11-20-2022

Investigating Small-Group Cognitive Engagement in General Chemistry Learning Activities using Qualitative Content Analysis and the ICAP Framework

Safaa Y. El-Mansy
Portland State University

Jack Barbera
Portland State University, jbarbera@pdx.edu

Alissa J. Hartig
Portland State University, ahartig@pdx.edu

Citation Details

Check for updates

Investigating small-group cognitive engagement in general chemistry learning activities using qualitative content analysis and the ICAP framework

Safaa Y. El-Mansy, [ID][a] Jack Barbera [ID][a] and Alissa J. Hartig [ID]*[b]

The level of students' engagement during active learning activities conducted in small groups is important to understanding the effectiveness of these activities. The Interactive–Constructive–Active–Passive (ICAP) framework is a way to determine the cognitive engagement of these groups by analyzing the conversations that occur while student groups work on an activity. This study used qualitative content analysis and ICAP to investigate cognitive engagement during group activities in a General Chemistry course at the question level, a finer grain size than previously studied. The analysis determined the expected engagement based on question design and the observed engagement based on group conversations. Comparisons of expected and observed engagement showed cases of mismatch, and further analysis determined that incorrect model use, unfamiliar scientific vocabulary, and difficulty moving between molecular representations were all contributing themes to the observed mismatches. The implications of these findings with regard to teaching and research are discussed.

Introduction

Active learning (AL) strategies have been shown to enhance student success beyond traditional methods (Kuh *et al.*, 2005; National Research Council, 2012; Freeman *et al.*, 2014), often improving outcomes for students who have been historically underrepresented within science, technology, engineering, and mathematics (STEM) fields (Lorenzo *et al.*, 2006; Haak *et al.*, 2011; Eddy and Hogan, 2014). For these reasons, AL strategies have been at the center of national calls for the adoption of evidence-based instructional practices to transform education in STEM fields (National Research Council, 2012; President's Council of Advisors on Science and Technology (PCAST), 2012).

At the same time, evidence supporting the effectiveness of a given strategy can be inconsistent (*e.g.*, Andrews *et al.*, 2011) and simply adding AL strategies to a learning environment does not necessarily lead to the same performance outcomes across groups (Shortlidge *et al.*, 2019). Likewise, a 2019 meta-analysis of peer-reviewed studies on the effectiveness of a wide range of AL strategies within chemistry found that the effect size of these practices varied widely, in some cases resulting in no positive impact (Rahman and Lewis, 2020). As Cooper (2016) points out, the umbrella of AL also covers a wide range of

classroom practices, making it difficult to define what specific aspects of AL are effective and under what conditions such strategies work.

At a minimum, the effectiveness of any AL strategy depends on learners' meaningful cognitive engagement with the learning materials (Bonwell and Eison, 1991). While there is little dispute that learners benefit more from active compared to passive learning (Freeman *et al.*, 2014), a broader hierarchy of cognitive engagement has been proposed (Chi, 2009; Chi and Wylie, 2014). The ICAP framework (Chi, 2009; Chi and Wylie, 2014) offers a way to understand the varied outcomes in AL through a hierarchy of four levels of cognitive engagement: Interactive, Constructive, Active, and Passive. In this framework, simply being Active is one of the lower levels of engagement and is less likely to foster students' understanding than the higher level Constructive or Interactive modes (Chi and Wylie, 2014). In the ICAP framework, students' level of cognitive engagement is evaluated based on their overt physical and verbal behaviors (Fig. 1). For example, behaviors related to receiving information, such as reading a text or listening to instructions, would indicate Passive engagement. Active engagement would involve physical manipulations of information while learning, such as highlighting or underlining text. During Constructive engagement, students would perform the same physical manipulations that occur in Active engagement; in addition, they would generate output beyond the information provided in the learning materials. Examples

[a] Department of Chemistry, Portland State University, Portland, Oregon, USA
[b] Department of Applied Linguistics, Portland State University, Portland, Oregon, USA. E-mail: ahartig@pdx.edu
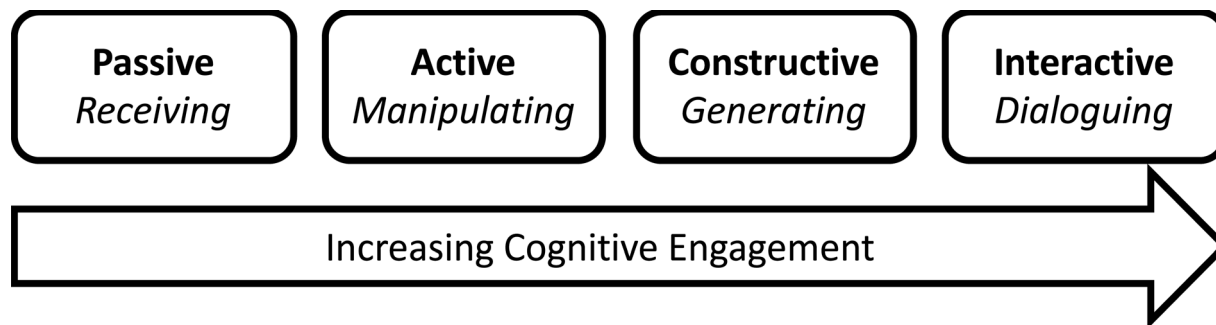
Fig. 1 Modes of cognitive engagement (in bold) and characteristic behavior (in italics) according to the ICAP framework (Chi *et al.*, 2018).

of Constructive engagement include summarizing a text or taking notes in one's own words. Similar to the Constructive mode, during Interactive engagement, students would generate new information; however, this generation would occur through dialoguing among students or between students and instructors.

Studies within the ICAP framework have operationalized cognitive engagement by observing students' physical behaviors (Villalta-Cerdas and Sandi-Urena, 2014; Wiggins *et al.*, 2017), categorizing activities by their broad instructional design features (Wiggins *et al.*, 2017; Henderson, 2019; Lim *et al.*, 2019; Menekse and Chi, 2019), and analyzing student conversations (Chi, 2009; Menekse and Chi, 2019; Liyanage *et al.*, 2021). Each of these approaches has strengths and weaknesses. ICAP studies that examine engagement in terms of students' physical behaviors have used large-scale observation of overt behaviors at regular intervals at a distance in order to capture whole-class data (*i.e.*, an observer seated in the back of the room with a chart, such as the "live coding" used by Wiggins *et al.* (2017) or the observation procedures used in Villalta-Cerdas and Sandi-Urena (2014)). While this approach may be able to distinguish Passive engagement from higher ICAP levels, the differences between Active, Constructive, and Interactive engagement are difficult to tease out at this level of granularity. For example, if a student is writing something on a worksheet, this could be simply Active engagement if it involves identifying relevant information on a graph and recording the answer. However, if the student is making inferences based on trends observed in the same graph, this student would be engaging at a Constructive level. What students are saying while engaging in these physical behaviors is essential to determining what level of engagement they reflect.

ICAP studies that rely on the instructional design features of the activity as a whole are based on the idea that the structure of the activity itself will constrain the ways that students can engage with it. For example, to assess the impact of cognitive engagement on learning, Henderson (2019) used a series of instructional conditions designed to reflect various ICAP levels, in which a lecture-based condition was used for Passive engagement, an individual writing activity was used to elicit Constructive engagement, and a peer instruction format was used to prompt Interactive engagement. This focus on coding engagement based on instructional design features assumes, however, that all students in a group will engage at the same level throughout an activity and does not distinguish among the levels of cognitive engagement required for different types of questions or phases within an activity.

These assumptions merit greater scrutiny. Research has shown that the type of activity students participate in can affect the nature of their conversation when working in small groups (Young and Talanquer, 2013). These differences in group conversations may reflect different modes of engagement. A study on small group activities using Peer-Led Guided Inquiry (PLGI) found that students' construction of arguments varied based on the number of students participating (Kulatunga *et al.*, 2013). It is possible that these students were engaging at different levels. Variations in conversation may also be important in Process-Oriented Guided Inquiry Learning (POGIL) activities (Farrell *et al.*, 1999; Hanson *et al.*, 2018). Through the lens of the ICAP framework, not every part of an activity may elicit the same level of cognitive engagement. For example, POGIL activities involve a three-step learning cycle (Atkin and Karplus, 1962) where students first explore information provided in a model, then identify trends and patterns during the concept invention step, and finally apply the learned concept to new situations (Hanson *et al.*, 2018). The direct questions about a model during the exploration stage of this cycle are meant to ensure that students understand the model on which later parts of the activity are based. In terms of the ICAP framework, many of these questions rely primarily on Active engagement because they ask students to identify and/or reflect on information in a model that is provided for them and do not require the generation of additional information. By contrast, questions from the concept invention and application stages are more likely to elicit Constructive or Interactive engagement because they require students to make inferences that go beyond the information provided in the original model. This type of variation might be expected in any type of scaffolded learning activity.

ICAP studies that examine student conversations have generally used discourse analysis as a means for understanding student engagement during AL activities. Discourse analysis examines texts and talk in context in order to understand participants' actions (Wood and Kroger, 2000), and in education research, discourse analysis focuses on the role of spoken language in teaching and learning (Cole *et al.*, 2014). Discourse analysis research in chemistry education research has largely focused on patterns of interaction or argumentation in various

instructional settings (Kulatunga and Lewis, 2013; Xu and Talanquer, 2013; Young and Talanquer, 2013; Warfa *et al.*, 2014; Current and Kowalske, 2016; Moon *et al.*, 2016; Repice *et al.*, 2016; Shultz and Li, 2016; Stanford *et al.*, 2016; Dohrn and Dohn, 2018). The use of discourse analysis in ICAP studies both within and outside of chemistry education research has generally been oriented toward the coding of individual student conversational turns, for example, the frequency of specific discourse moves (*e.g.*, claim, accept, oppose) (Menekse and Chi, 2019), or the frequency, distribution, and engagement level evident in student conversational turns during small-group discussions (Liyanage *et al.*, 2021).

Discourse analysis can also be applied at a broader level, beyond individual turns. Because the highest two engagement levels outlined in the ICAP theory rely on distinctions that relate not just to what individual students are doing but to how they respond to one another during small-group conversations, coding longer exchanges is especially useful for distinguishing between Constructive and Interactive engagement. As noted above, there is a need to examine the extent to which actual student engagement in an activity matches the planned level of engagement based on the instructional design features of the activity itself. Therefore, using these ICAP levels as coding categories for both the activity design features and for students' observed engagement as evident in their conversations across different parts of an activity can provide a systematic way of investigating this alignment.

Whereas discourse analysis is useful in understanding *how* students interact with one another, an alternative method is needed to investigate *what* is being said, *i.e.*, the content of the conversation. Qualitative content analysis (QCA) is well suited to filling this gap. QCA offers a method for systematically coding the content of textual data, whether verbal or written, to identify patterns (Schreier, 2012). QCA includes both deductive approaches (directed content analysis) and inductive approaches (conventional content analysis) (Hsieh and Shannon, 2005). Conventional content analysis can provide insights into phenomena that are not yet well described (Hsieh and Shannon, 2005). Because little research to date has explored the alignment between the instructional design features of individual parts of an activity and the actual level of engagement that they generate, an inductive approach is better suited to developing an understanding of instances where mismatches occur. Where mismatches between the planned and actual levels of engagement are found, conventional content analysis can be used to examine the content of students' discussions during these parts of an activity in order to identify patterns or themes that explain these mismatches. Therefore, conventional content analysis can be used to identify patterns as to which specific aspects of question design seem to foster higher or lower engagement across different groups as well as any other relevant themes that arise in students' conversations.

Research questions

The purpose of this study is to investigate cognitive engagement during small-group activities at the question level. To do so, we used qualitative content analysis and the ICAP framework to answer the following research questions.

(1) What range of engagement modes are expected during a general chemistry AL activity based on the question design?

(2) What range of engagement modes are observed during a general chemistry AL activity based on students' physical and verbal behaviors during group conversations?

(3) If mismatches occur between the expected and observed levels of cognitive engagement, what themes account for this mismatch?

Methods

Setting

Students from the first and second terms of a three-term General Chemistry sequence at Portland State University in the Pacific Northwest of the United States participated in this study. This course consisted of 20–30 students who were enrolled in the Honors College. Students in these courses come from a variety of STEM majors, including biology, chemistry, physics, and the pre-professional tracks, such as pre-medical and pre-dental. The first term occurred during fall quarter 2020, the second term occurred during winter quarter 2021, and the fall and winter term courses were taught by two different instructors. Classes met three times per week for 65 minutes and were conducted remotely through Zoom due to the COVID-19 pandemic. Each activity day began with a short lecture introducing the new material. Students were then placed in groups of 3–4 students in breakout rooms to work collaboratively on an activity worksheet. These groups remained consistent over the course of the term.

Activity worksheets were developed in house and structured using a format which included a model containing conceptual material followed by key questions, exercises, and problems. Key questions (KQ) generally asked about information explicitly presented in the model, providing an opportunity for students to gain familiarity with the content. Exercises (EX) included questions which required students to apply the content and infer an answer either conceptually or by performing a calculation. Problems (P) were similar to exercises but tended to be more complex, generally involving multiple steps or novel applications of the model content. The completed activity worksheets were turned in through the learning management system, and a nominal number of points were awarded for participation and attendance during the activity.

Data collection

Institutional Review Board (IRB) approval for this research study was received from Portland State University (HRRP# 2007004-18). Students were recruited at the beginning of each term by author S. Y. E. During the fall term, seven students consented to participate and were divided into two groups: Group A consisted of four students and Group B consisted of three students (Table 1). Three students from the fall also consented to participate during winter term and formed a

Table 1 Groupings for study

Fall 2020		Winter 2021
Group A	Group B	Group C
Nani	Jacob	Nani
Beth	Helen	Helen
Katie	Grace	Grace
Leslie	—	—

new group: Group C. All student names reported in this manuscript are pseudonyms.

Three activities were observed during fall term and one activity was observed during winter term. The activities during the fall were evenly spaced, with the first one covering the concepts of mole and molar mass occurring near the beginning of the term, the second one covering concepts involving solutions and dilutions occurring near the midway point of the term, and the third activity covering electronegativity and polarity occurring near the end of the term. During winter term, the single activity occurred near the beginning of the term and covered concepts surrounding thermal energy and calorimetry. Each breakout room session was audio and video recorded. These recordings were transcribed verbatim by a transcription service. Transcripts were then reviewed and edited as needed by author S. Y. E. and pertinent physical actions from the participants (*e.g.*, nod of agreement) were added to the transcripts. Unclear conversation was denoted by [XXX] in the transcripts.

Data analysis

Most of the prior work done using ICAP to investigate engagement during group activities assumed a single engagement mode over an entire activity (Menekse *et al.*, 2013; Wiggins *et al.*, 2017; Henderson, 2019). As these activities may contain different types of questions, this assumption may not be correct. Therefore, for the four activities observed, a finer grain size was used. The unit of analysis was each question within an activity. At this level of analysis, each question was first coded according to the ICAP framework, where the intended engagement mode of students was identified based on the question design.

Previous work investigating group conversations using ICAP looked at quantitative measures such as frequency of conversational turns or discourse moves (Wiggins *et al.*, 2017; Menekse and Chi, 2019); however, this type of analysis does not provide insight into the relation between the group conversation and

the question design. To address this gap, a second round of coding applied the ICAP framework to the group responses to each question in an activity. Each group's response to a question was coded based on the content of the conversation and the definition of each of the ICAP modes. The codebook for both types of coding is presented in Table 2. Each question and group response in the transcripts was coded deductively based on features of the levels of engagement outlined in the ICAP framework (Chi, 2009; Chi and Wylie, 2014).

Question coding. Three of the four engagement modes of ICAP (Fig. 1) were applied to each question in an activity (Table 2). For multi-part questions, each part was assigned a separate code. Passive engagement was not used to code questions because the questions were designed to be used in a group activity with the *intent* for students to engage actively at a minimum. Questions were coded as Active (A) if the information to answer the question could be found in the presented materials; it was assumed that students would use this information in their response. For the higher engagement modes (*i.e.*, Constructive and Interactive), the difference between these modes is determined by whether the generation of new information occurs through dialogue. Since it is not possible to distinguish this difference based on the structure of the questions alone, Constructive and Interactive engagement were collapsed into a single code, Constructive/Interactive (C/I).

Group response coding. Each group response to a question (or part of a question, for multi-part problems) was coded separately, resulting in a response code for each question answered in each activity. Passive engagement was not used as a code because by virtue of conversation simply occurring, students were manipulating information, and therefore, the lowest mode of engagement students could participate in at the whole-group level would be Active. Although it is possible for *individual* students to be engaging passively, the group response code was based on the conversation that occurred among all group members. The response was coded as Active (A) if the students in a group explicitly referred to the information presented in the activity in their response. The Constructive (C) code was defined by the conversation generating new information to respond to the question; this new information was generated by a single student. Conversation may still occur between students with other students agreeing with the student generating information; however, this type of dialogue does not constitute *co*-generation of information and therefore would still be coded as Constructive. This contrasts with the

Table 2 Codebook for question and group response codes

Question codes	
Active (A)	Information to answer the question can be found in the provided materials
Constructive/Interactive (C/I)	New information needs to be generated to answer the question prompt
Group response codes	
Active (A)	Conversation reflects that an answer was taken from information provided
Constructive (C)	One person provides the answer, generating new information. Can include forms of agreement from other group members (*e.g.*, head nods, "yeah", "uh-huh", *etc.*)
Interactive (I)	Participants generate information to answer the question based on one another's responses. Other participants' contributions of off-topic talk or forms of agreement are not included in this code

Chemistry Education Research and Practice

Paper

Interactive (I) code, where new information is generated through dialogue between two or more students. During the dialogue, each student contributed new information and each contribution built upon information previously generated in the conversation.

Mismatch between question and group response codes. Across all four activities and three groups, group responses were observed, coded, and compared to the corresponding question code. When the question code and the group response code were not the same, this was identified as an instance of mismatch. Since Constructive and Interactive engagement were a single code (*i.e.*, C/I) for the questions, if the corresponding group response was coded as Constructive or Interactive, either of these was considered a match. For each case of mismatch, the group conversation was examined inductively using conventional content analysis (Hsieh and Shannon, 2005) to determine if there were any themes that may explain the cause of the mismatch. To identify potential causes, each question and group response showing mismatch was read by two researchers. The researchers then independently identified specific phrases which were thought to contribute to the cause of the mismatch. The researchers then discussed these mismatch causes and combined common causes into themes.

Trustworthiness. Trustworthiness of the findings in this study was established through the evaluation of quality criteria such as qualitative reliability and credibility (Korstjens and Moser, 2018; O'Connor and Joffe, 2020). To enhance reliability in coding the questions and responses, a secondary coder was employed to evaluate the application of the codes in a two-stage process. The author S. Y. E. developed the codebook (Table 2), and both author S. Y. E. and the secondary coder first each individually coded each question and group response in a single activity. The coders met, discussed and resolved differences in coding, and came to consensus. Through the discussion to achieve consensus, the coders agreed that no modifications to the codebook were needed. The two coders then coded all the questions and group responses across the remaining activities. Inter-rater reliability (IRR) at each stage was evaluated by calculating Cohen's kappa (Cohen, 1960). During the first stage, the IRR values for question and group response coding of the single activity were 0.88 and 0.56, respectively. The IRR values for the subsequent question and group response coding across all remaining activities during the second

stage were 1.00 and 0.99, respectively. Kappa values greater than 0.8 are generally considered to have good reliability (Landis and Koch, 1977). For the identification of themes related to mismatched engagement levels between the questions and group responses, investigator triangulation (Lincoln and Guba, 1985) was used to establish credibility. Two of the authors (S. Y. E. and A. J. H.) used conventional content analysis to identify patterns in the transcripts and worked together to combine these patterns into themes.

Results and discussion

Question coding

Questions were coded as either Active (A) or Constructive/Interactive (C/I) based on how the information to answer would be derived (Table 2). Fig. 2 presents a portion of the model from the Solutions and Dilutions (SD) activity.

For example, Key Question 6 from the Solutions and Dilutions activity (SD-KQ6) was coded as Active because the information in the model (Fig. 2) explicitly states the required information in the text blurb and in the equation in the gray box at the top of the table.

(SD-KQ6) *When making a dilute solution, which of the following remains constant? (i) The concentration (ii) The moles of solute (iii) The volume of the solution.*

However, Key Question 9 from the same activity (SD-KQ9) asks students to provide an algebraic expression for M_D (*i.e.*, the molarity of the dilute solution). Since this question asks students to manipulate the equation in the model (Fig. 2), they would be generating new information. Therefore, SD-KQ9 was coded as a Constructive/Interactive question.

(SD-KQ9) *In preparing for an experiment, you need to know what the concentration of a dilute solution (M_D) will be. Provide an algebraic solution using the relation in the model for this concentration.*

In total, 68 questions were coded across the four activities (Table 3). Since the groups did not complete the activities in their entirety during the time allotted, the data includes only those activity questions which had a corresponding group response. Additionally, questions which were answered by both Groups A and B were counted only once. The overall results

---

Excerpt of Model from Solutions and Dilutions (SD) Activity

What are stock and dilute solutions? – Often times in lab you are given a stock solution, this solution typically has a higher concentration than you will need for your experiment. In running an experiment, you typically use dilutions made from this stock solution. *Dilutions* (or dilute solutions) are solutions with <u>lower concentrations</u> than the stock solution.

In a dilution, an aliquot of the stock aqueous solution is diluted by <u>adding more water</u>. In doing this, the volume of the dilute solution increases but the number of moles of solute remains constant. As the number of moles of solute remain constant, the following equality can be made.

Relation between concentrated and dilute solutions

$$M_C \times V_C = moles\ of\ solute = M_D \times V_D$$

M_C = molarity of **concentrated** solution	M_D = molarity of **dilute** solution
V_C = volume of **concentrated** solution	V_D = volume of **dilute** solution

Fig. 2 Portion of model from Solutions and Dilutions activity.

Table 3 Frequency of Active (A) *vs.* Constructive/Interactive (C/I) question coding by activity. Percentage of question codes per activity are given in parentheses

Activity	Question type	
	A	C/I
Mole and Molar Mass	2 (11)	16 (89)
Solutions and Dilutions	3 (23)	10 (77)
Electronegativity and Polarity	3 (17)	15 (83)
Thermal Energy and Calorimetry	5 (26)	14 (74)
Total	13 (19)	55 (81)

show that 13 questions were Active and 55 questions were Constructive/Interactive.

In general, the majority of questions (81%) were Constructive/Interactive questions across all activities. Table 3 shows that within the different activities, the percentage of questions coded as Active can vary, consisting of up to around one quarter of the total coded questions. Such variation was not captured in previous studies which coded at the activity level (Menekse *et al.*, 2013; Wiggins *et al.*, 2017; Henderson, 2019).

Group response coding

Group responses were coded as Active (A), Constructive (C), or Interactive (I) based on if more than one student contributed to the answer and whether their response(s): (1) generated new information, and (2) involved students building upon each other's statements to develop a final answer. In the conversation excerpts that follow, line numbers are used to allow for easy identification of pertinent portions of the text, information in parentheses refers to non-verbal actions, and information in square brackets has been added to the transcripts for clarity.

Excerpt 1 illustrates a group response that was coded as Active. In this excerpt, members of Group A are responding to SD-KQ6. Beth's comment (line 261) mentions looking at the equation which is a reference to the model (Fig. 2); therefore, this group response was coded as Active.

(SD-KQ6) *When making a dilute solution, which of the following remains constant? (i) The concentration (ii) The moles of solute (iii) The volume of the solution*

> Excerpt 1: Group response to SD-KQ6, coded as Active
>> 260 KATIE: Okay. So, key question six: "When making a dilute solution, which of the following remains constant?" Circle your response: "One, the concentration, two, the moles of the so-, solute, or three, the volume of the solution."
>> 261 BETH: It looks from the equation [in the model] that the moles of the solute stay constant.
>> 262 NANI: Yeah.

Excerpt 2, on the other hand, illustrates a group response that was coded as Constructive. This excerpt focuses on Group C's response to Key Question 3 from the Thermal Energy and Calorimetry activity (TEC-KQ3), where students are asked to explain the difference in heat capacity between two blocks. Fig. 3 presents a portion of the model from the Thermal Energy and Calorimetry (TEC) activity.
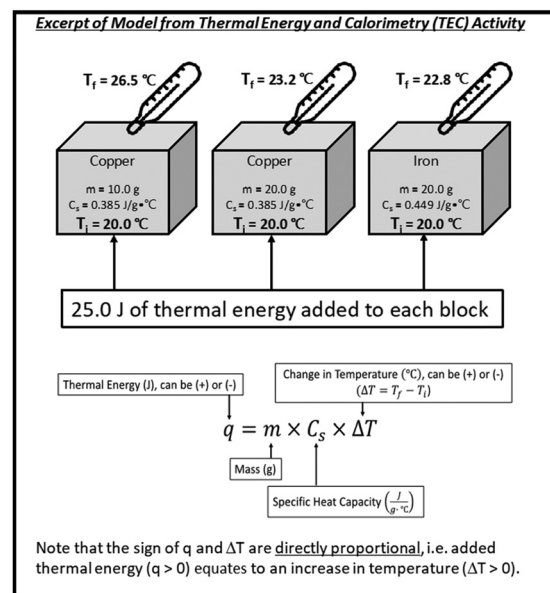

Fig. 3 Portion of model from the Thermal Energy and Calorimetry (TEC) activity.

In Excerpt 2, Helen provides the answer to the question associated with this portion of the model (line 52), and the contributions from Nani and Grace are forms of agreement (lines 53 and 54). Therefore, Helen is the only student generating new information and this group response was coded as Constructive.

(TEC-KQ3) *How does the difference in specific heat capacity between blocks 2 and 3 relate to their final temperature? Briefly explain.*

> Excerpt 2: Group response to TEC-KQ3, coded as Constructive
>> 51 GRACE: So, "How does the difference in specific heat capacity between blocks two and three relate to their final temperature?"
>> 52 HELEN: So it, it's the same as mass, right? So, like a greater specific heat capacity will result in a lower final temperature.
>> 53 GRACE: Yeah.
>> 54 NANI: (nods).
>> 55 HELEN: So, so block two will have a greater final temperature.
>> 56 GRACE: Mm hmm.

Excerpt 3 gives an example where the coding of the group response was ambiguous. In this excerpt, students from Group A respond to Key Question 7 from the Solutions and Dilutions activity (SD-KQ7). Although the answer is present in the model (Fig. 2), and Katie gives the correct answer (line 271), it is unclear from the conversation whether Katie's response was based on the information in the model (Active) or she generated new knowledge (Constructive). In the absence of evidence that the response came from the model, it was assumed that she generated new knowledge and the group response was coded as Constructive.

(SD-KQ7) *When making a dilute solution, which of the following* ***decreases****? Circle your response. (i) The concentration (ii) The moles of solute (iii) The volume of the solution*

Excerpt 3: Ambiguous group response to SD-KQ7, coded as Constructive

> 270 BETH: Okay. The sec-, or the seventh quest-, seven, seventh key question is, um, "When making a dilu-, dilute solution, which of the following decreases, circle your response? Um, one, the concentration, two, the moles of the solute, or three, the volume of solution."
>
> 271 KATIE: Wouldn't it be the concentration since we're diluting it?
>
> 272 BETH: Yeah, I think so.

Excerpt 4 shows an example of Interactive engagement, where Katie, Leslie, and Beth all contribute new information to solving the calculation in Exercise 5 from the Solutions and Dilutions activity (SD-EX5). Leslie and Katie start by determining what variable they are solving for (lines 400 and 401). Leslie then builds on this by identifying the numerical value for M_C, and Katie further contributes new knowledge by mentioning the form of the equation they should use to solve (lines 403 and 404). Beth further builds on this knowledge by providing the numerical solution (line 408). Since Leslie, Katie, and Beth all contribute pieces of information to answer the question and each of their statements builds upon the previous student's comment, this response was coded as Interactive.

(SD-EX5) *What is the concentration when 11.75 mL of 0.375 M sucrose is diluted to 50.0 mL?*

Excerpt 4: Group response to SD-EX5, coded as Interactive

> 400 KATIE: Ok, "What is the concentration when 11.75 milliliters of 0.375 molarity or mole?" I don't even know. Sucrose is diluted to 50 milliliters. Okay. So now we're trying to find MC. Again, MC.
>
> 401 LESLIE: No, we're find-, we're trying to find, MD now.
>
> 402 BETH: Yeah. I think MD.
>
> 403 LESLIE: Cause MC is that 0.375.
>
> 404 KATIE: Oh yeah, so we're finding...so we would do our MC times VC divided by VD then?
>
> 405 BETH: Yeah.
>
> 406 LESLIE: Did you guys get there?
>
> 407 KATIE: Just about...Oh geez!
>
> 408 BETH: Do you guys get 0.0881?
>
> 409 LESLIE: Mm hmm.

In total, 101 group responses were coded (Table 4). Groups A and B have a different number of response codes for each activity because they moved at different speeds and therefore did not answer the same number of questions. As with the question coding, since students did not complete the activities during the time allotted, coded responses are only for completed questions, not all questions in the activity. Overall, group responses were distributed across the three engagement modes with 8 responses coded as Active, 32 responses coded as Constructive, and 61 responses coded as Interactive. Results indicate that Interactive group responses ranged from 64% to 87% for Group A and from 39% to 77% for Group B across the Mole and Molar Mass, Solutions and Dilutions, and Electronegativity and Polarity activities. Only Group C completed the Thermal Energy and Calorimetry activity, and only 58% of their

Table 4 Frequency of group response codes by activity and group. Percentages of group response codes by activity and group are given in parentheses

Activity	Group	Response type A	C	I
Mole and Molar Mass	A	0 (0)	1 (13)	7 (87)
	B	1 (6)	7 (39)	10 (55)
Solutions and Dilutions	A	1 (9)	3 (27)	7 (64)
	B	1 (8)	2 (15)	10 (77)
Electronegativity and Polarity	A	0 (0)	5 (36)	9 (64)
	B	2 (11)	9 (50)	7 (39)
Thermal Energy and Calorimetry	C	3 (16)	5 (26)	11 (58)
Total		8 (8)	32 (32)	61 (60)

responses during this activity reached the level of Interactive engagement. Overall, observed engagement levels across groups and across questions within an activity varied widely.

Matches between question and group response codes

A total of 68 questions (Table 3) and 101 group responses (Table 4) were coded across the three groups and four activities. We began the comparison between coding groups by examining the questions coded as Constructive/Interactive and their corresponding group responses. Table 5 shows the breakdown of the frequency of Constructive/Interactive coded questions by activity and group. It also shows how the group responses were distributed across the Constructive and Interactive codes. These results indicate that when the question was coded as Constructive/Interactive, all the group response codes were either Constructive or Interactive, indicating a match with this question code but different levels of engagement. Across all groups and activities, the portion of group responses coded as Interactive ranged from 40% to 90%. In total, just over two-thirds of the responses were coded at the level of Interactive engagement.

In addition to variation in response coding seen across activities, variation was also observed across groups (Table 6). For groups A and B, who completed the same three activities, several of the response codes differed across the two groups on the questions that both groups completed. For example, Table 6

Table 5 Breakdown of frequency of Constructive (C) and Interactive (I) question and group response codes by activity and group. Percentages of Constructive (C) and Interactive (I) group responses are given in parentheses

Activity	Group	Question codes C/I	Response codes C	I
Mole and Molar Mass	A	8	1 (13)	7 (87)
	B	16	7 (44)	9 (56)
Solutions and Dilutions	A	8	1 (13)	7 (87)
	B	10	1 (10)	9 (90)
Electronegativity and Polarity	A	11	4 (36)	7 (64)
	B	15	9 (60)	6 (40)
Thermal Energy and Calorimetry	C	14	5 (36)	9 (64)
Total		82	27 (33)	55 (67)

Table 6 Distribution of response matches between groups A and B across the Constructive/Interactive (C/I) questions that were answered by both groups (noted in parentheses)

	Response codes	
Activity	C	I
Mole and Molar Mass (8)	0	6
Solutions and Dilutions (8)	1	5
Electronegativity and Polarity (11)	2	3

Table 7 Breakdown of frequency of Active (A) question and corresponding group response codes by activity and group

		Question codes	Response codes		
Activity	Group	A	A	C	I
Mole and Molar Mass[a]	A	0	0	0	0
	B	2	1	1	0
Solutions and Dilutions	A	3	1	2	0
	B	3	1	1	1
Electronegativity and Polarity	A	3	0	1	2
	B	3	2	0	1
Thermal Energy and Calorimetry	C	5	3	0	2
Total		19	8	5	6

[a] Groups A and B have different numbers of active questions because Key Questions 1–4 were assigned prior to class, and Group A did not discuss them while Group B went over them as a group before proceeding.

shows that on the 8 completed questions coded as Constructive/Interactive in the Mole and Molar Mass activity, the responses of groups A and B only overlapped on 6 question responses, all coded as Interactive. The fewest matches between groups were observed on the 11 Electronegativity and Polarity questions, with only 5 of the response codes matching.

Upon comparison of question codes to the response codes of each group, mismatches were found exclusively in questions coded as Active. A breakdown of the frequency of questions and group responses coded as Active is shown in Table 7. While 19 total questions were coded as Active, only 8 responses were also coded as Active, a 42% match. This means that more than half of the questions coded as Active had a mismatch with their

corresponding group response codes, where students were responding at a higher engagement mode than was indicated by the question design. Among the 11 Active questions which showed a higher group response engagement mode, the responses split almost evenly between Constructive (5) and Interactive (6) engagement.

To further investigate these mismatches, conventional content analysis was used to identify the potential causes by examining each mismatched question and group response for specific phrases that identified the source of the mismatch. Causes were then collected into common themes. Table 8 summarizes these results. Each of the questions in these mismatched cases was coded as Active because the information to answer the question was explicitly available in the activity.

Themes relating to mismatch

Conventional content analysis was used to investigate each of the group responses for details that explain the higher level of engagement displayed by the conversation compared to the question. The analysis suggested three possible themes: model use, unfamiliar vocabulary, and molecular representations. Although Key Question 7 from the Solutions and Dilutions activity (SD-KQ7) and Key Question 4 from the Electronegativity and Polarity activity (EP-KQ4) showed a mismatch, our inductive analysis did not suggest that the cause of mismatch in these cases falls into one of the identified themes. The group responses on these items were deemed to be ambiguous because it was not clear from the conversation if the students' response was taken from the activity material.

Theme 1: Model use. Three of the 11 instances of mismatch were due to improper model use. These cases occurred during the Thermal Energy and Calorimetry (TEC) and the Mole and Molar Mass (MM) activities. Because the answers to these questions were explicitly stated in the model, it was expected that the students would use the model to answer these questions, and that the group conversation would show evidence of this.

For example, in Excerpt 5, Group C responds to Key Questions 4 and 5 from the Thermal Energy and Calorimetry activity (TEC-KQ4 and TEC-KQ5). Since the answers to both these questions are explicitly stated in the model (Fig. 3), these

Table 8 Frequency of mismatched question and response codes with associated themes

Theme	Activity	Group	Mismatched cases
Model use	Thermal Energy and Calorimetry	C	2
	Mole and Molar Mass	B	1
Unfamiliar vocabulary	Solutions and Dilutions[a]	A	1
		B	1
Molecular representations	Electronegativity and Polarity[a]	A	2
		B	1
Ambiguous	Solutions and Dilutions[a]	A	1
		B	1
	Electronegativity and Polarity	A	1

[a] The mismatched cases in these activities occurred on the same questions in Groups A and B.

questions are coded as Active. Although the group response to TEC-KQ4 did refer to the model and was coded as Active, the response was incomplete. The correct response should have included ΔT and q, but Helen and Grace used the model to decide that the answer should only include ΔT (lines 68–71). Because of this incomplete use of the model, Helen and Grace engaged interactively to answer the next question in the activity, TEC-KQ5, which built upon the aspects of the model high-lighted in TEC-KQ4. This interaction starts from line 72 and Grace's realization that they need two variables. From there, Helen builds upon this, suggesting the two variables are T_i and T_f (line 73). Although the final answer they come to is incorrect, one can see that it is the incomplete use of the model in TEC-KQ4 which prompts the Interactive engagement in TEC-KQ5.

TEC-KQ4: *When mathematically determining q, which variables can be positive or negative?*

TEC-KQ5: *How are the two variables in KQ4 related?*

    Excerpt 5: Example of incomplete model use

68 GRACE: It s-, it shows at the top model [referring to the model in Box 2], which ones. So…

69 HELEN: Yeah, it does. So only ΔT.

70 GRACE: Yeah. ΔT, and if you want to include the thermal energy, you could say that, but we're already talking about it, so…

71 HELEN: Yeah. I don't think you would include q.

72 GRACE: And then, "How are the two variables related?" Um…Oh, they said the two variables. Okay. So, you can't include q. It's the same thing…

73 HELEN: Okay. No, no. So, it's temperature final and temperature initial.

74 GRACE: Oh, those are the two variables. Ohhh…

75 HELEN: Yeah.

76 GRACE: Okay. Never mind. Um, But the temp…Oh yeah. The temp can be negative.

77 GRACE: Well if one's, if one's, it depends on which one, if the final's higher than the initial, then you get a positive number. If the initial's higher than the final, you get a negative number. So I suppose that's how it's related…right.

    Theme 2: Unfamiliar vocabulary. Two of the 11 instances of mismatch involved students' use of unfamiliar vocabulary, specifically the scientific term "aliquot" in Key Question 8 of

the Solutions and Dilutions activity (SD-KQ8). Although this question is coded as Active because the information to answer the question is explicit in the model (Fig. 2), responses from both Groups A and B display a higher mode of engagement due to unfamiliarity with the term "aliquot". For example, in Excerpt 6, the higher engagement mode of Group B's response is prompted by Helen's question about the meaning of "aliquot." Jacob responds and Grace looks up the definition, ostensibly on Google (lines 166–168). It is evident that the Interactive engagement resulted from unfamiliarity with the term "aliquot".

(SD-KQ8) *In a dilution, which is always **larger**? Circle your response. (i) The volume of the aliquot (ii) The volume of the final solution.*

    Excerpt 6: Example of unfamiliar vocabulary

166 HELEN: I know it's the second one, but what exactly is the ali- aliquot? Cause I know [XXX] fairly small, so small sample or whatever.

167 JACOB: I guess the aliquot would be, do you think it would be the given volume?

168 GRACE: I'm just looking it up.

169 JACOB: Fair enough.

170 HELEN: What does Google say?

171 GRACE: A portion of a larger whole, a specific sample taken for chemical analysis or other treatment. I think it's like a portion of the sample. So the portion is obviously going to have less.

172 JACOB: So in a dilution, which is yeah, the volume of final solution will be larger.

    Theme 3: Molecular representations. Communicating complex scientific ideas is dependent on using multiple "languages of science", which may include symbolic, graphical, or mathematical representations (Osborne, 2010). Four of the 11 instances of mismatch involved students' struggles in moving between different representations in the Electronegativity and Polarity activity. Fig. 4 depicts a portion of the model from this activity.

    Key Question 8 from the Electronegativity and Polarity activity (EP-KQ8) asks students to explain why DL_2 is a polar molecule. Since this information is depicted in the model (Fig. 4), this question is coded as Active. Students in Groups A and B seemed to have difficulty moving between the Lewis structure representation and bond dipole representation of molecules. In Excerpt 7, the Interactive engagement of Group A is prompted by Beth asking about the number of arrows that
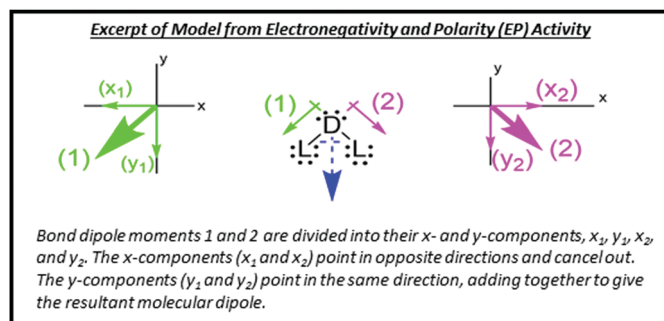

Fig. 4   Portion of model from the electronegativity and polarity (EP) activity.

should be drawn (line 285). Leslie builds on the question explaining that she drew three arrows (line 286), and the instructor (INST in excerpt) builds further, adding new information that there should be two component arrows for each bond dipole (line 290).

EP-KQ8: *Using the blank Cartesian coordinate system, draw the x- and y-components of each bond and use them to explain why DL₂ is a polar molecule.*

Excerpt 7: Example of molecular representations (Group A)

285 BETH: For this one, do you only draw two arrows or should there be more than two?

286 LESLIE: I'm doing three for that one. So like the two going on the X and then the one going down for the Y.

287 BETH: OK.

288 INST: [Key Question] Eight. Okay. And are you looking at it or have you talked about it?

289 BETH: Um, we've talked about how many arrows to draw and um, I think we decided on drawing like three arrows. Uh, I drew like two, um, on the X axis, like going different directions and then one down on the Y axis.

290 INST: Okay. So for each of the diagonal arrows, they have both an X and a Y component. Yeah. So the downward, yes. I see what, you're what you're drawing, Katie. So you, so you have for each of the diagonals, you have an X and a Y. And so for this one, you have an X and a Y. So you actually have two downward arrows on the Y axis.

291 BETH: Two downwards? Ok.

While the Interactive engagement in Excerpt 7 was prompted by difficulty in translating between the Lewis structure and the representation depicting bond dipoles, in Excerpt 8, we see a desire to understand more deeply the role of specific features of the Lewis structure (*i.e.*, lone pairs of electrons) in the dipole representation is the trigger for the Interactive engagement. In Excerpt 8, Group B engages interactively to try to gain a deeper understanding of what the vector model of dipoles represents. Their response to the same question begins with a discussion of the Lewis structure to identify the molecular geometry (lines 368–371). From there, they reference the model to determine how to draw the components of the bond dipoles (lines 373–386). Lines 387–396 show the group generating new information as they attempt to make the connection between the lone pairs of electrons in the Lewis structure and the bond dipoles. In lines 385 and 386, both Helen and Jacob directly refer to Fig. 4 in the model, stating that the answer is there (Active engagement). However, Grace's desire to understand how the lone pairs fit into the vector representation causes the group to engage at the higher Interactive mode (lines 387 and 393). In both groups' conversations, it is apparent that the students attempting to move from the Lewis structure representation of the molecule to the vector model of bond dipoles is the trigger for the higher mode of engagement.

Excerpt 8: Example of molecular representations (Group B)

368 GRACE: Oh, and this one has lone pairs. What kind of structure does that make?

369 JACOB: The chart's...DL2, lone pairs.

370 JACOB: It's bent.

371 GRACE: I think bent?

372 JACOB: Yeah.

373 JACOB: Cause if we're looking at the model, um, the model gives like the best description of it above, uh, for the DL2. So net molecular dipole due to bent geometry. And it shows you below what that bent geometry looks like on the planes.

374 GRACE: So for this we're doing four.

375 JACOB: And it's asking us why it's polar.

376 GRACE: Oh, I'm assuming they don't cancel each other out.

377 HELEN: The left and right aspects do, but they still have a net, like, down.

378 GRACE: Wait, what?

379 JACOB: It has a net molecular dipole.

380 GRACE: Yeah. No...Have you guys started drawing the coordinate? I don't know how they'll look. Are they pointing down Y?

381 JACOB: They're pointing down, yeah, Y.

382 GRACE: Okay. At what angle?

383 HELEN: Like 45ish each in the third and fourth quadrant.

384 JACOB: Like it's coming out of the origin.

385 HELEN: I mean, it's just like the green and pink arrows in Fig. 4 is what I drew. But on one axis or like one....

386 JACOB: Same. I simply, I literally don't know why, like I know, but I also like don't know, so I just looked at Fig. 4 that has the answer so... Well, it has like what we're supposed to be gathering from it.

387 GRACE: What about the lone pairs?

388 JACOB: Um, it shows in Fig. 4, like kind of, uh, the lone pairs are kind of like on the arrows or like, do you see Fig. 4?

389 GRACE: Oh yeah.

390 JACOB: So that's kind of what Fig. 4 does with the...

391 GRACE: So I've got two of them with the arrows pointing opposite ways in the third and fourth quadrants.

392 JACOB: Yes.

393 GRACE: What are the ones for the lone pairs?

394 HELEN: It just says of each bond. I don't think you have to worry about the lone pairs.

395 JACOB: And then ask why it's polar. And um, like Helen said, it doesn't cancel because of the net molecular dipole.

396 HELEN: I said it's polar because though the dipoles cancel out in the x-direction, they have a net downward dipole moment. I don't think that's like correct language, but...

397 JACOB: I mean, I think it's it, but it gets your point across.

Conclusion

Previous studies using the ICAP framework of cognitive engagement to investigate active learning environments assumed a single

engagement mode for the entire activity (Wiggins *et al.*, 2017; Henderson, 2019). However, the data examined above suggest that students may engage differently with different parts of an activity. In addition, some studies have also assumed an engagement mode based on the activity design instead of overt student behavior (Menekse *et al.*, 2013; Wiggins *et al.*, 2017). ICAP identifies engagement modes based on student behavior, and as seen above, it may not be accurate to assume the expected engagement mode based on activity design would be the same as the observed engagement mode based on student behaviors. To address these concerns, we used ICAP to investigate cognitive engagement of student groups during AL activities in answering the following research questions.

RQ1: *What range of engagement modes are expected during a general chemistry AL activity based on the question design?*

This study used a finer grain size, *i.e.*, identifying engagement modes at the question level rather than the activity level. Results indicated that across the four activities observed, the majority of questions (81%) were designed to elicit Constructive or Interactive engagement. Investigation at this finer grain size confirms that not all questions were designed with the same mode of engagement in mind, and therefore studies which assume a single engagement mode for the entire activity may miss insights that can be seen when looking at engagement at the question level.

RQ2: *What range of engagement modes are observed during a general chemistry AL activity based on students' physical and verbal behaviors during group conversations?*

The study also identified observed engagement modes of student groups by using ICAP to examine group conversations. Results indicated that within a single activity, the engagement of the group based on their conversation varied from Active to Interactive, with the majority of the group responses (60%) showing Interactive engagement. Additionally, within each group, the percentage of Interactive responses was not consistent across all activities (64–88% for Group A; 39–77% for Group B). These results provide further evidence that coding engagement at the question level for both questions and responses can give insight into students' engagement which is lost when coding at the activity level.

RQ3: *If mismatches occur between the expected and observed levels of cognitive engagement, what themes account for this mismatch?*

By comparing the expected engagement mode based on the question design with the observed engagement mode based on the group responses, cases of mismatch were identified. The group conversations were then further investigated using qualitative content analysis for common themes that caused the mismatches. Results suggested that the causes of the higher than expected observed engagement levels were related to three themes: model use, unfamiliar vocabulary, and struggles with different molecular representations.

Limitations

Due to the small sample size used in this study, these results are not generalizable to large populations. Additional studies

are being conducted in author Barbera's research group to provide more generalizable insights into students' engagement in small group learning activities. Since the observed groups were recorded through Zoom, we were unable to see what students were writing unless papers were held up to the camera. Because of this limitation, engagement modes of groups were based solely on the group conversation. However, being able to see what students were writing on their worksheets could have provided additional insight into their cognitive engagement. Future data collections will take place in person and will be able to account for these actions. Finally, the coding of activity questions according to ICAP was based solely on design features present in each question and not explicitly on any stated intention on the part of the activity designers. Therefore, although the activity questions may have been written to elicit a specific type of thinking or engagement on the part of the students, the questions could only be coded based on specific features that were present in the questions themselves.

Implications for instructors

Results of this study showed that there were multiple instances of Constructive or Interactive engagement occurring in Key Questions where Active engagement was expected. Incomplete or lack of model use was one reason for this. In some cases, this resulted in students engaging at a higher level but obtaining an incorrect answer. While many instructors discuss the structure of and expectations for these types of learning activities at the start of a term, we would suggest that instructors regularly remind students to read through the model prior to answering any questions in the worksheet and to refer back to it in their responses. This would reinforce the purpose of the models and may focus the groups' conversations on the data and details within the materials.

Use of new and potentially unfamiliar scientific terms can possibly promote students' curiosity and potentially lead to higher modes of engagement. This idea was supported in this study where use of the unfamiliar term "aliquot" resulted in more conversation and a higher engagement mode. Although there is the danger that discussion of such vocabulary could result in unhelpful, tangential conversations, group discussions around the term "aliquot" seemed to help students reason out an answer to the question. In addition, learning relevant new vocabulary is essential to students' growth as scientists. Therefore, use of unfamiliar vocabulary that is relevant to the concept being taught can be a useful tool to promote student learning.

It should be noted that although ICAP states that cognitive engagement increases as one moves from Passive to Active to Constructive to Interactive, it should not be inferred that Interactive engagement is always the most desirable. As shown in this study, these higher than expected modes were due to a variety of factors that could provide insight to future improvements in the activities or instructional practices. Worksheets for these activities were structured such that students begin with Key Questions which are designed to orient students to the

pertinent information in the model (*i.e.*, Active engagement), followed by Exercises and Problems, which allow students to manipulate and apply the information in a more advanced manner (*i.e.*, Constructive or Interactive engagement). By scaffolding worksheets in such a manner, students use knowledge gained at the lower engagement modes to foster a deeper understanding during the more complex Exercises and Problems.

Implications for research

Investigation of student conversations using qualitative content analysis has opened avenues of further exploration. While this study looked at the engagement mode of the group as a whole, it is apparent that not all participants within a group are engaging to the same degree. For example, in Group A, Nani was a very quiet student who rarely contributed to conversations but was always writing on her worksheet and nodding along with other students' statements. Exploring the individual students' engagement could provide insight into how a student's engagement correlates with learning outcomes. Other factors such as group dynamics and how these dynamics change over time may also be understood by analyzing the engagement of each individual. In addition, further exploration into the root causes of the identified mismatch themes can be explored. For example, the unfamiliar vocabulary theme could be due to differences in prior knowledge that students bring to the activity. Research in this area could increase understanding of how prior knowledge affects students' engagement in small-group activities.

Conflicts of interest

There are no conflicts of interest to declare.

Acknowledgements

References

Andrews T. M., Leonard M. J., Colgrove C. A. and Kalinowski S. T., (2011), Active learning not associated with student learning in a random sample of college biology courses, *CBE-Life Sci. Educ.*, **10**(4), 394–405.

Atkin J. M. and Karplus B., (1962), Discovery or invention?, *Sci. Teach.*, **29**(5), 45–51.

Chi M. T. H., (2009), Active-constructive-interactive: A conceptual framework for differentiating learning activities, *Top. Cogn. Sci.*, **1**(1), 73–105.

Chi M. T. H. and Wylie R., (2014), The ICAP framework: Linking cognitive engagement to active learning outcomes, *Educ. Psych.*, **49**(4), 219–243.

Chi M. T. H., Adams J., Bogusch E. B., Bruchok C., Kang S., Lancaster M., Levy R., Li N., McEldoon K. L., Stump G. S., Wylie R., Xu D. and Yaghmourian D. L., (2018), Translating the ICAP theory of cognitive engagement into practice, *Cogn. Sci.*, **42**(6), 1777–1832.

Cohen J., (1960), A coefficient of agreement for nominal scales, *Educ. Psych. Meas.*, **20**(1), 37–46.

Cole R. S., Becker N. and Stanford C., (2014), Discourse analysis as a tool to examine teaching and learning in the classroom, in Bunce D. M. and Cole R. S. (ed.), *ACS Symposium Series*, Washington, DC: American Chemical Society, ch. 4, pp. 61–81.

Cooper M. M., (2016), It is time to say what we mean, *J. Chem. Educ.*, **93**, 799–800.

Current K. and Kowalske M. G., (2016), The effect of instructional method on teaching assistants' classroom discourse, *Chem. Educ. Res. Pract.*, **17**, 590–603.

Dohrn S. W. and Dohn N. B., (2018), The role of teacher questions in the chemistry classroom, *Chem. Educ. Res. Pract.*, **19**, 352–363.

Eddy S. L. and Hogan K. A., (2014), Getting under the hood: How and for whom does increasing course structure work? *CBE Life Sci. Educ.*, **13**, 453–468.

Farrell J. J., Moog R. S. and Spencer J. N., (1999), A guided-inquiry general chemistry course, *J. Chem. Educ.*, **76**(4), 570–574.

Freeman S., Eddy S. L., McDonough M., Smith M. K., Okoroafor N., Jordt H. and Wenderoth M. P., (2014), Active learning increases student performance in science, engineering, and mathematics, *Proc. Natl. Acad. Sci. U. S. A.*, **111**, 8410–8415.

Haak D. C., Lambers J. H. R., Pitre E. and Freeman S., (2011), Increased structure and active learning reduce the achievement gap in introductory biology, *Science*, **332**, 1213–1216.

Hanson D. M., Goodwin J. and Phillips M., (2018), *Foundations of chemistry: Applying POGIL principles*, Pacific Crest Publishing.

Henderson J. B., (2019), Beyond "active learning": How the ICAP framework permits more acute examination of the popular peer instruction pedagogy, *Harv. Educ. Rev.*, **89**(4), 611–634.

Hsieh H. F. and Shannon S. E., (2005), Three approaches to qualitative content analysis, *Qual. Health Res.*, **15**(9), 1277–1288.

Korstjens I. and Moser A., (2018), European Journal of General Practice Series: Practical guidance to qualitative research. Part 4: Trustworthiness and publishing, *Eur. J. Gen. Pract.*, **24**(1), 120–124.

Kuh G., Kinzie J., Schuh J. and Witt E., (2005), *Student sucess in college: Creating conditions that matter*, Washington DC: Association for the Study of Higher Education.

Kulatunga U. and Lewis J. E., (2013), Exploration of peer leader verbal behaviors as they intervene with small groups in college general chemistry, *Chem. Educ. Res. Pract.*, **14**, 576–588.

Kulatunga U., Moog R. S. and Lewis J. E., (2013), Argumentation and participation patterns in general chemistry peer-led sessions, *J. Res. Sci. Teach.*, **50**(10), 1207–1231.

Landis J. R. and Koch G. G., (1977), The measurement of observer agreement for categorical data, *Biometrics*, **33**(1), 159–174.

Lim J., Ko H., Yang J. W., Kim S., Lee S., Chun M.-S., Ihm J. and Park J., (2019), Active learning through discussion: ICAP framework for education in health professions, *BMC Med. Educ.*, **19**(1), 1–8.

Lincoln Y. S. and Guba E. G., (1985), *Naturalistic inquiry*, Sage.

Liyanage D., Lo S. M. and Hunnicutt S. S., (2021), Student discourse networks and instructor facilitation in process oriented guided inquiry physical chemistry classes, *Chem. Educ. Res. Pract.*, **22**(1), 93–104.

Lorenzo M., Crouch C. H. and Mazur E., (2006), Reducing the gender gap in the physics classroom, *Am. J. Phys.*, **74**(2), 118–122.

Menekse M. and Chi M. T. H., (2019), The role of collaborative interactions versus individual construction on students' learning of engineering concepts, *Eur. J. Eng. Educ.*, **44**(5), 702–725.

Menekse M., Stump G. S., Krause S. and Chi M. T. H., (2013), Differentiated overt learning activities for effective instruction in engineering classrooms, *J. Eng. Educ.*, **102**(3), 346–374.

Moon A., Stanford C., Cole R. and Towns M., (2016), The nature of students' chemical reasoning employed in scientific argumentation in physical chemistry, *Chem. Educ. Res. Pract.*, **17**(2), 353–364.

National Research Council, (2012), *Discipline-based education research: Understanding and improving learning in undergraduate science and engineering*, Washington, DC: The National Academies Press.

Osborne J., (2010), Science without literacy: A ship without a sail?, *Cambridge J. Educ.*, **32**(2), 203–218.

O'Connor C. and Joffe H., (2020), Intercoder reliability in qualitative research: Debates and practical guidelines, *Int. J. Qual. Methods*, **19**, 1–13.

President's Council of Advisors on Science and Technology (PCAST), (2012), *Engage to excel: Producing one million additional college graduates with degrees in science, technology, engineering, and mathematics*.

Rahman T. and Lewis S. E., (2020), Evaluating the evidence base for evidence-based instructional practices in chemistry through meta-analysis, *J. Res. Sci. Teach.*, **57**(5), 765–793.

Repice M. D., Sawyer R. K., Hogrebe M. C., Brown P. L., Luesse S. B., Gealy D. J. and Frey R. F., (2016), Talking through the problems: A study of discourse in peer-led small groups, *Chem. Educ. Res. Pract.*, **17**, 555–568.

Schreier M., (2012), *Qualitative content analysis in practice*, Sage.

Shortlidge E. E., Rain-Griffith L., Shelby C., Shusterman G. P. and Barbera J., (2019), Despite similar perceptions and attitudes, postbaccalaureate students outperform in introductory biology and chemistry courses, *CBE-Life Sci. Educ.*, **18**(1), 1–14.

Shultz G. V. and Li Y., (2016), Student development of information literacy skills during problem-based organic chemistry laboratory experiments, *J. Chem. Educ.*, **93**, 413–422.

Stanford C., Moon A., Towns M. and Cole R. S., (2016), Analysis of instructor facilitation strategies and their influences on student argumentation: A case study of a process oriented guided inquiry learning physical chemistry classroom, *J. Chem. Educ.*, **93**, 1501–1513.

Villalta-Cerdas A. and Sandi-Urena S., (2014), Self-explaining effect in general chemistry instruction: Eliciting overt categorical behaviours by design, *Chem. Educ. Res. Pract.*, **15**(4), 530–540.

Warfa A.-R. M., Roehrig G. H., Schneider J. L. and Nyachwaya J., (2014), Role of teacher-initiated discourses in students' development of representational fluency in chemistry: A case study, *J. Chem. Educ.*, **91**, 784–792.

Wiggins B. L., Eddy S. L., Grunspan D. Z. and Crowe A. J., (2017), The ICAP active learning framework predicts the learning gains observed in intensely active classroom experiences, *AERA Open*, **3**(2), 1–14.

Wood L. A. and Kroger R. O., (2000), *Doing discourse analysis: Methods for studying action in talk and text*, Thousand Oaks: Sage.

Xu H. and Talanquer V., (2013), Effect of the level of inquiry on student interactions in chemistry laboratories, *J. Chem. Educ.*, **90**, 29–36.

Young K. K. and Talanquer V., (2013), Effects of different types of small-group activities on students' conversations, *J. Chem. Educ.*, **90**, 1123–1129.