Chemistry critical friendships: Investigating for chemistry-specific discourse within a domain-general discussion of best practices for inquiry assessments

Journal:	*Chemistry Education Research and Practice*
Manuscript ID	RP-ART-10-2019-000245.R1
Article Type:	Paper
Date Submitted by the Author:	03-Dec-2019
Complete List of Authors:	Schafer, Adam; Miami University, Chemistry & Biochemistry Yezierski, Ellen; Miami University, Chemistry & Biochemistry

SCHOLARONE™
Manuscripts

Chemistry critical friendships: Investigating for chemistry-specific discourse within a domain-general discussion of best practices for inquiry assessments

Adam G. L. Schafer, Ellen J. Yezierski*

Miami University, Department of Chemistry and Biochemistry, Oxford, OH

ABSTRACT

High school chemistry teachers struggle to use assessment results to inform instruction. In the absence of expert assistance, teachers often look to their peers for guidance and support; however, little is known about the assessment beliefs and practices of high school chemistry teachers or the discourse mechanisms used as teachers support one another. Presented in this paper are the results from analyzing a discussion between five high school chemistry teachers as they generated a set of best practices for inquiry assessments. To analyze the discussion, a novel representation called a discourse map was generated to align the analyses conducted on chemistry teacher discourse as they temporally occurred. Results show the utility of the discourse map for evidencing critical friendship and assessment practices evoked by the teachers during the discussion of best practices. Implications for the structural considerations of materials and chemistry teacher professional development are presented as well as potential future investigations of teacher discourse regarding the use of data to inform instruction.

INTRODUCTION

Chemistry educators face increasing pressure from federal, state, and local levels to improve student achievement and better monitor student growth (Knapp *et al.*, 2005; Darling-Hammond *et al.*, 2012; NGSS Lead States, 2013). Educators spend a significant amount of time on the difficult tasks of designing assessments and interpreting assessment results (Stiggins, 1988; Towndrow *et al.*, 2010; Remesal, 2011; Smith, 2013; Harshman and Yezierski, 2016). However, school districts and teachers claim that teachers' ability to develop assessments, interpret assessment results, and use assessment data to guide instruction is limited by inadequate teacher preparation (Buck *et al.*, 2010; Towndrow *et al.*, 2010; Smith, 2013). Few content-specific materials exist to assist teachers in developing skills using data to inform day-to-day instruction (Hamilton *et al.*, 2009; Harshman and Yezierski, 2017). Without proper support, chemistry teachers are left to develop assessment skills through trial and error. Alternatively, teachers' peers could lend support in the form of advice or materials. Little is known about how teachers support each other in facilitating changes to assessment practices; however, many agree that tapping peer knowledge can be a productive way for high school chemistry teachers to improve assessment skills (Stiggins, 1988; Black and Wiliam, 1998; Heady, 2000; Bell and Cowie, 2001; Swaffield, 2004; Towndrow *et al.*, 2010; Fletcher *et al.*, 2016). The goal of this investigation is to deepen the understanding about United States high school chemistry teachers' domain-general and chemistry-specific beliefs and practices about assessments, as revealed through their perceived best practices, to help teachers consider potential chemistry-specific improvements to practice, even when faced with domain-general guidance and feedback.

Critical Friends

Productive professional development enables teachers to interact with their peers to test practices and beliefs, a process that occurs optimally when teachers have developed critical friendships. Critical friends are members of a professional community aimed at fostering members' instructional improvement by collegial conversations about teaching and learning (Dunne and Honts, 1998; Swaffield, 2004; Schuck and Russell, 2005; Curry, 2008; Taylor and Storey, 2013; Moore and Carter-Hicks, 2014). Critical friends evaluate each other's work, participate in intellectually engaging discourse, and/or collectively engage in learning (Dunne and Honts, 1998; Swaffield, 2004; Schuck and Russell, 2005; Curry, 2008; Taylor and Storey, 2013; Moore and Carter-Hicks, 2014). A few studies, such as one by Baskerville and Goldblatt that investigated teachers' transition from personal indifference to critical friendship, characterize the nature of the critical friendship, but few studies have explored the benefits of teacher interaction once critical friendship is evidenced and few studies investigate the domain-specific nature of interactions between critical friends. Schuck and Russell posit that critical friendship is essential for the critique and restructuring of existing practices by providing support and constructive feedback (Schuck and Russell, 2005). We concur and posit that critical friendship is essential for chemistry teachers to share in-depth, discipline-specific beliefs and practices with their peers. Additionally, characterizing beliefs and practices is difficult, and providing critical friends a forum to question, think, and share what practices they believe to be the "best practices" could elicit more contextualized revelations about individual beliefs and practices (Dunne and Honts, 1998).

Teacher Discourse

Teacher-teacher interactions and teacher-facilitator interactions are the most influential factors for changing participants' views and teaching practices (Graham, 2007; Akerson *et al.*, 2009; Vangrieken *et al.*, 2017). Studies regarding teacher discourse often investigate the impact of discourse during professional development on instruction or the quality of teacher discourse during instruction (O'Connor and Michaels, 1993; Moje, 1997; Childs and McNicholl, 2007). There is limited research regarding the nature of teacher-teacher discourse during professional development, particularly chemistry-specific investigations of teacher-teacher discourse during professional development. Such research would be valuable because of the depth of contextualized data generated regarding teacher practices and beliefs among critical friends.

During long-term professional development, teachers have the opportunity to build trust and comradery with critical friends. Trust and experience can lead to new learning opportunities between critical friends as they engage in critical discussions of beliefs and practices (Dunne and Honts, 1998; Curry, 2008; Attard, 2012). Critical friends' discussions may include conflict and differing perspectives that can stretch beyond their comfort zones, but critical friends are generally able to resolve conflict by collegial means (Snow-Gerono, 2005; Curry, 2008; Attard, 2012). The familiarity of other critical friends' classroom environments and practices promotes discourse that moves beyond superficial topics to underlying theory that can improve pedagogy. By creating an environment of trust and collective growth, teachers can discuss their chemistry students and chemistry-specific instructional practice in challenging ways. Pairing the real-life experiences of teachers with research and practical inquiry is the most productive integration of educational theory and practice (Richardson, 1996). By engaging critical friends in discussion about chemistry assessment, the nuances of teacher-teacher discourse can be explored to uncover teachers' chemistry assessment beliefs and practices.

THEORETICAL FRAMEWORKS

Logic of Inquiry

Much can be learned about teachers' assessment beliefs and how teachers support one another to facilitate changes to assessment practices by bringing teachers together to work on improving assessment practices. However, an important consideration for professional development is whether or not the teachers working together to improve assessment practices are critical friends. Since critical friends are often more willing to share analytical and discriminating ideas and feedback, discourse between critical friends while they engage in the process of proposing or revising theories about assessment can surface the nature of teacher beliefs and practices. However, discourse between critical friends can be complex. To evidence critical friendship, fundamental aspects of discourse, such as contributions by individuals, interactions between individuals, and how discourse is constructed, need to be extracted from records of the complex discourse. The logic of inquiry framework synthesized by Kaartinen and Kumpulainen structures the dissection of discourse processes and explanation-building using the four parallel analysis categories as shown in Table 1 (Barnes and Todd, 1995; Gee and Green, 1998; Kumpulainen and Mutanen, 1999; Kaartinen and Kumpulainen, 2002). The disaggregation of the many facets of discourse allows for an in-depth characterization of teacher statements so that evidence of critical friendship can be uncovered from records of teacher discourse.

Table 1. Logic of Inquiry Framework

Category	Description (Category investigates…)
Discourse Moves	the nature of the discursive exchanges
Logical Processes	the logical relationship between contributions to discourse as a social interaction
Nature of Explanation	how the contribution to discourse was constructed
Cognitive Strategies	how the individual approached the contribution they made to discourse

Discourse moves shed light on the participatory roles of an individual during social interaction by characterizing their conversational turn (Kaartinen and Kumpulainen, 2002). When engaging in social interaction, an individual may take a conversation turn by making a discourse move, such as proposing an idea or continuing conversation based on the idea of a peer. For example, a teacher may initiate a conversation about how best to assess a concept a teacher may ask their peers how they assess the concept. The act of initiating a conversation can be viewed as a single conversational turn.

Logical processes are related to the relationship between conversational turns and how social interaction contributes to collective understanding (Kaartinen and Kumpulainen, 2002). Essentially, logical processes characterize what occurred during a conversational turn. For example, a person who initiates a conversation can do so in a number of ways, such as asking a question or proposing an idea.

The nature of explanation characterizes how the discourse during a conversational turn was constructed (Kaartinen and Kumpulainen, 2002). For example, if a person were to initiate a conversation by asking a question they may do so formally (such as when a student asks their teacher) or informally (such as when a student asks a friend). Although the nature of explanation can change as an individual interacts with other individuals, it can also differ based on the content an individual is discussing or the type of conversational turn they are engaging in. For example, a student may ask a question to their friend in a more formal tone to ensure precision of language but respond to questions from their friend informally.

Cognitive strategies are related to how an individual frames their contribution to social interaction (Kaartinen and Kumpulainen, 2002). Similar to logical processes, cognitive strategies seek to characterize what is occurring during a conversational turn. However, whereas logical processes characterize the relationship between conversational turns, cognitive strategies characterize the relationship of the individual taking the turn to the conversational turn. Cognitive strategies essentially describe how they individual engages in social interaction (Kaartinen and Kumpulainen, 2002). For example, if a student initiated a conversation by asking a question, they could do so by relating to an everyday experience or asking if a concept from class applies in a different situation. Further description of how the Logic of Inquiry framework was used in the study may be found in the Methods section.

RESEARCH QUESTIONS

A clear gap in the literature exists regarding the nuances of teacher-teacher discourse. Although interactions with critical friends has been found to improve instructional practices, little is known about the nature of interactions between critical friends. The structure of teacher-teacher discourse, as well as the nature of teacher contributions to discourse, is complex. A representation of individual teacher contributions, interactions between individuals, and the ways interactions are constructed can be used to organize complex teacher discourse to investigate for the presence of critical friendships. Additionally, a lack of content-specific materials about assessment practices is available for high school chemistry teachers to develop assessments, interpret results, and use results to guide instruction. Thus, the goal of this study is to investigate not only assessment practices valued by United States high school chemistry teachers, but also the structure and chemistry-specific nature of discourse between high school chemistry teachers to deepen our understanding of chemistry teacher assessment practices. The following research questions frame the study:

1) To what extent does discourse from a facilitated discussion about assessment best practices reveal characteristics of critical friendships between high school chemistry teachers?

2a) What best practices about assessment are revealed through a facilitated discussion about best practices for chemistry assessments among high school chemistry teachers?

2b) How do teachers' reports of best practices revealed through a facilitated discussion about best practices for chemistry assessments align to those cited in the literature?

3) What are the chemistry-specific features of the best assessment practices revealed during a facilitated discussion about best practices for chemistry assessments among high school chemistry teachers?

METHODS

This research was approved by the university's Institutional Review Board as an investigation into the alignment between high school chemistry teachers' practices and beliefs about assessment. All methods were in compliance with the university's policies on ethics. Informed consent was obtained for all participants prior to participation. The "laboratory" for this investigation took the form of a long-term professional development program for high school chemistry teachers offered during the spring 2018 semester. The professional development consisted of four 5-hour sessions held over five months, each emphasizing a different component of the process of data-driven inquiry (DDI): Day 1 (goals), Day 2 (evidence), Day 3 (conclusions), and Day 4 (actions) (Harshman and Yezierski, 2015). The professional development employed collaborative action research, which is a method of facilitating productive peer interaction and structuring long-term professional development that promotes meaningful instructional change. In this method, time and support are provided to teachers and researchers as they collaborate to find solutions to practical problems teachers experience in the classroom (Lieberman, 1986; Clift *et al.*, 1990). Teachers in the study described herein are working collaboratively to improve their practices generating assessments and interpreting assessment results. The research questions for the study described herein are answered by focusing on discourse among members of the teacher cohort during the first day of the professional development. The main goal for the first day was to have the teachers develop and refine a set of best practices for assessment that they could apply to their work throughout the rest of the professional development.

Sample

Teachers who had previously participated in the Target Inquiry at Miami University (TIMU) professional development were invited to participate in this extension to the original TIMU professional development (Herrington and Yezierski, 2014). As a result, five high school chemistry teachers participated in the extension professional development who had already completed a significant amount of professional development in improving the quality and frequency of inquiry instruction in their classrooms. All five chemistry teachers have been teaching for at least 10 years and currently teach in Ohio (US) public high schools (grades 9-12). Demographic information for the teachers is provided in Table 2.

Table 2. Teacher Demographic Information

Participant	Years of Experience	School Size (# Students)
Anne	11	400
Ashton	25	1600
Celine	18	650
Claude	16	450
Emmerson	10	750

Data Collection

All five teachers participated in the discussion of best practices. There was no time limit placed on the discussion, but discussion ended after about one hour. Teachers sat together at tables arranged in a U-shape, with two video cameras placed so that each would capture physical interactions from half of the group of teachers. An audio recorder was placed in the middle of the teachers to capture verbal interactions.

Both authors facilitated the best practices discussion by asking the teachers to generate assessment practices using the following prompts:

1) What are the characteristics of a quality assessment?
2) How do you know the assessment is high quality?
3) How do you know if the assessment met the goals of the lesson?

While one facilitator mediated the best practices discussion with relevant follow-up questions, another worked as a scribe to record proposed assessment practices by writing them on the board. The scribe consistently conducted member checking with the teachers by asking clarification questions to ensure that written assessment practices accurately reflected the teachers' ideas. Discussion continued on the first question until the teachers stated they had nothing left to add to their list of practices, at which point the second question was asked. This iterative process continued for about an hour, until the teachers had generated three lists that summarized what they believed to be essential assessment practices for inquiry activities. After the professional development day concluded, the practices were summarized into a set of guidelines for constructing assessments, evaluating assessment quality, and evaluating alignment between the assessment and lesson goals. The guidelines were given back to the teachers during the second professional development day to confirm that the guidelines aligned with the teachers' original intents. Returning the best practices to the teachers to confirm that the identified practices align to the teachers' original intents serves as evidence for trustworthiness through member checking (Maxwell, 2013).

Analysis of Teacher Discourse

Audio data from the discussion were transcribed verbatim. Video data from the discussion supplemented the transcripts by allowing researchers to identify and note who was talking to as well as the presence of teachers' nonverbal forms of participation (e.g., nodding). When the speaker changed during a line in the transcript, a new line was started for the new speaker. Transcripts from the discussion were subjected to three parallel analyses modified from the logic of inquiry framework (Table 1) (Kaartinen and Kumpulainen, 2002). For each analysis, the transcript (including non-verbal forms of discourse) was deductively coded for types of *discourse moves*, *logical processes*, and *nature of explanation*. Since the analysis categories of the logic of inquiry framework are parallel, but not corequisite, the analytical decision to exclude the *cognitive strategies* category does not impact the ability to investigate for themes using the other three components of the framework. Dedoose software was used to manage data and visualize patterns among codes (Dedoose Version 8.0.35, 2018). Interrater analysis was conducted by having two researchers separately code portions (about 10% at a time) of the transcripts. This process resulted in an interrater agreement of 75-82% for all three categories. Disagreements in code application were discussed, resulting in minor modifications to code descriptions and reapplication of the codes throughout the data. The complete codebook is provided in Appendix 1.

The first research question seeks evidence of critical friendship among secondary chemistry teachers. To address this question, several methods of tabulating and organizing frequencies of code applications were considered; however, few offered the opportunity to characterize code occurrences over time. Summaries of overall code applications and cross-tabulations of code applications did not meet the standard for the first research question; however, these were essential benchmarks for addressing the first research question. To simultaneously examine all coding schemes and to represent the temporal relationship among code occurrences, a new representation was developed called a discourse map. Line numbers from the transcript were used to track code location on the discourse map. Code applications were organized using a spreadsheet with each column representing a line of the transcript and each row representing a participant.

The second research question investigates what chemistry assessment best practices were revealed during a facilitated discussion of best practices and how such practices align to high-quality assessment practices cited in the literature. Best practices generated by the teachers were compared to published papers that communicate what teachers should consider when generating assessments, evaluating assessment quality, and determining alignment between the assessment and the lesson goals.

The third research question investigates the chemistry-specific features of best assessment practices revealed during the discussion. To address this question, the discourse contributing to the development of each assessment practice was analyzed for chemistry-specific considerations the teachers discussed when generating their set of best practices.

RESULTS AND DISCUSSION

Investigation of Discourse for Evidence of Critical Friendship

Overall discourse moves. To address the first research question, discourse patterns across the entire intervention were identified by examining frequencies of code applications. The *discourse moves* coding scheme characterizes each conversational move and turn taken by the teachers. Table 3 shows the overall number of code applications for the *discourse moves* for each teacher as well as the total code applications. Inspection of Table 3 reveals that the "agree/disagree" code was applied most frequently. Although the code identifies instances of agreement and disagreement by teachers, the transcripts show that there were only a few applications of direct disagreement, with an overwhelming majority of applications for this being "agree." The only instances of direct disagreement by the teachers consisted of minor pedagogical decisions around the assessment environment. For example, teachers disagreed about when to implement pre-tests. Some of the teachers did not agree with the idea of giving the students a quiz before being introduced to the content while others thought assessing content before introducing the students to it provided a good baseline of student prior knowledge. All other disagreements stemmed from the time teachers spend on a single chemistry topic. The lack of direct disagreement evidences the nature of critical friendship between these teachers. Prior investigations about critical friendship shows that teachers are able to resolve conflict by collegial means, finding ways to express their differing views in a more productive manner than direct disagreement (Dunne and Honts, 1998; Swaffield, 2004; Schuck and Russell, 2005; Curry, 2008; Taylor and Storey, 2013; Moore and Carter-Hicks, 2014).

Table 3 Overall Code Applications of Discourse Moves

Code	Ashton	Celine	Claude	Anne	Emmerson	Total
Initiating	38	31	6	2	0	77
Continuing	125	118	98	18	1	360
Agree/Disagree	85	56	65	59	40	305
Replying	5	4	1	1	0	11
Concluding	0	0	5	0	0	5
Referring Back	2	6	6	1	0	15
Commenting	4	10	6	9	0	29
Total	259	225	187	90	41	802

Emmerson and Anne made fewer contributions compared to their peers (41 and 90 *discourse moves*). Ashton, Celine, and Claude each more frequently engaged in discourse throughout the best practices discussion (187-259 moves). Since Emmerson and Anne did not contribute as frequently as the others, the assessment practices proposed by these teachers may not align as well to their beliefs. However, Emmerson and Anne's consistent agreement with assessment practices proposed by other teachers is evidence that the practices proposed align to their own. Previous studies about critical friends speak of a dual nature of the relationship (Dunne and Honts, 1998; Schuck and Russell, 2005; Taylor and Storey, 2013). Critical friends support the growth and development of their peers while also often recognizing discussion with critical friends as an opportunity to reflect on their own practice (Dunne and Honts, 1998; Swaffield, 2004; Schuck and Russell, 2005; Taylor and Storey, 2013). While Emmerson and Anne were not vocally contributing to discourse with their peers, they were often writing in their notebooks. Since they were often writing, Emmerson and Anne still seemed engaged.

Code co-occurrence of logical processes and nature of explanation. The *logical processes* characterize the type of interactions the teachers are engaging in, while the *nature of explanation* codes can be viewed as the way the interaction was constructed. Table 4 shows a cross-tabulation of the logical processes and nature of explanation coding schemes that identifies overall frequencies and code co-occurrences to reveal underlying features of how the teachers interacted.

Table 4. Cross Comparison of Logical Processes and Nature of Explanation

		Nature of Explanation Codes				
		Action-Oriented	Descriptive	Causal	No NoE	Total
Logical Processes Codes	Provides Reasoning	15	41	48	1	105
	State a Goal	15	2	3	0	20
	State a Result	3	20	15	0	38
	Refines	8	23	2	0	33
	Presents	20	28	14	5	67
	Evaluates	6	8	10	6	30
	Contradicts	0	5	1	1	7
	Gives Example	44	33	21	2	100
	Total	111	160	114	15	400

Logical processes codes were applied to verbal discourse to identify the relationship between discursive moves. The total counts for the *logical processes* show that a majority of teacher interactions involved either "providing reasoning" (105 applications) or "giving examples" (100 applications). The "provides reasoning" code was applied to statements of evidence or reasoning for performing a practice or including an assessment characteristic.

The *nature of explanation* coding scheme helped researchers examine the way statements were constructed. "Action-oriented" codes (111 applications) were applied to statements that encouraged an identifiable action to be performed, such as what to do when receiving confusing results from assessments (Celine: *What you have to do is reassess the goal.*). "Descriptive" codes (160 applications) were applied to statements that about the qualities of a practice. For example, when clarifying an assessment practice to the group Celine described an overall assessment goal that *…it's not about facts, it's about understanding.* "Causal" codes (114 applications) were applied to statements that stated a cause-effect implication to the implementation of a practice, such as when Claude was interpreting results from an assessment: *…like oh my goodness only 20% of my kids got this right, it must be a bad question.* "Descriptive" statements were the most frequently encountered *nature of explanation* code (160 applications), with only 15 statements receiving "no *nature of explanation*" code (abbreviated as No NoE in Table 4).

Conducting a statistical test of the frequencies of code applications is inappropriate, since the method of extracting code occurrence results in statements that vary in length. Examining code frequencies qualitatively reveals overall patterns regarding the nature of teacher discourse about assessment development. For example, much of the discourse involved providing context for teacher assessment practices (giving examples) and defending presented practices to critical friends (providing reasoning). Since teachers construct much of their discourse through examples with evidence, they may benefit from materials situated in the classroom. For example, providing a rich description of the environment that pedagogical practices were enacted so that teachers can better reason with how they would apply to their own classrooms. The high frequency of the *nature of explanation* codes illustrates the variety of support and reflection throughout discussion as teachers consider their own practices and encourage the development of their peers' practices.

Table 4 shows that teachers more frequently used "action-oriented" (15 co-occurrences) statements than "descriptive" (2 co-occurrences) or "causal" (3 co-occurrences) statements when discussing goals. Alternatively, when discussing the results of implementation teachers were more frequently using "descriptive" (20 co-occurrences) or "causal" (15 co-occurrences) statements than "action-oriented" (3 co-occurrences) statements. The contrast between the *nature of explanation* code occurrences for "stating goals" and "stating results" can possibly be explained by patterns observed in literature about the use of assessment data to guide instruction. The use of “action-oriented” statements while discussing goals is likely due to the fact that teachers attempt to design assessments to investigate measurable outcomes (Pellegrino, 2012; Sandlin *et al.*, 2015). Alternatively, literature about using assessment results to guide day-to-day instruction is often coarse-grained and does not provide specific guidance on how to use assessment results to guide instruction (Knapp *et al.*, 2005; Irons, 2008; Hamilton *et al.*, 2009; Suskie, 2009; Witte, 2012). The lack of materials about how to use assessment results could negatively influence the teachers’ use of “action-oriented” statements while discussing assessment results.

Although investigating individual (and paired) code occurrences provided insight about the overall nature of critical friendship, a different representation of the data was necessary to reveal features of critical friendship over time. A novel representation of all three coding schemes as applied to all participants, called a discourse map, was used to synchronize data among logic of inquiry coding schemes for each participant, while also representing the code occurrences temporally. A sample selection from the discourse map is shown in Figure 2. The complete discourse map for the best practices discussion is found in the Appendix 2.

Each cell of the discourse map may contain a symbol, color, and/or shading gradient to indicate code application(s) in that line in the transcript. If a cell is empty, the person in that row was not participating in discourse for that line of the transcript. *Discourse moves* codes are indicated by symbols, shown in Figure 1. *Logical processes* codes are indicated by colors, shown in Figure 1. *Nature of explanation* codes are indicated by shading gradients, shown in Figure 1.

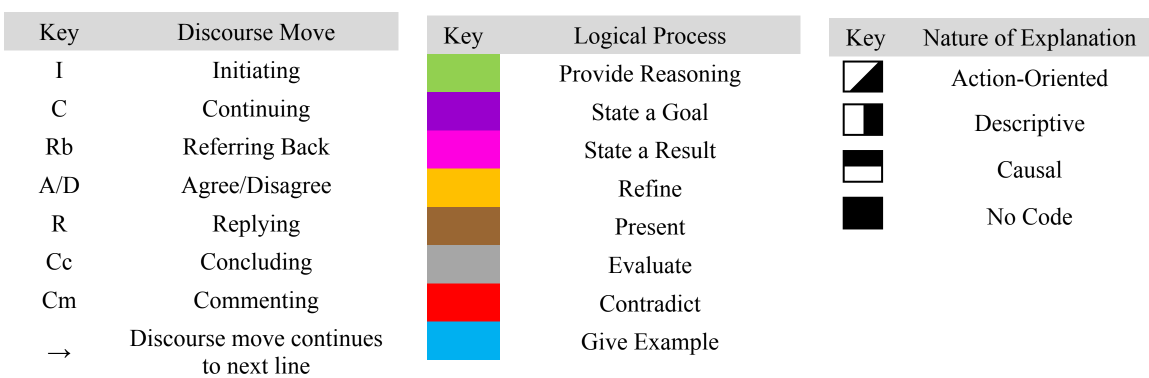


Figure 1. Indicators for the three coding schemes on the discourse map.

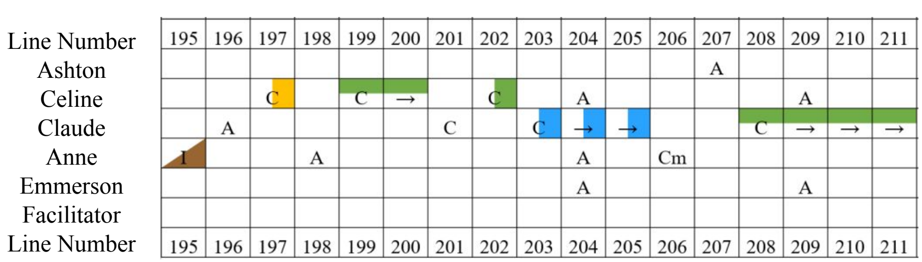


Figure 2. Segment of the best practices discussion discourse map.

Examination of the discourse map reveals that the codes for "agree/disagree" and "continuing" occur more frequently than others. The "agree/disagree" code was applied when teachers verbally or nonverbally indicated that they agreed or disagreed with another teacher or facilitator. The only non-verbal forms of agreement coded were when a teacher nodded in direct response to a statement made by another teacher. No nonverbal forms of disagreement were detected. In instances when a teacher would nod and provide a verbal remark only one code of "agreement" was applied. The "continuing" code was applied when one teacher extended discourse about the same idea presented by the previous teacher. The high frequency of code applications for the "agree/disagree" and "continuing" codes speak to the relationship among the individuals in the best practices discussion. All teachers have participated in multiple years of professional development together and seemed willing to disagree with their peers or continue the conversation with a differing perspective. For example, some teachers felt they had time to walk question-by-question with students to provide feedback about lab experiments, while others communicated that their school environment necessitated that they find less time-consuming ways to provide feedback to students. Even through disagreement, the teachers were supportive of each other's practices, recognizing the environmental differences that allow for pedagogical differences. Literature on the role of critical friends states that critical friends serve to support and critique each other (Schuck and Russell, 2005; Curry, 2008; Taylor and Storey, 2013; Moore and Carter-Hicks, 2014; Fletcher *et al.*, 2016). Schuck and Russell reflected that the value of critical friendships stems from encouraging the reconsideration of practices and creating space and opportunity to nourish that reconsideration (Schuck and Russell, 2005). The high frequency of "agreement/ disagreement" and "continuation" between teachers during the best practices discussion is likely an illustration of the support and ongoing consideration of the beliefs and practices of their peers.

The discourse map shows that a majority of teacher interactions involved either the "provide reasoning" or "giving examples" *logical processes*. The "provide reasoning" code was applied to statements of evidence or reasoning for performing a practice or including an assessment characteristic. Teachers provided reasoning in various contexts, such as crafting assessments (Celine: *The reason it worked is because it forced me to pay attention to what they wrote.*), interpreting assessment results (Claude: *Like oh my goodness only 20% of my kids got this right…*), and the benefits of critical friends (Ashton: *…because I gotta be honest with you, for me writing questions is painfully hard.*). "Gives example" codes were applied to actual or hypothetical examples teachers stated to illustrate an idea. Teacher examples were often very long and situated discourse in the context of the classroom. For instance, when Claude was talking about goals for his test questions, he stated:

> I have a test question where I give data uhm and the data is of four alkaline earth metal salts and what you observe when it's in water. So, one is dissolved and one is a little cloudy and one is thick precipitate. A thick white mix and then I give I think it was three and then I give a fourth and I say with question marks. and I say *What are you gonna see?* Alright, and they have to make a prediction based on periodic trend. And then, I ask a question and I've been fighting with this question for years. I wanna get to the point whether of 'is this a trend of the atom or is this a trend of the ion formed by that metal?'

Critical friends often require context to adequately discuss each other's aims and perceptions regarding their practice (Swaffield, 2004; Schuck and Russell, 2005). Once critical friends are oriented to their peer's perspective, they encourage each other to think deeper about their practice, giving opportunity for reasoning and examples to arise that emerge through reflection, thus explaining the frequency of lengthy examples.

One of the main benefits of the discourse map is the ability to investigate for discourse patterns over time. For example, examination of the discourse map reveals that teachers often transitioned between several logical processes and/or natures of explanation when elaborating on a single *discourse move*. In Figure 3, Ashton is contradicting a statement made by one of the other teachers (indicated by the red shading of cells 146-149). Ashton uses a descriptive explanation for contradiction (shown as vertical shade gradient). Potentially as an alternative, Ashton proposes (brown shading) an assessment characteristic or practice at the conclusion of his remarks. In Figure 4, Claude is continuing the discourse about an assessment practice proposed by another teacher with an example (blue shading). Claude begins with a descriptive example (vertical gradient) which slowly transitions to a causal explanation (horizontal gradient). Throughout the course of his contribution, Claude's statement even contains a descriptive goal (vertical, purple shading) embedded within his example.

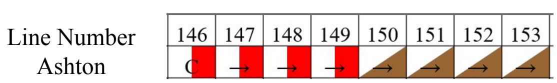


Figure 3. Ashton changing logical processes and nature of explanation over a single discourse move.

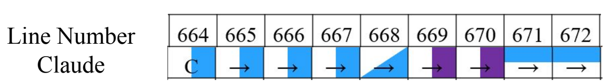


Figure 4. Claude changing logical processes and nature of explanation over a single discourse move.

Segments of complex discourse like those shown in Figures 3 and 4 are able to be identified using the discourse map. Critical friend groups offer space and opportunity to reconsider the aims and practices behind day-to-day instruction that is often a rarity for teachers (Dunne and Honts, 1998; Swaffield, 2004; Schuck and Russell, 2005; Fletcher *et al.*, 2016). Similar to the complexities of their day-to-day instruction, the knowledge and reasoning guiding their practice is likely equally complex. The value of critical friendships is rooted in the context of day-to-day instruction because the beliefs guiding assessment practices are situated with other experiences or knowledge (Dunne and Honts, 1998; Swaffield, 2004; Schuck and

Russell, 2005). For instance, the discourse map shows that classroom examples are provided within every segment of discourse and are rarely shorter than three lines of transcript. The frequent use of examples is representative of the importance of context as teachers consider how to change or improve practice.

Overall, the discourse map represents the discourse in a manner that allows for the characterization of discourse trends that are consistent with those of critical friends (Dunne and Honts, 1998; Swaffield, 2004; Curry, 2008; Baskerville and Goldblatt, 2009; Taylor and Storey, 2013). The number of years participating in professional development together serves as additional evidence that these teachers are critical friends. The critical friendship between the teachers participating in the best practices discussion is embodied in the depth and vulnerability of the ideas and practices shared during the discussion (Dunne and Honts, 1998; Schuck and Russell, 2005). The key findings in the study thus far are summarized in Table 5.

Table 5. Key Findings and Data Sources for Research Question 1

Research Question	Data Sources	Key Findings
1) To what extent does discourse from a facilitated discussion about assessment best practices reveal characteristics of critical friendships between high school chemistry teachers?	Tables 3 & 4, Discourse Map	• Teachers in this investigation have discourse patterns aligned with critical friends as described in the literature. • Discourse map was more useful than overall code frequencies for revealing discourse characteristics.

Best Practices for Chemistry Assessments

The second research question asks what best practices for assessment are generated by the teachers and how those practices align with assessment literature. Table 6 contains the best practices generated and approved by the teachers during the professional development. Literature best practices were viewed as evidence-based practices that are shown to improve the quality of assessment design or reliability and validity of assessment results.

Table 6. Teacher-Generated Best Practices for Assessment

Assessment Component	Best Practice
Creation Considerations	a. Assessment Items must align to learning goal b. State the learning objective in terms that explicitly state what the student should explain/do. c. The learning objective should be clear and provided to the students. d. The assessment should address different levels of knowledge within one instrument. e. Identify the appropriate "level" for demonstrating competency of the learning objective i. The level should consider both Bloom's Taxonomy and Johnstone's Triangle f. Identify the type of data needed from students to give good evidence of understanding g. Focus on conceptual understanding, not just recall of facts h. Assessment items should consider student prior knowledge
Evaluation of Quality	a. Clarity i. Students should understand the point of the item ii. Item should not be "tricky" and should actually get at what you want iii. Not a reading comprehension item, but a chemistry item 1. Present no more than 3 ideas per sentence 2. At least one idea connects to next sentence 3. Consider removing extraneous information iv. Strong links to what has been emphasized during instruction b. Ease of grading i. Needs to be relatively easy to grade c. Consider if the item detects common misconceptions. i. Use both formative and summative to inform teaching and detect misconceptions early.
Evaluation of Meeting Goals	a. In thinking about a collection of items you should see a diversity of levels i. The level should consider both Bloom's Taxonomy and Johnstone's Triangle b. When comparing student performance on items to goals: i. Need to triangulate evidence from multiple items to draw conclusions regarding instruction i. One sketchy item may be a problematic item ii. Verbal student feedback can inform how item is being interpreted by students iii. Consider the distribution of scores on items and who gets the items correct c. Expert validation i. Get insights and evaluation of items from other teachers

The generated best practices for assessment contain a mix of simple-to-apply rules (e.g., limiting the number of ideas per sentence to a maximum of three and removing extraneous information) as well as broad guidelines without specific actions to follow (e.g., the imperative to align the assessment items to a learning goal). Whether the best practice stated specific actions to follow or not, this set of guidelines aligned to practices cited as high-quality considerations in assessment literature.

To discuss the alignment to assessment literature, the teacher-generated best practices are summarized as three themes:

1) The importance of assessing for conceptual understanding
2) Including clear, direct assessment item stems and learning goals
3) Assessing learning goals multiple times over a variety of conceptual and representational levels

Teachers' guidelines stating the importance of assessing for conceptual understanding align with several investigations of high-quality assessment practices (Black and Wiliam, 1998; National Research Council, 1999, 2001, 2014; Bell and Cowie, 2001; Stiggins, 2001; Gibbs and Simpson, 2004; Lyon, 2011). The National Research Council states that providing students with assessment opportunities to reveal what they have learned and understood to themselves, their peers, and instructors is more beneficial to achievement than simple recall (National Research Council, 2014). Embedded within the literature is the acknowledgement that teachers are tempted to include assessment items that assess recall (or statements of facts) over items that require application and conceptual knowledge. Recall items can be simple to generate, obtain from assessment sources, and grade, but often do not intellectually engage the student enough to collect adequate data about the students' true understanding (Stiggins, 2001; Towns, 2014). Advice from the literature says that writing high-quality assessment items can be difficult, and teachers should take advantage of evidence-based resources as well as their peers to generate high-quality, conceptual items (Dunne and Honts, 1998; Towns, 2014).

When discussion about best practices for assessment began, the teachers quickly leapt to the importance of clarity and transparency of the learning goals guiding assessment. Generating clear learning goals, including them in the assessment process, and articulating them to the students are very prominent practices in the literature (Stiggins, 1988, 2001; Black and Wiliam, 1998; Bell and Cowie, 2001; Gibbs and Simpson, 2004; Sato *et al.*, 2008; Hamilton *et al.*, 2009). Learning goals are an expectation for well-crafted learning activities, and a significant body of literature exists depicting the importance of crafting clear, concise, and measurable goals used to link assessment and instruction. The teachers in this study were very aware of the benefits of well-crafted and usable learning goals as evidence by the guidelines they generated.

The third theme from the teacher-generated best practices involved the use of multiple measures of student knowledge to improve the precision of claims made about student understanding. Similar to the other two themes, the idea of assessing learning goals multiple times through a variety of conceptual levels aligns with literature for desirable assessment practices (Black and Wiliam, 1998; Bell and Cowie, 2001; Stiggins, 2001; Gibbs and Simpson, 2004; Hamilton *et al.*, 2009; Lyon, 2011; National Research Council, 2014; Towns, 2014). The use of multiple measures is recommended from several sources, including the National Research Council which calls the use of multiple, related questions with a variety of tasks to assess a single learning goal "multicomponent tasks"(National Research Council, 2014). Studies promoting the use of multiple measures state that the additional data allow the teacher and student to gain insight to student-specific challenges as well as knowledge gained (Black and Wiliam, 1998; Bell and Cowie, 2001; National Research Council, 2001; Stiggins, 2001).

Although teachers recognize the importance of using multiple measures to investigate student competency, they do not always know what to do with the data they obtain, shown as Claude discusses triangulating evidence.

Claude:	And so, if I can get away from that question for a second because something's not working, and I've tweaked my instruction.
Celine:	Sure. Right, right.
Claude:	To try to emphasize that, and it's not working. And so, then I'm like in my brain. If 70% of my kids are missing a question. If 70% of my kids aren't getting the question right. And this may go all the way back to college, right? A 70% a 70 on a chemistry test in college was a low B or a C right?
Anne:	That was an A!
Claude:	And uhm. If uhm you get the whole bell curve whatever. If 70% of my kids aren't getting it right. Somethings wrong. Somethings wrong with the question something's wrong with the instruction. Something's wrong with something. Uhm. My question is, you know when only 20% are getting it right, like okay, not only is something wrong, something's really wrong. Right? uhm and so what do I do with that? Is that a point where I say, uhm you know, this is a losing battle I can't win it no matter what I do? Is this the point where I say I've gotta change and spend half a week just on the difference between the symbolic?

Claude expresses struggles with understanding how to use results from assessments to guide his instruction. In his remarks, Claude states that he can clearly tell that "something's really wrong" when his students underperform on an assessment but lacks the certainty to identify what that "something" is. Claude poses many questions that illustrate his frustration, leading to the idea that he is "fighting a losing battle." If 70% of students get an item incorrect, Claude believes that an instructional intervention is needed, but he cannot interpret the 70% value to enact class-level decisions. Previous research showed that when faced with poor student performance, teachers stated they would reteach or re-cover the content, valuing repetition as a means of closing knowledge gaps (Harshman and Yezierski, 2015). Here, Claude is drawing the conclusion that "something" did not work. He believes that a different instructional method is required and that simply repeating the learning activity would possibly result in a similar student performance. Claude's inability to rely on student data to guide instruction was a gap in assessment practice shared by all teachers in the best practices discussion. A recent review found that using assessment results to guide day-to-day instruction is not addressed in current literature (Harshman and Yezierski, 2017). Often sources state that teachers *should* use assessment results to guide instruction but lack the guidance for *how* teachers could use data from assessment results to guide instruction (Heady, 2000; Martone and Sireci, 2009; Clinchot *et al.*, 2017). Other sources state that teachers need content-specific professional development opportunities to learn how to use data in their classrooms (Smith, 2013; Herrington and Yezierski, 2014; Banilower *et al.*, 2018). This gap in the literature aligns with the gap in best assessment practices exhibited by the teachers in the best practices discussion. Table 7 summarizes teachers' best assessment practices uncovered during the teacher discussion.

Table 7. Key Findings and Data Sources for Research Question 2

Research Question	Data Source	Key Findings
2a) What best practices about assessment are revealed through a facilitated discussion about best practices for chemistry assessments among high school chemistry teachers? 2b) How do teachers' reports of best practices revealed through a facilitated discussion about best practices for chemistry assessments align to those cited in the literature?	Table 6. Teacher-Generated Best Practices for Assessment	• Teacher-generated practices contain mix of broad guidelines and simple-to-apply rules. • Teacher-generated practices organize into themes aligned with relevant assessment literature. • Discourse of assessment practices revealed gaps in teacher practices not addressed by literature

Chemistry-Specific Features of Best Practices About Assessments

Any teacher would likely be able to generate a set of practices for assessment. However, the critical friendship that has developed between these experienced teachers over years of high-quality professional development has afforded a forum where all were comfortable to share ideas with greater depth and vulnerability. The comfort shared by the teachers in this forum led to content-specific discussions that acknowledged the needs of each individual's learning environment, as evidenced by Claude's sharing about his struggles to interpret assessment data.

Claude	So here's the issue, right? I have a test question where I give data and the data is of four alkaline earth metal salts. And what you observe when it's in water. So, one is dissolved and one is a little cloudy and one is thick precipitate. A thick white mix and then I give, I think it was three, and then I give a fourth and I say with question marks. And I say, "What are you gonna see?" Alright, and they have to make a prediction based on periodic trend. And then I ask a question and I've been fighting with this question for years. I wanna get to the point whether of is this a trend of the atom or is this a trend of the ion formed by that metal? Right? and in such in such expression that's what I ask. Okay? because quite honestly for 95% percent of my kids can make the prediction because I've even got it arranged as they are on the periodic table. Uhm, but the last question I'm trying to get at: are these the metals or are these the ions? Uhm and tell me how you know. And
Anne	You're talking the trend is the trend for the metal or for when it's an ion?
Claude	Is the trend is the are we working with magnesium metal, calcium metal, etcetera or are we working with magnesium ion, calcium ion, going down the list. Uhm and inevitably 80% of my kids tell me it's of course it's the metal. Right?
Facilitator	Even though there's no metals in the system.
Claude	Even though there's no metals there! And we've done the experiment. They've seen this. They know they're not working with the metal, but its but this is on my semester exam, whatever. Uhm and I'm like okay. So, I've taught it, we've talked about it, they've worked with it, and they still are missing it. Why? is it because the question in some way is confusing? Is it because I didn't do a good? Is it because, you know why are they missing this question so badly? Uhm and I just don't know, and I don't know how to answer that.

Claude raises uncertainty how to interpret assessment results, as was evidenced before, but Claude shares chemistry-specific details with his critical friends. Claude recognizes that he can share his struggles teaching and assessing periodic trends, whether it stems from his instruction, assessment, or students. Non-chemists would likely need further information about periodic trends to engage in a discussion of pedagogic struggles involving periodic trends, and/or related content. Claude's quotation exemplifies how chemistry teachers could benefit by having discipline-specific critical friends who would understand their chemistry-specific assessment and instructional problems.

Although domain-general discourse was far more common, chemistry-specific conversations occurred during the best practices discussion. For example, when discussing the importance of assessing for conceptual understanding, Claude shared,

Anne	Avoiding over-emphasis of facts but that's, I think that could be added.
Claude	Yeah
Celine	Yeah, it's not about facts it's about understanding.
Anne	Yeah.
Celine	And, sometimes we get caught up in the trap of it's an easy question, everybody should be able to answer it
Claude	Yeah, I remember the first
Celine	And if it's a bad question, that's meaningless.
Claude	The first test I got, I ever looked at when I started teaching about atomic structure. Uhm it was 90% counting protons, neutrons, electrons, right? Uhm. Why would you do that?
Ashton	Right?
Claude	Why would you do that? You test that about, you can do that in about 4 or 5 questions and then let's get to actually what the, what we know about the atom as opposed to just that because it's easy to do next. I mean that's the thing, its super easy to test.
Anne	9, 9, 10
Claude	Yes. Yeah, and its super easy to write the question.
Anne	And super boring to grade.
Celine	That's so true
Claude	There's hundreds of examples of question you could write without batting an eye. It's just an easy to write and then it's an easy test to grade, but have you tested that the kids understand anything?

While elaborating on Anne's suggestion to avoid an over-emphasis of assessing factual recall, Claude reflects on how early in his career he fell into the trap of making assessments that were easy to write and grade but did not assess students' conceptual understanding of chemistry topics. The best practice of assessing for conceptual understanding could be domain-general, however this group of critical friends generated chemistry-specific considerations for the nature and depth of chemistry assessments.

As stated above, the importance of generating clear and direct learning goals is common throughout educational literature. In this discussion of best practices, the topic of learning goals arose several times. Teachers spoke of the importance of being direct, being specific, and potentially giving the learning goals to the students. In one instance from Celine, she discusses how she uses the learning goals to help students recognize knowledge gaps that become evident as students attempt to apply common chemistry symbols.

Celine When you asked, "How do you make the student aware of the goal?" Okay, so here's an example. We're doing a simple lab. We're getting into reactions again. And uhh reactant sodium bicarbonate. You know one of the things they had to do was write down the equation and translate every symbol from the equation. The arrow, what does it mean? The *aq*, what does it mean? Then, [the students] had to answer a question after that. This is a pre-lab, and I made them write down the three objectives of the lesson. And one of them was "I can translate the symbols into words for this reaction." That was. They wrote it. So, you know *aq* means aqueous right? And the kids were like "means dissolved in water" and they have that answered in just a little chart. Then they have the question, "Looking at the reaction above, what do you predict the reactants will look like with your eyes?" Okay, so we have aqueous baking soda and aqueous hydrochloric acid and they all can tell me the definition of'em in the table, *aq* means dissolved in water. 75% answered the question in the pre-lab, predict what this will look like with your eyes in the macroscopic, baking soda will be a powder. So, when the kids came up to my desk and I had to stamp it for the pre-lab. I clearly with every kid who got it wrong said, "What's the goal?" and I did this and it was painfully long and awful. This is the goal and you see where you wrote *aq*, I said, "What does that mean?" and they said it means it's dissolved in water. And I said, "So, this baking soda and this is water. What does it mean?" I said you said powder and I looked at the students and I said, "What I'm telling you and you can tell me what it means, but you can't translate it and apply it to my lab. You haven't met the goal."

After presenting the importance of generating and articulating clear learning goals, Celine shared a practice she engages in to help her students clarify content by using the learning goal. In the quote, Celine noted that knowing the definition of the symbols used in chemistry are not enough and that students should be assessed on their ability to apply chemistry knowledge in a laboratory environment. Celine's example led teachers to indicate that assessing learning goals using a variety of conceptual and representational levels was important. After Celine's example about how she communicates to students their progress about meeting chemistry-specific expectations within learning goals, she mentioned that the learning goals need to represent specific levels of knowledge expected of the student, as well as the representational levels to be assessed.

Celine	Well, I was also talking about a kid can puke up a definition.
Ashton	Yes!
Celine	They know *aq* means dissolved in water.
Ashton	But, to get that?
Celine	And they can do that, but to apply.
Ashton	Yes.
Celine	So, I'm talking about the application of the definition.
Ashton	Yes.
Celine	So, when I'm. So, yes it was Johnstone's related, but it was both. It was Bloom's and Johnstone's [levels]

During this dialogue between Celine and Ashton, they shared that although assessing student knowledge of definitions is important, so is assessing the students' ability to apply the definition. The teachers admitted that assessing student knowledge at a variety of representational levels was difficult, as shown by Ashton's quote,

Ashton	So, just as an aside. I had kids put a puddle of water on a piece of paper like a half page sheet. Conductivity tester, put it in. Nothing. Yeah. Right? They could figure that out. So, I said let's take some sodium carbonate and drop it to the side, push a little bit in. Okay start the conductivity on this side and slowly get closer and they see the light get brighter and brighter and brighter as it gets closer. Okay, now let's try to look at this through Johnstone's triangle. It was like pulling teeth. I mean it was like their brain exploded. Then what happened was after we did that with two items, on one side they put sodium carbonate on the other side they put copper (II) sulfate and they pushed a bit in it and I said just wait. And in the middle, a line appears. We could've spent a month on that. I mean the amount of misconceptions, the amount of you know the ions you know about this is what I saw. And then saying okay you saw this. Give me the symbol what does *aq* mean or what is that line you saw? Is that aqueous? You know is that. It was just like that simple experiment was just like *explosion noise*.

The length and detail of the teacher examples emphasizes the need for critical friends who can interact in content-specific ways. These teachers would not be able to provide such chemistry-rich examples with peers who were not knowledgeable in the content. Although literature about best practices for assessment is discussed in a domain-general manner, the teachers in this study had chemistry-specific conversations about how the practices influenced their ability to generate and use assessments in their classroom. Table 8 provides a summary of the chemistry-specific features of their best-practice discussion.

Table 8. Key Findings and Data Sources for Research Question 3

Research Question	Data Sources	Key Findings
3) What are the chemistry-specific features of the best assessment practices revealed during a facilitated discussion about best practices for chemistry assessments among high school chemistry teachers?	Teacher discourse	• Teachers value assessing for conceptual understanding of chemistry knowledge and identified chemistry-specific pitfalls teachers encounter when designing assessments. • Teachers shared possible ways to use chemistry experiences to help students better understand learning goals. • Teachers shared the importance of assessing at different levels of chemical representations and chemistry-specific experiences to enact high-quality practices that engage students with multiple levels of representations.

CONCLUSIONS

The discourse from a facilitated discussion about assessment best practices revealed characteristics of critical friendship between high school chemistry teachers, as represented in the discourse map. By representing the *discourse moves*, *logical processes*, and *nature of explanations* as they temporally occurred, the discourse map allowed for a more thorough investigation for evidence of critical friendship than characterization of individual coding results. The map revealed the complexity of teacher discourse as participants transitioned between several *logical processes* and *nature of explanations*, even throughout a single *discourse move*. Teachers made use of this forum of critical friends to discuss beliefs and practices *in situ, a*s evidenced by an example within every discourse fragment with examples often longer than three lines of transcript. The frequent use of examples gave teachers concrete instances to manipulate and perceive as they reasoned with their own assessment beliefs and practices and those of their critical friends.

Additionally, teachers often provided reasoning to accompany their examples, evidencing the importance of situating examples within the context of specific classroom environments. Reasoning provided could indicate reflection and discourse about the beliefs driving teachers' assessment practices. Although the discourse map served as a useful tool to investigate discourse at a bird's eye view, investigation of the best practices generated by the teachers required a thorough examination of the transcript.

The best practices for assessment generated in this study are provided in Table 6. Their generated best practices align to guidelines set by several evidence-based investigations of high-quality assessment practices. Three themes emerged from the teacher best practices: (1) assessing for conceptual understanding, (2) including clear and direct learning goals, and (3) assessing learning goals over a variety of conceptual levels are present throughout the literature. Themes identified represent the beliefs and practices this group of teachers deem important for generating assessments, evaluating assessment quality, and evaluating alignment between instructional and assessment materials. Teachers stated that using assessment data *should* be used to inform day-to-day instruction, however they reported practice gaps limiting their ability to implement this best practice.

Teachers revealed several chemistry-specific considerations for the generated best practices (as well as gaps in assessment practices) during the best practices discussion. Teachers discussed several seemingly domain-general practices such as developing items that assess conceptual knowledge, generating clear and direct learning goals, and assessing knowledge at a variety of representational levels in a chemistry-specific manner. The chemistry-specific discourse that these teachers engaged in evidences the benefit of having discipline-specific critical friends who are familiar with the content anchoring the discussion of beliefs and practices.

LIMITATIONS

While these findings introduce a novel representation of discourse and suggest a need to support teacher interpretation of class-level assessment data to inform instruction, this study has several limitations. Participants in this study have multiple years of teaching experience and professional development about improving the quality and frequency of inquiry instruction. The small sample size ($N = 5$) and experience of the teachers of this study could potentially hinder transferability of the findings. The teachers in this study proposed many assessment practices that relate to literature-supported practices, which could be the result of their previous experiences. The possibility exists that teachers with fewer years of experience and/or less professional development about inquiry instruction that those in our sample may not reflect the same practices.

The characteristics of this group align to characteristics of critical friends as described in the literature. The critical friendship shared by this group of high school chemistry teachers was both a limitation and an advantage to this study. Here critical friendship could limit the transferability of study findings, because findings may only apply to teachers who have developed a comfort sharing with one another over several years of interaction. However, critical friendship in this study facilitated greater elicitation and depth of data about teachers' struggles using assessment results to inform instruction, since teachers shared openly with others who were familiar with their classroom practices.

A limitation of the discourse map is that the line of text as a representational unit is an artifact the data transcription method. Some of the lines of text are sentence fragments while others are complete teacher statements. Additionally, some lines of transcript contain brief (or nonverbal) discourse contributions by from non-speaking teachers. For these reasons, readers are cautioned against interpreting the discourse map quantitatively. Rather, the discourse map should be interpreted qualitatively as a representation of high school chemistry teacher discourse about assessment beliefs and practices.

IMPLICATIONS AND FUTURE WORK

The findings presented here have important implications for teacher change and the design of materials and professional development to improve assessment practices. Professional development materials and opportunities could better support teacher change if time and support are provided for critical friendships to develop. The frequent use of examples implies the need for context, so teachers may benefit from vignettes or stories attached to guides for interpreting the results of assessments. Black and Wiliam state that teachers find the process of translating research into practice difficult (Black and Wiliam, 1998). Since the teachers often provided reasoning to justify their proposed practices, incorporating classroom context within assessment materials would improve teachers' likelihood to adopt evidence-based assessment practices.

Results also contribute to a greater understanding of the factors used by high school chemistry teachers while changing and discussing assessment practices. Teachers often rely on the support of their peers while working to improve pedagogy (Black and Wiliam, 1998; Bell and Cowie, 2001; Towndrow *et al.*, 2010). Recognizing the productive discourse patterns used by these critical friends not only helps the design of materials and professional development but also can educate teachers on more productive ways to talk about their beliefs and practices.

The discourse map serves as a tool to characterize discourse patterns over time and can be used to organize complex analyses of content. Toulmin's argumentation pattern has been previously used in chemistry to uncover student ideas embedded in argumentative discourse (Cole *et al.*, 2012). In this study the logic of inquiry framework was used to analyze discourse. Regardless of the discourse framework (such as logic of inquiry or Toulmin), the discourse map can be used to represent the structure of discourse as it occurs over time and/or to identify key discourse fragments.

Additionally, the findings presented here highlight the need for future research investigating gaps in teacher assessment practices regarding the ability to interpret class-level data to inform future instruction. Although the guidelines generated by these critical friends may serve useful for other chemistry educators, the frustration exemplified by Claude and other teachers speaks to the need for quality professional development and materials for high school chemistry teachers regarding the use of data. With more professional development opportunities and materials available to them, teachers would be able to better use assessment results to guide their instruction.

APPENDICES

Appendix 1 – Logic of Inquiry Codebook

Table 9: Discourse Moves Codes

Discourse Moves	Codes describe the conversational turn taken by the teachers/facilitators
Initiating	Teacher introduces a new idea (within the topic proposed by the presenter) not previously mentioned

Continuing	Teacher/facilitator progresses conversation within an initiated idea by elaborating with an example, qualifier, criteria, question
Referring back	Teacher/facilitator extends to a previously stated idea, links current discourse to previously mentioned idea, or asks a clarifying question about a previously stated idea
Agree/ disagree	Teacher/facilitator doesn't progress coversation, but expresses agreement with the previous statement (can be verbal or nonverbal)
Replying	Teacher/facilitator responds directly to a question from another teacher or facilitator
Concluding	Teacher/facilitator makes an ending remark for the current discussion allowing for the initiation of new ideas
Commenting	Teacher doesn't extend conversation, but adds a personal anecdote to discourse

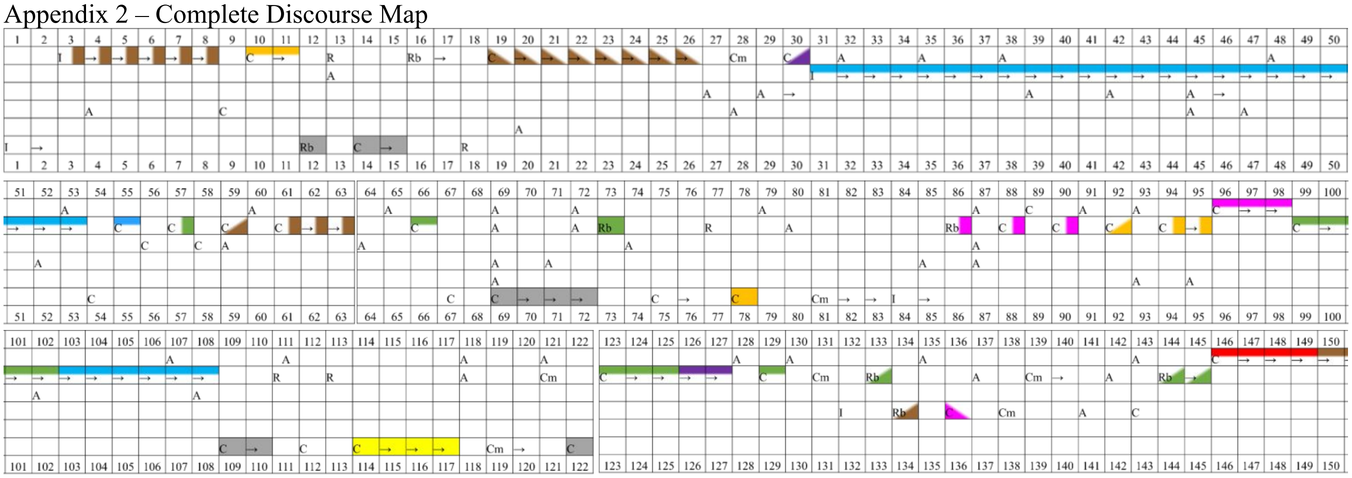
Table 10: Logical Processes Codes

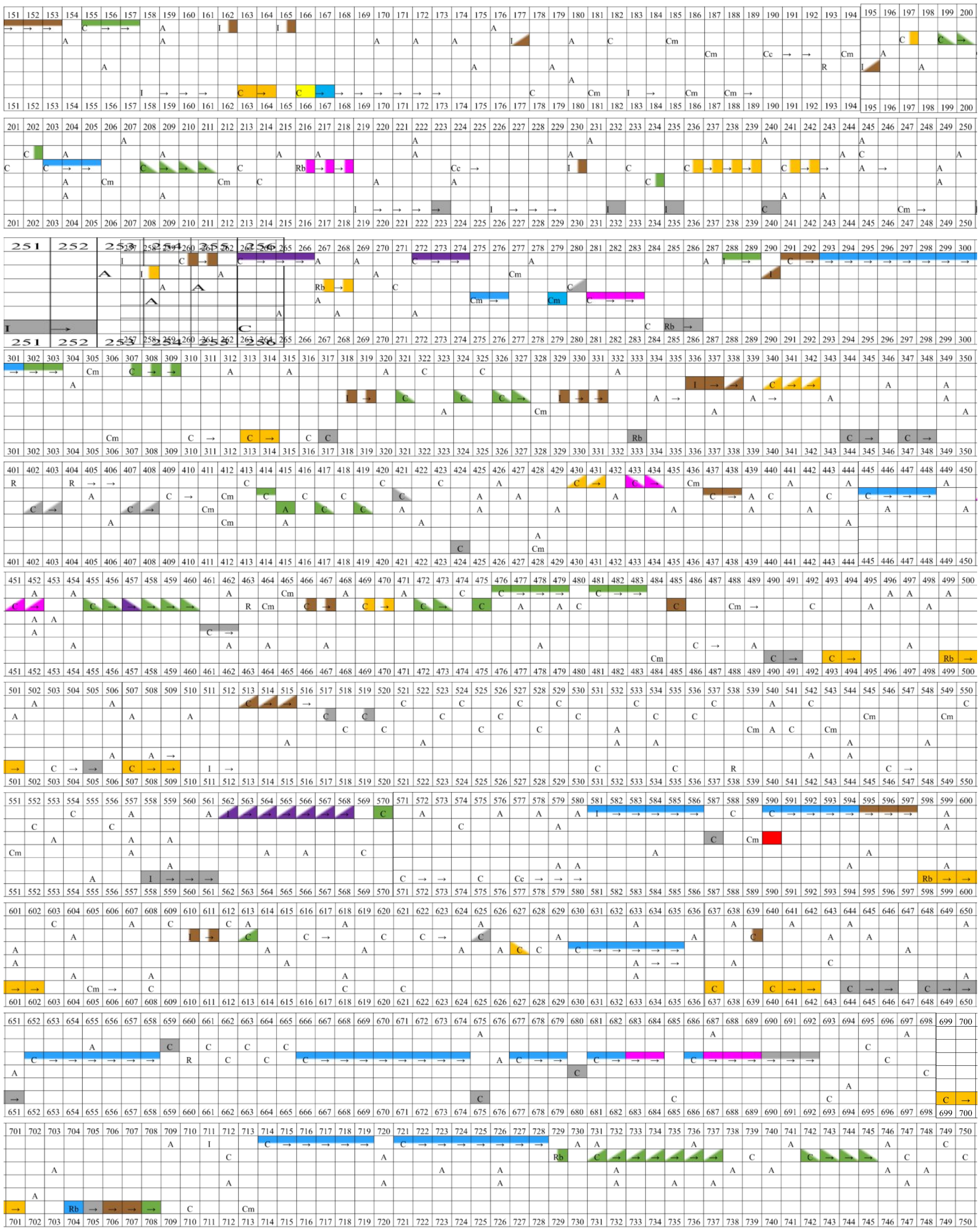
Logical Processes	Describes the structure of the teachers' statements.
Provide Reasoning	Teacher gives reasoning or evidence for including an assessment practice or characteristic or why practice or characteristic should be performed a certain way
State a Goal	Teacher states what they are attempting to accomplish when performing an assessment practice or including an assessment characteristic
State a Result	Teacher gives a possible outcome or student results from the implementation of an assessment practice/ characteristic (can be hypothetical)
Refines	Teacher refines assessment technique/assessment characteristic/pedagogical action already stated by proposing a qualifying statement or a restriction OR (connects/defines the boundary between) the assessment technique/assessment characteristic/pedagogical action to a previously stated one
Presents	Teacher provides a possible assessment technique/assessment characteristic/pedagogical action OR asks the group a question about (the importance of/if) an assessment technique/assessment characteristic/pedagogical action should be included in the best practices.
Evaluates	Teacher/facilitator questions or inquires about an assessment technique/assessment characteristic/pedagogical action made by another to increase depth of explanation, clarify details, or set a boundary from another goal
Contradicts	Teacher expresses disagreement with a statement from another teacher or expresses that they in some way cannot do a stated assessment practice or include an assessment characteristic
Gives example	Teacher provides an classroom example to illustrate a practice or characteristic (can be hypothetical)

Table 11 – Nature of Explanation Codes

Nature of Statement	The manner in which teachers discuss certain topics. (Reasoning/characteristic/practice stated as…)
Action oriented	… an action to be performed by the teacher/student/assessment without reference to an identifiable cause or outcome (including practices that they do/try to implement).
Causal	… either a direct cause or result of an observable phenomena.
Descriptive	… a quality of the assessment/teacher/student.

Appendix 2 – Complete Discourse Map





751	752	753	754	755	756	757	758	759	760	761	762	763	764	765	766	767	768	769	770	771	772	773	774	775
				A				A			C	→												A
		C	→														A							
C					C	→	→			C				C	→	→		C	→	→	→		C	→
						A			C				A									C		
	C																							
751	752	753	754	755	756	757	758	759	760	761	762	763	764	765	766	767	768	769	770	771	772	773	774	775

776	777	778	779	780	781	782	783	784	785	786	787	788	789	790	791	792	793	794	795	796	797	798	799	800
	A																							A
		A	→			C		C		C			C											A
→	→	→	→	→	→		C		C													A		
						A					C	→												
														C	→	→	→	→	→	C	→	→	→	→
776	777	778	779	780	781	782	783	784	785	786	787	788	789	790	791	792	793	794	795	796	797	798	799	800

801	802	803	804	805	806	807	808	809	810	811	812	813	814	815	816	817	818	819	820	821	822	823
	A			C		C					A		C	→	→	→	→	→	→			
	A	A																				
					A							A										
								A														
				A					A													
→	→		C				C	→	→	→										Cc	Cc	Cc
801	802	803	804	805	806	807	808	809	810	811	812	813	814	815	816	817	818	819	820	821	822	823

AUTHOR INFORMATION

Corresponding Author

*E-mail: yeziere@miamioh.edu

ACKNOWLEDGMENTS

We thank the high school chemistry teachers for participating in this project. We also thank the Yezierski and Bretz research groups at Miami University for their feedback and guidance. This material is based upon work supported by the U. S. National Science Foundation under Grant No. (DRL-118749).

REFERENCES

1. Akerson V. L., Cullen T. A., and Hanson D. L., (2009), Fostering a community of practice through a professional development program to improve elementary teachers' views of nature of science and teaching practice. *J. Res. Sci. Teach.*, **46**(10), 1090–1113.
2. Attard K., (2012), Public reflection within learning communities: An incessant type of professional development. *Eur. J. Teach. Educ.*, **35**(2), 199–211.
3. Banilower E. R., Smith P. S., Malzahn K. A., Plumley C. L., Gordon E. M., and Hayes M. L., (2018), Report of the 2018 National Survey of Science and Mathematics Education. in *Report of the 2018 NSSME+*. Chapel Hill, NC: Horizon Research, Inc.
4. Barnes D. R. and Todd F., (1995), Communication and learning revisited: Making meaning through talk, Portmouth, NH: Boynton/Cook Publishers.
5. Baskerville D. and Goldblatt H., (2009), Learning to be a critical friend: From professional indifference through challenge to unguarded conversations. *Cambridge J. Educ.*, **39**(2), 205–221.
6. Bell B. and Cowie B., (2001), The Characteristics of Formative Assessment. *Sci. Educ.*, **85**(5), 536–553.
7. Black P. and Wiliam D., (1998), Inside the Black Box: Raising Standards Through Classroom Assessment. *Phi Delta Kappan*, **80**(2), 139–148.
8. Buck G. A., Trauth-Nare A., and Kaftan J., (2010), Making formative assessment discernable to pre-service teachers of science. *J. Res. Sci. Teach.*, **47**(4), 402–421.
9. Childs A. and McNicholl J., (2007), Investigating the Relationship Between Subject Content Knowledge and Pedagogical Practice Through the Analysis of Classroom Discourse. *Int. J. Sci. Educ.*, **29**(13), 1629–1653.
10. Clift R., Veal M. L., Johnson M., and Holland P., (1990), Restructuring Teacher Education Through Collaborative Action

Research. *J. Teach. Educ.*, **41**(2), 52–62.

11. Clinchot M., Ngai C., Huie R., Talanquer V., Banks G., Weinrich M., et al., (2017), Better Formative Assessment: Making formative assessment more responsive to student needs. *Sci. Teach.*, **84**(3), 69–75.
12. Cole R., Becker N., Towns M., Sweeney G., Wawro M., and Rasmussen C., (2012), Adapting a methodology from mathematics education research to chemistry education research: Documenting collective activity. *Int. J. Sci. Math. Educ.*, **10**(1), 193–211.
13. Curry M. W., (2008), Critical Friends Groups: The Possibilities and Limitations Embedded in Teacher Professional Communities Aimed at Instructional Improvement and School Reform. *Teach. Coll. Rec.*, **110**(4), 733–774.
14. Darling-Hammond L., Amrein-Beardsley A., Haertel E., and Rothstein J., (2012), Evaluating Teacher Evaluation. *Kappan*, **93**(6), 8–15.
15. Dedoose Version 8.0.35, (2018), web application for managing, analyzing, and presenting qualitative and mixed method research data.
16. Dunne F. and Honts F., (1998), "That Group Really Makes Me Think!" Critical Friends Groups and the Development of Reflective Practitioners. *Annu. Meet. Am. Educ. Res. Assoc.*
17. Fletcher T., Ní Chróinín D., and O'Sullivan M., (2016), A Layered Approach to Critical Friendship as a Means to Support Pedagogical Innovation in Pre-service Teacher Education. *Stud. Teach. Educ.*, **12**(3), 302–319.
18. Gee J. P. and Green J., (1998), Discourse Analysis , Learning , and Social Practice : A Methodological Study. *Rev. Res. Educ.*, **23**, 119–169.
19. Gibbs G. and Simpson C., (2004), Conditions Under Which Assessment Supports Students' Learning. *Learn. Teach. High. Educ.*, **1**(1), 3–31.
20. Graham P., (2007), Case Study Professional Learning COPs. **31**(1).
21. Hamilton L., Halverson R., Jackson S. S., Mandinach E., Supovitz J. A., Wayman J. C., et al., (2009), Using Student Achievement Data to Support Instructional Decision Making, Washington, DC.
22. Harshman J. and Yezierski E., (2017), Assessment Data-driven Inquiry: A Review of How to Use Assessment Results to Inform Chemistry Teaching. *Sci. Educ.*, **25**(2), 97–107.
23. Harshman J. and Yezierski E., (2016), Characterizing high school chemistry teachers' use of assessment data via latent class analysis. *Chem. Educ. Res. Pr.*, **17**(2), 296–308.
24. Harshman J. and Yezierski E., (2015), Guiding teaching with assessments: high school chemistry teachers' use of data-driven inquiry. *Chem. Educ. Res. Pract.*, **16**(1), 93–103.
25. Heady J. E., (2000), Assessment - A Way of Thinking About Learning - Now and in the Future: The Dynamic and Ongoing Nature of Measuring and Improving Student Learning. *J. Coll. Sci. Teach.*, **29**(6), 415–421.
26. Herrington D. G. and Yezierski E. J., (2014), Professional development aligned with AP chemistry curriculum: Promoting science practices and facilitating enduring conceptual understanding. *J. Chem. Educ.*, **91**(9), 1368–1374.
27. Irons A., (2008), Enhancing Learning Through Formative Assessment and Feedback, New York, NY: Routledge.
28. Kaartinen S. and Kumpulainen K., (2002), Collaborative inquiry and the construction of explanations in the learning of science. *Learn. Instr.*, **12**(2), 189–212.
29. Knapp M., Swinnerton J., Copland M., and Monpas-Huber J., (2005), Data-Informed Leadership in Education, Seattle,

WA.

30. Kumpulainen K. and Mutanen M., (1999), The situated dynamics of peer group interaction: An introduction to an analytic framework. *Learn. Instr.*, **9**(5), 449–473.
31. Lieberman A., (1986), Collaborative research: Working with, not working on. *Educ. Leadersh.*, 28–32.
32. Lyon E. G., (2011), Beliefs, Practices, and Reflection: Exploring a Science Teacher's Classroom Assessment Through the Assessment Triangle Model. *J. Sci. Teacher Educ.*, **22**(5), 417–435.
33. Martone A. and Sireci S. G., (2009), Evaluating Alignment Between Curriculum, Assessment, and Instruction. *Rev. Educ. Res.*, **79**(4), 1332–1361.
34. Maxwell J. A., (2013), Qualitative Research Design: An Interactive Approach, 3rd ed. Knight V. (ed.) Thousand Oaks, California: SAGE Publications.
35. Moje E. B., (1997), Exploring Discourse, Subjectivity, and Knowledge in Chemistry Class. *J. Classr. Interact.*, **32**(2), 35–44.
36. Moore J. A. and Carter-Hicks J., (2014), Let's Talk! Facilitating a Faculty Learning Community Using a Critical Friends Group Approach. *Int. J. Scholarsh. Teach. Learn.*, **8**(2), 1–17.
37. National Research Council, (2014), Developing assessments for the next generation science standards, Washington, DC: The National Academies Press.
38. National Research Council, (2001), Knowing what students know: The science and design of educational assessment, Pelligrino J., Chudowsky N., and Glaser R. (eds.) Washington, DC: National Academy Press.
39. National Research Council, (1999), The Assessment of Science Meets the Science of Assessment: Summary of a Workshop, Washington, DC: The National Academies Press.
40. NGSS Lead States, (2013), Washington, DC: The National Academies Press.
41. O'Connor M. C. and Michaels S., (1993), Aligning Academic Task and Participation Status through Revoicing : Analysis of a Classroom Discourse Strategy. *Anthropol. Educ. Q.*, **24**(4), 318–335.
42. Pellegrino J. W., (2012), Assessment of science learning: Living in interesting times. *J. Res. Sci. Teach.*, **49**(6), 831–841.
43. Remesal A., (2011), Primary and secondary teachers' conceptions of assessment: A qualitative study. *Teach. Teach. Educ.*, **27**(2), 472–482.
44. Richardson V., (1996), The case for formal research and practical inquiry in teacher education. in Murray F. B. (ed.), *The teacher educator's handbook: Building a knowledge base for the preparation of teachers*. San Fransisco, CA: Jossey-Bass, pp. 715–737.
45. Sandlin B., Harshman J., and Yezierski E., (2015), Formative Assessment in High School Chemistry Teaching: Investigating the Alignment of Teachers' Goals with Their Items. *J. Chem. Educ.*, **92**(10), 1619–1625.
46. Sato M., Wei R. C., and Darling-Hammond L., (2008), Improving Teachers' Assessment Practices Through Professional Development: The Case of National Board Certification. *Am. Educ. Res. J.*, **45**(3), 669–700.
47. Schuck S. and Russell T., (2005), Self-Study, Critical Friendship, and the Complexities of Teacher Education. *Stud. Teach. Educ.*, **1**(2), 107–121.
48. Smith P. S., (2013), 2012 National Survey Of Science and Mathematics Education, Chapel Hill, NC.
49. Snow-Gerono J. L., (2005), Professional development in a culture of inquiry: PDS teachers identify the benefits of

professional learning communities. *Teach. Teach. Educ.*, **21**(3), 241–256.

50. Stiggins R. J., (1988), Revitalizing Classroom Assessment: The Highest Instructional Priority. *Phi Delta Kappan*, **69**(5), 363–368.
51. Stiggins R. J., (2001), The Unfulfilled Promise of Classroom Assessment. *Educ. Meas. Issues Pract.*
52. Suskie L., (2009), Assessing Student Learning: A Common Sense Guide, 2nd ed. San Fransisco, CA: Jossey-Bass.
53. Swaffield S., (2004), Critical friends: Supporting leadership, improving learning. *Improv. Sch.*, **7**(3), 267–278.
54. Taylor R. T. and Storey V. A., (2013), Leaders, critical friends, and the education community. *J. Appl. Res. High. Educ.*, **5**(1), 84–94.
55. Towndrow P. A., Tan A.-L., Yung B. H. W., and Cohen L., (2010), Science Teachers' Professional Development and Changes in Science Practical Assessment Practices: What are the Issues? *Res. Sci. Educ.*, **40**(2), 117–132.
56. Towns M. H., (2014), Guide to developing high-quality, reliable, and valid multiple-choice assessments. *J. Chem. Educ.*, **91**(9), 1426–1431.
57. Vangrieken K., Meredith C., Packer T., and Kyndt E., (2017), Teacher communities as a context for professional development: A systematic review. *Teach. Teach. Educ.*, **61**, 47–59.
58. Witte R., (2012), Classroom Assessment for Teachers, New York, NY: McGraw-Hill.