higher interactional quality scores.

One-way ANOVA was conducted to evaluate the relationship among the four groups: The main effect was significant, $F(3, 44) = 19.31$, $p < .001$, $\eta^2 = 0.57$. Figure 2 shows the normalised gain scores for individuals in these four groups.

Follow-up tests were conducted to evaluate the pairwise differences among the means of these four groups. Since the variances among the four groups were different, we chose not to assume the

Figure 1. Scatterplot showing the dyads’ interactional quality scores and their average normalised gain scores.
variances were homogenous and conducted post hoc comparisons by using the Dunnett’s C test. The most important finding is that there was no significant difference between the high performers within the dyads with lower interactional quality scores ($M = 0.63$) and low performers within the dyads with higher interactional quality scores ($M = 0.70$). An alternative way to look at this is that the learning gap between students within each dyad is greater for the dyads with lower interactional quality scores (a difference of 0.32) than dyads with higher interactional quality scores (a difference of 0.18). This suggests that low performers can achieve higher learning gains if they interacted well with their partners, given that they all had similar pretest scores.

5.2.2. Individual student performance across conditions
Lastly, we were interested in comparing how students within the dyads with lower interactional quality scores performed compared to students in the individual constructive condition. An ANOVA found that the normalised learning gains were not significant for the comparison, $F(1, 46) = 2.47$, $p = .12$, $\eta^2 = 0.05$, indicating that when dyads are not truly interactive, there is no difference between students’ performance across collaborative and individual conditions. This finding may explain the inconsistency in some collaborative learning studies that found no benefit of collaboration over individual learning conditions. While Menekse et al. (2013) showed that overall students in the interactive condition performed significantly better than the students in the constructive condition, when we looked at the two sub-groups of the interactive condition; dyads that have higher and lower interactional quality scores, the dyads with lower interactional quality scores are not performing better than the students in the constructive condition. Figure 3 shows the normalised gain scores for individual students in the solo condition, individual students within collaborative dyads with lower interactional quality scores, and individual students within collaborative dyads with higher interactional quality scores.

5.3. Relation between scaffolding duration and learning outcomes
We hypothesised that the total time of scaffolding duration should be correlated with dyads’ normalised gain scores since the scaffolding prompts are good indicators of joint interaction in dyadic
communication. Correlation analyses were conducted by using Pearson product moment correlation coefficient. The correlation between the scaffolding duration and the normalised gain scores was significant; \( r(22) = 0.51, p < .01 \). This result shows the prominence of the substantive involvement of both students on decision-making for what response to write on the activity sheet.

5.4. Relation between discourse moves and learning outcomes

The same two raters separately coded 10 of the 24 transcripts. The initial percent agreement was 85% for the discourse moves. The disagreements between raters were discussed and resolved. The rest of the transcripts were coded by one of the raters. Figure 4 shows the distribution of all the discourse moves in terms of percentages.

![Figure 3](image-url) Normalised gain scores for individual students in the solo (constructive) condition, individual students within dyads with lower interactional quality scores, and individual students within dyads with higher interactional quality scores.

![Figure 4](image-url) The percentages of all discourse moves.
We conducted a multiple regression analysis to evaluate how well the discourse moves predicted the normalised gain scores. Most of the discourse moves were observed in the accept (22.90%), elaborate (17.70%), and expand (22.20%) categories. On the other hand, the oppose (3.10%) and off-task (2.44%) type moves were rarely observed in students’ dialogue. Since the main goal was to investigate the effects of discourse moves on students’ learning outcomes and there were very few of oppose and off-task categories, we excluded these categories for this multiple regression analysis. Also, the correlation between question and response categories was high (0.97), so one of these predictors was redundant. Therefore, we kept the question moves but excluded the response moves from the regression analysis since the frequency of question moves was slightly more than the response moves.

The linear combination of discourse moves was significantly related to the normalised gain scores, $F(5, 18) = 5.51, p < .01$, normalised $R^2 = 0.49$. The sample multiple correlation coefficient was 0.78, indicating 61% of the variance of the normalised learning gains can be accounted for by the linear combination of discourse moves.

Among the discourse moves, accept and elaborate moves were significantly correlated with the normalised gain scores. Based on these results, we created two scatterplots (Figures 5 and 6) to illustrate the relation between normalised gain scores and the frequency of accept moves; and the relation between normalised gain scores and the frequency of elaborate moves. As expected, the first scatterplot indicated a negative correlation between dyads’ average learning outcomes and the frequency of accept moves in a dialogue. On the contrary, Figure 6 indicates a positive relation between dyads’ average learning outcomes and the frequency of elaborate moves in a dialogue.
In other words, these results confirmed that the dyads that completed or added new ideas to peer’s claims and explanations learned significantly because they were being constructive; whereas the dyads that simply accepted peer’s claims and explanations without adding anything new did not learn significantly, because they were being active, as predicted by the ICAP framework.

6. Summary and discussion

In this study, we primarily investigated the role interactional factors in an unstructured face-to-face collaborative learning environment by examining the dialogue patterns, discourse moves, quality and duration of interactions between collaborative partners while working on the cognitively challenging engineering activities. We also compared students learning gains across two conditions as collaborative dyads (interactive condition) and individual students (constructive condition). Overall, students in the interactive condition performed significantly better than the students in the constructive condition. In other words, this finding indicates that learner-learner interactions allow learners to construct knowledge beyond what one can accomplish in a learner-task interaction, even though the individual students engage with cognitively challenging engineering tasks. However, our analyses of the interactional quality indicated that this is particularly true for the highly interactive dyads.

In contrast, when there is little substantive discussion and only one of the students contributes very little or nothing, there was no meaningful gain for the non-contributing partner (Chi and Menekse 2015). However, the contributing partner was still benefiting which aligns with the

Figure 6. Scatterplot showing the correlation between the frequency of elaborate type moves in dyads’ dialogue and normalised gain scores.
findings of the ‘peer teaching’ literature as well as the prediction of ICAP (e.g. Roscoe and Chi 2007). Moreover, further analysis revealed that the students within the dyads with lower interactional quality scores did not perform significantly better than the students in the individual constructive condition. This supports the discrepant findings that students do not always benefit from collaborative work and that working in groups does not automatically make engagement interactive and beneficial (Dillenbourg 1999; Kuhn 2015; Summers and Volet 2010; Webb 1989).

We also explored students’ dialogues before and during the writing episodes and measured the time spent on scaffolding related to what response to write on the worksheet. The results showed that the scaffolding duration is significantly correlated with student learning. In some ways, the coding for scaffolding duration overlaps with what the interactional quality scores represent. However, the main difference between the two is that the scaffolding durations indicates a measure of the total time spent on the substantive discussion when one student scaffolds her partner regarding what to write on the worksheet, whereas the interactional quality scores indicate an ordinal classification of all discussion for the entire question.

Our analysis at a finer grain size, by classifying each utterance based on the discourse moves, revealed that frequency of the elaborate type moves in which students complete or add new ideas to peer’s claims and explanations positively correlated with the learning outcomes; whereas frequency of the accept type moves in which students simply accept peer’s claims and explanations without adding anything new negatively correlated with learning outcomes. In other words, as predicted by ICAP, the results show that coordinated co-constructive discourse with contribution by both students is crucial for both students to learn individually.

There are a few prior studies that have similar research design as ours, which compared small groups to individuals for the learning tasks that can be classified as interactive or constructive based on the ICAP framework (e.g. Azevedo et al. 2008; Kramarski and Dudai 2009; Okada and Simon 1997). For example, Okada and Simon (1997) compared the performances of dyads to individuals for a molecular genetics task and found, on average, dyads outperformed individuals when both partners of dyads contributed substantially in the discussions. Kramarski and Dudai’s (2009) compared the students who self-explained and received feedback from each other with the students who self-explained but did not receive any feedback. Based on the ICAP framework, the first condition can be classified as interactive, whereas the second condition can be classified as constructive. Results showed the students in the interactive condition performed significantly better that the ones in the constructive condition. Similarly, Czerniak and Haney (1998) compared the effect of collaborative concept mapping and individual concept mapping, and found students who collaborated on creating concept maps performed significantly better than the individual students. However, Czerniak and Haney (1998) study did not include any process data of collaboration, so it is not clear what interactional factors in the collaborative concept mapping condition lead to better performance. Finally, Azevedo et al. (2008) studied two conditions labelled as self-regulated learning (SRL) and externally facilitated self-regulated learning conditions (ERL). Students in both conditions were asked to think aloud while working on to learn about the circulatory system with hypermedia. The only difference between the two conditions was that the students in the ERL condition had access to a human tutor who used a tutoring script to provide adaptive scaffolding and feedback. Based on the ICAP framework, we can classify the SRL condition as constructive and the ERL condition as interactive. And the results showed students in the ERL condition outperformed the ones in SRL condition based on a biology test about the circulatory system. Taken as a whole, our findings were on par with the findings from these prior studies.

It is important to note some of the limitations of our study. The first limitation is comparing group and individual learning gains at the individual level by ignoring the group level variables. So, assuming independence and treating all observations as independent from each other is a methodological concern (Kenny, Kashy, and Cook 2006). So, one solution to overcome the independence assumption can be using the multilevel approach for the analysis. Multilevel analysis for the nested data structures like the ones in collaborative learning studies addresses the problems that researchers
encounter in the mixed factorial analysis of variance (Raudenbush and Bryk 2002). On the other hand, using a multilevel approach also has limitations in this type of research design that compares dyads with individuals. First, there will be no within group variability for the individual condition since each student in this condition should be categorised as a unique group for the multilevel analysis. Also, treating individuals as groups artificially increases the number of groups (i.e. sample size at level two) in multilevel analyses. One student per group obviously yields no variance within these groups and the lack of degrees of freedom to estimate $\sigma^2$ makes it impossible to run models without adding the groups with dyads into the model.

In addition, multilevel analysis requires large sample sizes and group sizes. For example, Scherbaum and Ferreter (2009) investigated the statistical power and required sample sizes for multilevel modelling, and they showed that at least 40 groups with 6 individuals in each group (which makes a total sample size of 240 students in the collaborative conditions) is needed to achieve a power of 0.80 at a medium effect size. However, these large sample sizes have been rarely used in collaborative learning studies and especially the numbers of individuals within groups (i.e. level one) often range between 2 (dyads) to 6 (small groups). Furthermore, the fine grain coding for conceptual ideas and structures from verbal data are extremely labour and time intensive. Verbal coding and qualitative analyses for studies with large sample sizes also require resources such as hiring several well-trained coders. As a result, conducting multilevel analyses is challenging for experimental collaborative learning studies that involve verbal coding and qualitative analysis. Simulation studies are needed to explore how badly the violation of non-independence assumption affects standard errors in a research design that compares dyads and individuals.

Another limitation was this was a lab study and the results could have been different in an actual classroom study due to multiple factors. First, in classrooms, it is difficult to control for possible confounding factors such as the level of student interaction, time spent on activities, or feedback from instructors. Second, in a regular course, grade is a major factor for students’ performance, whereas this is not the case in most lab studies. Third, there is typically an instructor or teaching team in a classroom environment, but in our study, students were directly engaged with the activities without any interference/support from an instructor or research team. Fourth, it is not easy to isolate dyads in real classroom settings and collect video or audio data. Fifth, in a classroom environment, most students typically know each other and have prior interactions, whereas in this study there is a very little chance for students in dyads to know each other. Since we scheduled pairs in the interactive condition whenever two participants were available at the same time, it is most likely that dyads were ‘strangers’ to each other. Our intervention asked them to communicate face-to-face while collaboratively working on the activities. So, working with a stranger may have an influence on participants’ behaviours. In addition, other factors such as participants’ attitudes, beliefs, and prior experiences about collaboration could have influence on how they work with others. However, we don’t have data in this study to understand the role of these factors on students’ interactions.

Overall, the results of the present analyses have important practical implications for educators and researchers from both engineering and other STEM disciplines. Our findings suggest that peer interactions in which peers contribute in constructive or cognitively demanding ways, deliver the best learning outcomes. Thus, instructors or curriculum designers should select activities that allow students to jointly generate new ideas that go beyond the information given in order to observe higher level reasoning. Learning activities should not only function as a rich tool for studying the content domain but they should also facilitate the development of collaborative practices when needed. Also, providing training to students that emphasises the value of certain types of discourse moves (i.e. elaboration versus accept) in collaborative learning settings could be effective in terms of learning outcomes.

In conclusion, our fundamental goal in this study was to explore the role of interactional factors in terms of quality of interactions, scaffolding instances, and discourse moves in an unstructured face-to-face collaborative learning environment with cognitively challenging activities in the context of
the ICAP framework. While the meta-review studies reveal the collaboration is effective when there is scaffolding, training or feedback component involved, the results for unstructured collaboration with no support systems are inconsistent. So, we focused on investigating what kinds of dialogue patterns occur in an unstructured collaborative condition and how these dialogue patterns affect students learning outcomes when there is no external support involved. The finer-grained analysis of verbal data revealed how certain types of dialogue patterns affect learning outcomes, as well as the significance of frequency of the scaffolding instances and certain discourse moves to predict students’ performance. In other words, the verbal data analyses specified when and how collaboration provides advantage over individual learning. In addition, interactional quality scores, scaffolding instances, and discourse moves explained the variation between the collaborative dyads’ performances within the interactive condition. Overall, our findings contribute to the current knowledge base on productive engagement in collaborative learning environments.

Disclosure statement
No potential conflict of interest was reported by the authors.

Funding
This material is based upon work supported by the USA National Science Foundation under Grant No. 0935235.

Notes on contributors
Muhsin Menekse is an Assistant Professor at Purdue University with a joint appointment in the School of Engineering Education and the Department of Curriculum & Instruction. Dr. Menekse’s primary research investigates how classroom activities affect conceptual understanding in engineering and science for all students. His second research focus is on verbal interactions that can enhance productive discussions in collaborative learning settings. And his third research focus is on metacognition and its implications for learning. Much of this research focuses on learning processes in classroom settings. Dr. Menekse is the recipient of the 2014 William Elgin Wickenden Award by the American Society for Engineering Education.

Michelene T. H. Chi is Foundation Professor and the Dorothy Bray Endowed Professor of Science and Teaching in the Mary Lou Fulton Teachers College at Arizona State University. A cognitive and learning science researcher, Dr. Chi’s overall approach to understanding how students learn is student-centered. She has developed the ICAP theory, focusing on what active learning activities students can do to engage that can lead to improved learning. She also has developed a theory of collective causality to explain why students have robust misconceptions and how they might be remove them. Chi has published over 125 papers and her work has been cited over 50,000 times. Dr. Chi was elected into the National Academy of Education in 2010, and the American Academy of Arts and Sciences in 2016. She received the Distinguished Contributions to Research in Education Award from the American Educational Research Association on 2016, and the David E. Rumelhart Prize from the Cognitive Science Society in 2018.

ORCID
Muhsin Menekse http://orcid.org/0000-0002-5547-5455

References


Appendices

Appendix 1. Learning activity and activity sheet

Characteristics of three pure metals, metal A, metal B and metal C, are pictured on four figures. Please answer the following questions based on the given information from these figures.

1) What characteristics/properties is shown in each figure? What information does each figure provide?
   Figure 1:

   Figure 2:

   Figure 3:

   Figure 4:

2) Compare the values for metal A, metal B, and metal C from each figure by using equal to (=), greater than (>), or less than (<) signs. Also, explain your reasoning for each comparison.
   - $E_A \quad \underline{\quad} \quad E_B \quad \underline{\quad} \quad E_C$
     Explanation:

   - Bond Energy$_A \quad \underline{\quad} \quad$ Bond Energy$_B \quad \underline{\quad} \quad$ Bond Energy$_C$
     Explanation:

   - $\alpha_A \quad \underline{\quad} \quad \alpha_B \quad \underline{\quad} \quad \alpha_C$
     Explanation:

   - $T_{M,A} \quad \underline{\quad} \quad T_{M,B} \quad \underline{\quad} \quad T_{M,C}$
     Explanation:
3) What kind of relations do exist between these characteristics/properties based on the figures? Come up with a rule that explains all the relations between three metals in these figures.

4) How do modulus, bonding energy, coefficient of thermal expansion and melting point affect bond strength? Explain your reasoning?

<table>
<thead>
<tr>
<th>Material</th>
<th>Coefficient of Thermal Expansion ($10^{-6} , (^{°}C)^{-1}$)</th>
<th>Melting Point Temperature (K)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Iron (Fe)</td>
<td>11.1</td>
<td>1811</td>
</tr>
<tr>
<td>Aluminum (Al)</td>
<td>23</td>
<td>933.47</td>
</tr>
<tr>
<td>Copper (Cu)</td>
<td>17</td>
<td>1357.77</td>
</tr>
<tr>
<td>Lead (Pb)</td>
<td>29</td>
<td>600.61</td>
</tr>
<tr>
<td>Tungsten (W)</td>
<td>4.5</td>
<td>3695</td>
</tr>
</tbody>
</table>

5) Table above contains linear coefficient of thermal expansions and melting point temperatures for iron, aluminum, copper, lead, and tungsten at room temperature. Based on the given information, select two possible sets of metals that can fit all the figures in place of metal A, metal B, metal C.

   Set 1: Metal A __________, Metal B __________, Metal C __________

   Explain your reasoning:

   Set 2: Metal A __________, Metal B __________, Metal C __________

   Explain your reasoning:
Appendix 2

The following examples illustrate how interactional quality scores were assigned. Each number represents a new turn by one of the partners. The first example is an episode with a Score of 1, second example is an episode with a score of 2, and third example is an episode with a score of 3.

(1) Exemplary episode with interactional quality score of 1:
   (1) Student A1: Have you seen this first graph before?
   (2) Student A2: No.
   (3) Student A1: I learned this in my material class before; strain and stress. It shows the relationship between these two.
   (4) Student A2: So, E is the energy?
   (5) Student A1: E is the elastic modulus.
   (6) Student A2: Oh elastic modulus.
   (7) Student A1: It is elastic modulus and it was elastic modulus in the pre-test.
   (8) Student A2: Yeah.
   (9) Student A1: But that number is just the relationship. It shows the relationship between these two. And so it is the slope. A has the higher elastic modulus because it has a greater slope.
   (10) Student A2: Yeah, relationship.
   (11) Student A1: Does that make sense?
   (12) Student A2: Yeah.
   (13) Student A1: So, it is a really easy graph if you know what it is looking for.
   (14) Student A2: Yeah. It was in the pre-test. I do not know what …
   (15) Student A1: Yeah, you do not know but it is really easy.
   (16) Student A2: OK …
   (17) Student A1: So the characteristics in figure 1. It shows the relationship between stress and strain.
   (18) Student A2: What information does each figure provide? Figure 1 [Reading the question 1 from worksheet].
   (19) Student A1: It is this one.
   (20) Student A2: Figure 1 is this one. The relationship between strain and stress, which is elastic modulus.
   (21) Student A1: Modulus, yeah. What characteristics …
   (22) Student A2: The relationship between strain and stress.
   (23) Student A1: yeah. [Starts writing] What information does each figure provide? This provides that A has the highest elastic modulus. Elastic modulus of A is greater than B which is greater than C.
   (24) Student A2: Okay …

The episode above is representative of a unit scored with the lowest interaction score. In this episode, student A1 is initiating all the ideas and student A2 is simply accepting the initiated ideas without discussing and expanding. Also, student A2 is asking only one question which is a yes/no type question that does not add anything new to discussion.

(2) Exemplary episode with an interactional quality score of 2:
   (1) Student B2: [Reading question 4] How do modulus, bond energy, coefficient of thermal expansion again modulus … I do not know. Ohh … uhhmmm … A greater modulus probably means the greater bond strength, right?
   (2) Student B1: Yeah
   (3) Student B2: Okay so, bonding energy lower that’s so except for coefficient of thermal expansion. The greater modulus, greater bonding energy and a greater melting point all relate to higher bond strength.
   (4) Student B1: Okay so, a greater modulus has greater bond energy and …
   (5) Student B2: uhhmmm will result in a higher melting point.
   (6) Student B1: Yeah … well yeah … will result in a higher melting point
   (7) Student B2: And this all relates to a higher bond strength, greater bond strength
   (8) Student B1: What? Okay … All characteristics …
   (9) Student B2: All relates to …
   (10) Student B1: Relates to …
   (11) Student B2: Higher bond energy bond strength
   (12) Student B1: All relates to higher bond energy, uhhmmm …
   (13) Student B2: But the coefficient of thermal expansion …
   (14) Student B1: It has an inverse relationship so that’s negative
   (15) Student B2: Yeah.
   (16) Student B1: As bond energy increases it decreases.
   (17) Student B2: Explain your reasoning.
   (18) Student B1: Yeah.
   (19) Student B2: Yeah.
(20) Student B1: Coefficient of thermal expansion decreases as the modulus …
(21) Student B2: As the bond energy increases.
(22) Student B1: As the bond … yeah …
(23) Student B2: Related to bond strength.
(24) Student B1: Yeah, as the bond strength increases.
(25) Student B2: So, as the bond strength increases. So, Here it can be reasoning it will probably mean that greater bond strength means more amount of temperature is required to break it
(26) Student B1: Okay …
(27) Student B2: And uhh … I do not know how to relate it modulus again I was not sure. It just means the thing is more elastic, I think.
(28) Student B1: Which one? This one?
(29) Student B2: No, relation between bond strength and elastic modulus.
(30) Student B1: Oh yeah.
(31) Student B2: Is that like that makes it more elastic? Because that would make sense if greater bond strength means
(32) Student B1: Haha …
(33) Student B2: Metal is more elastic
(34) Student B1: Okay.
(35) Student B2: Okay. The reasoning will be greater bond strength means that metal will be more elastic.
(36) Student B1: Yeah. That’s good.
(37) Student B2: And more temperature is required to break the bonds so a higher melting point.
(38) Student B1: Cool.
(39) Student B2: Did you write for this one? [Showing figure 3]
(40) Student B1: Ohh … I did
(41) Student B2: No like. Did you give any explanation for that?
(42) Student B1: Okay, okay uhhmm … so, it expands because the bond
(43) Student B2: Weaker the bond, more expansion
(44) Student B1: Weaker the bond more expansion yeah. The weaker the bonds have higher expansion. Thermal expansion.

The second example above is a representative episode which was scored with the score 2. In this episode, similar to the first example, one of the students, student B2, is initially proposing most of the ideas. Even though student B1 is not expanding or opposing most of student B2’s statements, she/he is not simply accepting the proposed ideas with comments like ‘yes’ and/or ‘I agree’, but restating and repeating B2’s proposed ideas. So, B1 is acting actively in terms of the ICAP framework rather than passively voicing agreement with B2. Also, student B1 is adding critical information at comment lines #14, #16, #24, and #42.

(3) Exemplary episode with an interactional quality score of 3:
(1) Student C2: [Reading Question 4] how do modulus, bonding energy, coefficient of thermal expansion and melting point affect bond strength? Explain your reasoning?
(2) Student C2: It is just intuitively, metal A is the strongest because it does not deform as much when you apply the same strain to it and it takes a lot more ripped part of a bond, I guess.
(3) Student C1: And its melting point, more energy is required to melt.
(4) Student C2: Make it destabilize, yeah.
(5) Student C1: So,
(6) Student C2: And when you heat it, it does not change its shape as easily as metal C.
(7) Student C1: So, how do we handle bond … metal A would be strongest per se. All four contributing the bond strength …
(8) Student C2: How about elastic modulus, bond energy and melting point all increase bond strength while high coefficient of thermal expansion decreases bond strength?
(9) Student C1: How this decreases bond strength? [Showing figure 3]
(10) Student C2: I am not sure it decreases it directly; I just notice it is the opposite of these three.
(11) Student C1: So, I guess thermal expansion does not contribute to the other three.
(12) Student C2: Possibly, I am remembering that the thing we read mentions that thermal expansion means the molecules are getting further apart, like I guess that would also means it is easier to tear down apart because there is like metallic attraction
(13) Student C1: Ok, so, these three would help, but this not …
(14) Student C2: You want these to be high and this to be low to maximize bond strength
(15) Student C1: Yes.
(16) Student C2: Alright [Writing on the worksheet]. And then for explaining that …
(17) Student C1: Less stress, more is energy is required, uhhm, more … more energy is required for this …
(18) Student C2: Take more energy for any change happens, whereas this means less energy is needed for change.
In the third example above, both students propose fairly equal amounts of substantive statements and responses and each student builds upon those of the other. For example, as student C2 initiates the statement ‘Metal A is the strongest because it does not deform as much when you apply the same strain to it and it takes a lot more ripped part of a bond, I guess’, student C1 is expanding this statement by adding ‘And its melting point, more energy is required to melt.’ Also, both students are asking information seeking questions by referring to each other’s comments like ‘How this decreases bond strength?’ Note that while the interactional quality score for each segment indicates the degree of shared line of reasoning between dyads, these scores are not exclusively dependent on the number of turns. In other words, although the number of turns between dyads is correlated with the interactional quality score, these scores were not assigned by counting the frequency of turns.