



Cite this: *Chem. Educ. Res. Pract.*,  
2014, 15, 777

## Development of a protocol to evaluate the use of representations in secondary chemistry instruction

Stephanie B. Philipp, Destinee K. Johnson and Ellen J. Yeziarski\*

Although observational protocols have been developed that assess different aspects of science teaching, none of the protocols existing in the literature address the principles of effective chemistry instruction guided by Johnstone's triangle of macroscopic, symbolic, and particulate representations of matter (Johnstone, 1991). We developed our own protocol, the Representations in Chemistry Instruction (RICI) protocol, to meet this need. RICI's research-based indicators include: (1) who used the representations (teacher or student) during instruction; (2) the role of representations in improving conceptual understanding; (3) the quality of discourse around the representations; and (4) the degree to which different representations (macroscopic, symbolic and particulate) were integrated in the lesson. The protocol was evaluated for face validity by a panel of chemistry education researchers and for reliability with evaluation of paired observations by two researchers, resulting in a Cohen's kappa of 0.71. The RICI protocol, used with an inquiry-based instruction observation protocol, like the similarly formatted EQUIP (Marshall *et al.*, 2009), can evaluate the quality of secondary chemistry instruction for education researchers and professional development leaders, provide feedback to chemistry teachers for professional growth, and serve as a model for effective use of representations in chemistry instruction.

Received 7th May 2014,  
Accepted 3rd August 2014

DOI: 10.1039/c4rp00098f

[www.rsc.org/cerp](http://www.rsc.org/cerp)

Observer documentation of instructional practices and the complex discourse that occurs in secondary chemistry classrooms is very challenging. One solution has been to create observation protocols framed around research-based teaching practices, such as the Reformed Teaching Observation Protocol (RTOP) (Sawada *et al.*, 2002). Over the past decade a number of such protocols have been developed for science instruction; however, no instrument has been either developed for or recommended specifically for chemistry. As such, the goal of this project was to develop an observation protocol having a particular interest in the features that best meet the unique needs of documenting and evaluating inquiry-based secondary chemistry instruction and classroom discourse. Moreover, we wanted to develop a protocol to use alongside previously published inquiry-based teaching observation protocols.

## Background

*Rising Above the Gathering Storm* (NRC, 2007) strongly recommended improving primary and secondary science and mathematics education in order to increase America's talent pool in science, technology, engineering, and mathematics (STEM), because these subject areas are critical to our country's global competitiveness. This report led to

increased discussion about measuring the quality of instruction in STEM classrooms because there is considerable agreement in the literature that instructional quality affects student learning more than any other in-school factor (Kane and Staiger, 2012). Previous research has demonstrated that teachers' self-report data on the frequency of reform-oriented instructional practices meet reasonable standards of validity and reliability, but teachers are clearly not in a position to judge the quality of their own instruction (Weiss *et al.*, 2003). Research and policy communities are more interested in learning about classroom practice through the eyes of the external observer. Therefore, it is imperative that researchers have the means to measure the quality of instruction being implemented in classrooms.

Research-based observation protocols that generate valid and reliable data are needed, but from a measurement perspective, this poses a huge challenge because of the complex, multi-dimensional nature of inquiry instruction that should be the framework of secondary STEM classes (Marshall *et al.*, 2009). In the U.S. National Science Education Standards, inquiry is defined as

*a multi-faceted activity that involves making observations; posing questions; examining books and other sources of information to see what is already known in light of experimental evidence; using tools to gather, analyze, and interpret data; proposing answers, explanations, and predictions; and communicating the results. Inquiry requires*

Department of Chemistry and Biochemistry, Miami University, Oxford, OH 45056, USA. E-mail: yeziarski@miamioh.edu

identification of assumptions, use of critical and logical thinking, and consideration of alternative explanations (NRC, 1996, p. 23).

The importance of inquiry instruction in learning science is reiterated in the blueprint for the U.S. Next Generation Science Standards:

*Standards and performance expectations that are aligned to the framework must take into account that students cannot fully understand scientific and engineering ideas without engaging in the practices of inquiry and the discourses by which such ideas are developed and refined. At the same time, they cannot learn or show competence in practices except in the context of specific content* (NRC, 2011, p. 218).

Two observation protocols that have been developed and are often used for assessing the quality of inquiry-based instruction in science classrooms are the Reformed Teaching Observation Protocol (RTOP) (Sawada *et al.*, 2002) and the Electronic Quality of Inquiry Protocol (EQUIP) (Marshall *et al.*, 2009). The EQUIP, in particular, was designed for fine-grained examination of inquiry practice. It not only evaluates the quality of discourse and the use of assessments in instruction, but also provides a holistic view of a teacher's lesson (Marshall *et al.*, 2009). The rationale for choosing these protocols for use in this study and a description of each protocol will be detailed in this paper's methods section.

Although the quality of inquiry-based instruction may be assessed by these published protocols, no protocol has been developed specifically for evaluating the quality of inquiry-based instruction in secondary chemistry classrooms. Unlike many topics in other science disciplines, only a few concepts in chemistry can be explained by a novice in an experiential way. For instance, exploring the behavior of atoms and molecules requires models. This adds complications for teachers attempting to enact inquiry-based instruction, which is already a complex endeavor.

## Theoretical framework

### Constructivism

This study is part of a larger project that aims to evaluate the quality of inquiry-based instruction in the classrooms of chemistry teachers participating in a long-term professional development program. The evaluation calls for research-based classroom observation instruments, such as the RTOP or EQUIP, that were created for use as research tools by third-party observers. The psychological perspective of constructivism framed both the development of these instruments and the quality of instruction that the researchers hoped to assess. Constructivism is based on the supposition that knowledge is constructed in the mind of the learner in a continuous process of building and testing

(von Glasersfeld, 1989; Bodner *et al.*, 2001). Learning environments framed around a constructivist perspective are characterized by the active engagement of the learner in the development of knowledge instead of rote memorization forced upon the learner (Marshall *et al.*, 2004).

The features of instruction that are reformed in a constructivist perspective differ in quality from those in traditional instruction (Brooks and Brooks, 1999). These differences are highlighted in Table 1. One key difference is the presentation of content. In traditional instruction, the course content is presented with an emphasis on small parts and basic skills, whereas reformed instruction focuses on big picture concepts. A second difference is that student activities involve working with primary sources of information and/or manipulative materials as a part of reformed instruction. In a traditional classroom setting, instruction is didactic in nature and the students have a passive role. An observation of traditional instruction will produce a lower score than instruction which reflects constructivist principles when constructivist-based observation instruments, such as the RTOP, are used (see Table 1).

### Johnstone's triangle

To elucidate or develop the features of a classroom observation instrument that specifically allows effective documentation and evaluation of chemistry instruction and discourse, we must not only understand the nature of chemical knowledge, but also what makes chemistry unique from other disciplines. Taber (2013) stated, "Much scholarship in chemical education draws upon the model of there being three 'levels' at which the teaching and learning of chemistry operates" (p. 156). Alex Johnstone (1991, 2000) illustrated these levels of chemical representations as a triangle with the points labeled as macroscopic, submicroscopic, and symbolic (Fig. 1). He suggested that this 'multi-level thought'

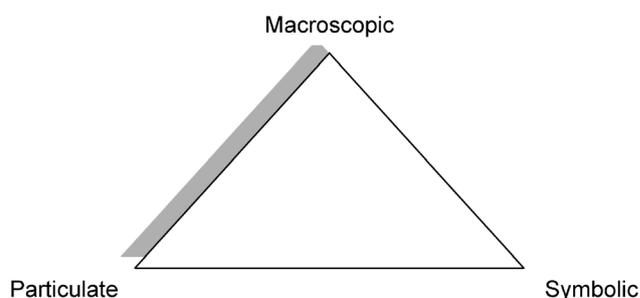


Fig. 1 Johnstone's triangle representing the three domains of chemical knowledge (Johnstone, 1991).

Table 1 Features of traditional and reformed learning environments (Brooks and Brooks, 1999)

Traditional classroom environment	Reformed classroom environment
<ul style="list-style-type: none"> <li>• Curriculum presented in small parts with a fixed pacing</li> <li>• Teachers as authority disseminate information to students for later recall</li> <li>• Teachers seek correct answers to questions</li> </ul>	<ul style="list-style-type: none"> <li>• Curriculum focuses on broad conceptual understanding</li> <li>• Activities involve students analyzing primary sources and using manipulative materials</li> <li>• Teachers seek student questions and points of view in order to improve instruction</li> <li>• Authentic assessment interwoven with teaching</li> </ul>
<ul style="list-style-type: none"> <li>• Assessment of learning is distinct from teaching</li> </ul>	

requirement is what makes all science difficult for students to understand, but chemistry specifically requires the coordination of (a) the macroscopic and tangible: what can be observed by the senses; (b) the submicroscopic: atoms, molecules, ions and structures; and (c) the symbolic: symbols, formulas, equations, molarity, mathematical manipulation, and graphs (as cited in Taber, 2013, p. 157). Students find it difficult to connect the three levels of representing matter and will build fragmented mental models of chemical concepts if they are not given the time and opportunity to integrate multiple representations of concepts (Gabel, 1999). Both Bodner and Domin (2000), and Copolo and Hounshell (1995) found that students who were taught to translate between multiple representations in solving a general or an organic chemistry problem were more likely to be successful problem solvers. Johnstone's triangle as a theoretical framework guides the criteria for the types of representations in chemistry instruction and is a useful framework for an observation instrument developed for chemistry instruction.

## Research questions

Based on the research goal – that is, developing an observation instrument that would work in evaluating features most salient for chemistry classroom – two related questions guide this study:

- (1) Based on previous research and instructional recommendations, what features should be included in an observational protocol to evaluate the representational elements of inquiry-based teaching in high school chemistry?
- (2) How does the evaluation of the instructional use of chemistry representations compare with the evaluation of reformed or inquiry based instruction in high school chemistry lessons?

## Methods

### Participants and data collection

The data for this project were obtained from the study of a professional development program Target Inquiry at Miami University (TIMU). TIMU offers 2.5 years of intense professional development to U.S. high school chemistry teachers with a primary goal of improving the quantity and quality of inquiry-based instruction facilitated in secondary chemistry classrooms (Yeziarski and Herrington, 2013). At the time of this study, only data from the first year of the TIMU project had been collected, hence only teacher data from the first year of the TIMU study are included in this paper. Institutional review board permission was given to the researchers to collect human subjects' data for this study. All research participants were provided with information detailing their rights as human subjects and informed consent was obtained from all of the participants.

During the 2012–2013 academic year, we observed 17 different teachers in southwestern Ohio one time as a baseline evaluation of instruction. The observations were scheduled in advance by both the research participant-teacher and the TIMU Principal Investigator, and were videotaped for future analysis by multiple

researchers. The 17 lessons observed were labeled as student-centered, laboratory-based, and/or inquiry-based by the teacher. Our purpose in observing these lessons was to capture the various instructional methods that experienced chemistry teachers at the beginning of a professional development program considered to be effective student-centered pedagogy. We did not want to influence the teachers' pedagogical choices by constraining their lesson topics, teaching method, or length of lesson. Because the teachers had just started the TIMU program, their knowledge and beliefs about inquiry were constructed during their years of experience prior to the TIMU program. The teachers emphasized various factors that they believed were important to label a lesson as inquiry-based. While many of these factors were quite similar (e.g., students exploring phenomena and taking responsibility for some part of the investigation while teachers provided resources and monitored activity), the teachers possessed varying skills and curricular materials for carrying out inquiry-based instruction. The length of each classroom observation varied based on school schedules (e.g., period or block); however, we observed the entire lesson, as defined by the teacher.

### Instruments

This study originated with a purposeful review of existing observation protocols (Yeziarski, 2014, in press). Protocol selection was based on the following standards: (a) aligns with a constructivist perspective, (b) generates data that not only provides helpful feedback for teachers, but also measures instructional quality, and (c) requires or provides training for observers of varying experience and instructional background. *Electronic Quality of Inquiry Protocol* (EQUIP) (Marshall *et al.*, 2009) and *Reformed Teaching Observation Protocol* (RTOP) (Sawada *et al.*, 2002) were selected for subsequent data analysis due to their alignment with these considerations. A description of each protocol follows.

**EQUIP.** The *Electronic Quality of Inquiry Protocol* was designed and validated by researchers at Clemson University (Marshall *et al.*, 2009). It was designed to measure the quantity and quality of inquiry-based instruction in elementary and secondary mathematics and science classrooms. The EQUIP is a qualitative and quantitative rubric-style instrument that consists of descriptive indicators and a numerical rating scale. The EQUIP has a total of 19 indicators that are grouped into four constructs: instruction, discourse, assessment, and curriculum. Each indicator is rated on a 4-point scale from 1 to 4 (pre-inquiry to exemplary inquiry). Each construct (four to five indicators) is then given a summative rating based on the essence of the indicators within that particular category. After all four constructs are rated, the lesson is assigned a summative score based on the essence of the overall lesson, which is not necessarily an average of the individual construct scores. Possible scores can range from a low 1 to a high 4.

**RTOP.** The *Reformed Teaching Observation Protocol* (RTOP) was designed by the Evaluation Facilitation Group of the Arizona Collaborative for Excellence in the Preparation of Teachers (ACEPT) (Sawada *et al.*, 2002). This instrument consists of 25 items that evaluate a lesson in terms of being (a) standards-based, (b) inquiry-oriented, and (c) student-centered.

The RTOP ultimately provides evidence for whether or not K-20 mathematics and science classes have been *reformed*, or are more constructivist-based and less traditionally didactic (Sawada *et al.*, 2002). The 25 RTOP items are grouped into three scales: lesson design and implementation, content, and classroom culture. Each item is rated on a 5-point Likert-type scale from 0 to 4 (never occurred to very descriptive). An overall RTOP score can range from 0–100, based on the total of all 25 items.

### Training of researchers to use the RTOP and EQUIP

All three researchers completed the training to use the RTOP and EQUIP. Two members of the research team were new to the RTOP and learned to use it by first carefully reviewing a collection of papers about the development and use of the RTOP (Piburn *et al.*, 2000; Sawada *et al.*, 2002; Yeziarski and Herrington, 2011). The RTOP training required the literature review so that researchers could understand the nature of reformed instruction as it is defined and captured by the instrument as well as each of the RTOP subscales and items. After the literature review, they watched a series of online training videos of instruction, rated them using a modified version of the instrument (Yeziarski and Herrington, 2011), and compared their scores to those of expert raters (RTOP developers). After completing the required online training, the researchers debriefed and negotiated the scores from their training videos together. Additional videos of high school chemistry teachers' lessons were rated, and the team debriefed and negotiated scores. This norming process continued in rounds of 2–3 videos until the gap between new and experienced raters was consistently no more than 5%. The EQUIP training was similar to that of the RTOP. All members of the research team were new to the EQUIP and started training with reviewing literature (Marshall *et al.*, 2009), continued with a webinar and practice scoring of selected training videos, and ended with rating the additional videos of the high school chemistry lessons that were used for RTOP training. It was important to complete training on both instruments with classroom videos that were similar to those in the study. By negotiating and norming with high school chemistry videos (rather than training videos that were either a different content area or educational level), we were able to focus on the disciplinary and developmental characteristics most closely related to the sample under study.

### Development of a chemistry instruction observation rubric

Although valid protocols have been developed to assess many different aspects of science teaching, none of the protocols existing in the literature (Yeziarski, 2014, in press) addressed the chemistry-specific criteria for effective instruction guided by Johnstone's triangle (Johnstone, 1991). We endeavored to develop our own protocol to meet this need and to function as a chemistry-specific supplement to an observation protocol for student-centered or inquiry-based instruction. Because the EQUIP (Marshall *et al.*, 2009) was designed and validated specifically for evaluating inquiry-based science and mathematics instruction, we used its descriptive rubric format in our own protocol design.

**Review of literature on chemical representations.** The development of a chemistry instruction observation protocol began with an examination of the literature focused on the teaching and learning of the representational domains of chemical knowledge, and comparisons between novice and expert understanding of chemistry representations. Chemistry representations can be categorized according to Johnstone's triangle as macroscopic (experiential materials or activities), submicroscopic (animations, diagrams, or models characterizing the particulate nature of matter), or symbolic (chemical formulas and symbols, graphs, and equations) (Johnstone, 2000; Taber, 2013). Chemistry representations are used regularly by working chemists to think about their investigations and to explain and justify their findings (Kozma, 2003). Representations of chemical phenomena, especially in the submicroscopic (particulate) domain, are also essential in teaching and learning chemistry in order to provide a framework for understanding chemistry concepts (Gabel *et al.*, 1987). Chemistry novices often focus on the surface features of representations, are most familiar with macroscopic representations, and exhibit difficulty in translating between representations of the three domains of chemistry knowledge (Wu *et al.*, 2001). Chemistry experts, on the other hand, can more competently connect representations from different domains from an understanding of underlying chemical concepts (Kozma and Russell, 1997). To guide student development of conceptual understanding, teachers should support students' connections between the representations and chemistry concepts by (a) explicitly linking representations with verbal descriptions in lessons and in classroom discussion; (b) encouraging student use and development of representations to understand and explain the dynamic nature of chemical phenomena; and (c) promoting understanding of the transformation between two dimensional and three dimensional models (Grosslight *et al.*, 1991; Kozma and Russell, 1997; Wu and Shah, 2004; Chandrasegaran *et al.*, 2008; Chittleborough and Treagust, 2008).

There are compelling reasons for crafting a protocol with a constructivist perspective to evaluate representations in chemistry instruction. For students to move toward expert-like understanding of representations, they must have opportunities to engage in connecting representation levels and negotiating meaning from representations, and not merely adopt or memorize teacher-given representations. For instance, Carolan *et al.* (2008) have published a framework of pedagogical principles to guide use of representations in mathematics and science teaching and learning which is based in student-centered activity involving student generated representations and evaluation of those representations. The foundation of this framework was grounded in the distributed cognition work of Giere and Moffatt (2003) and meta-representational work of DiSessa (2004). Giere and Moffatt relate that it takes both a person and an external representation to accomplish some cognitive tasks. In science learning, this is exemplified by students using representations as tools to develop conceptual understanding. Similarly, DiSessa valued student-centered teaching about representations to encourage students to do the things that scientists do, namely developing models that help explain science concepts (NRC, 2011). Nyachwaya *et al.* (2011) found that only by students creating their

own representations could teachers diagnose problems that the students were having in conceptualizing the particulate nature of matter.

Therefore, based on the review of literature, we created four indicators of representations used to illustrate underlying chemical concepts in chemistry instruction, taking into account the EQUIP's descriptive rubric structure, whereby an increasing score on each indicator indicates a move from more teacher-centered instruction [Pre-inquiry (1 point) and Developing inquiry (2 points)] to more student-centered approaches [Proficient inquiry (3 points) and Exemplary inquiry (4 points)]. Because the framework for the new U.S. Next Generation Science Standards (NRC, 2011), as well as previous U.S. science standards called for inquiry-based science instruction or student engagement in the practices of science, we have situated this protocol in an inquiry-based framework. Specifically, the indicators are: R1: a description of who actually used models in the classroom; R2: the extent to which models were connected to conceptual understanding; R3: the quality of classroom discourse around the representations used; and R4: the degree to which the three representation types (macroscopic, particulate, and symbolic) were integrated in the lesson.

The integration of representation types has been featured in much of the literature on use of representations in chemistry. Gabel *et al.* (1987) recommended an increased emphasis on representing matter at the particulate level. Gabel (1993) reviewed many studies that supported the observation that most chemistry instruction takes place at the symbolic level, and very little connection between macroscopic and particulate levels were observed in typical instruction (Novick and Nussbaum, 1981; Pickering, 1990; Sawrey, 1990; Bunce and Gabel, 1991). In talking with the teachers in this study, most do not recall being taught chemistry with a focus on particulate representations and most have not taught that way with their own students. For its Advanced Placement Chemistry course, the College Board (2013) recently highlighted the importance of emphasizing the particulate level. Therefore, the protocol we have developed values explicit connections at the particulate level with either or both of the other two levels, because explanation occurs at the particulate level in chemistry. Johnstone (1991) warned that integrating all three levels too early in a lesson may result in a cognitive overload, so we have given the option of integration of particulate with only one other domain, depending on the judgment of the teacher. Certainly integration of all three levels in a lesson may be desirable and would be recognized as exemplary inquiry.

Although integration of Johnstone's levels in a lesson are not necessarily a constituent of inquiry-based instruction (integration could occur in a teacher-centered lesson), it would be difficult to imagine an exemplary inquiry-based lesson in which students were responsible for conceptual development without also students making connections between a particulate model and either macroscopic phenomena or symbolic representation, or both (Harrison and Treagust, 1998; Bridle and Yezierski, 2011).

**Development of the representations in chemistry instruction (RICI) protocol.** The qualitative description of every indicator at each level (Pre-inquiry through Exemplary Inquiry) was developed in a highly iterative process involving the three chemistry

education researchers. Two researchers developed the initial indicators and qualitative descriptions for each level of each indicator. A third researcher joined the effort and independently confirmed the alignment of the indicators with the chemical representations literature cited in the previous section. The qualitative descriptors for each indicator were refined as the researchers returned to the literature, seeking specific research-based recommendations for effectively using representations in inquiry-based chemistry instruction. Then two of the researchers jointly applied the draft protocol to six of the 17 videotaped lessons from teachers participating in the TIMU program, carefully discussing and editing word choice to clearly describe each indicator and negotiating scores for 100% agreement.

The face validity of the protocol, now titled the Representations in Chemistry Instruction (RICI) protocol, was assessed by the three researchers, two of whom were former high school chemistry teachers and currently hold doctorates in curriculum and instruction (science/chemistry education). The third researcher was an accomplished undergraduate chemistry major. The researchers also enlisted six chemistry education graduate students, a post-doctoral research fellow and a full professor in chemistry education research (ranging from proficient to expert in chemistry education instrument development) to examine the protocol for validity in measuring inquiry-based use of representations in secondary chemistry classes. Suggestions for clarity were made and incorporated into the protocol. The final protocol validated and employed in this study is located in Table 2.

The lessons rated were fairly homogeneous, but we wanted the RICI to be able to capture the best possible practices for using representations in an inquiry-based chemistry lesson. We considered the research-based features that an exemplary lesson might have and described such features at the Level 4 for each indicator. Following is a vignette that describes an instructional example which would score at this highest level for all indicators on the RICI.

*A teacher suspected that her students did not understand the symbol (aq) in a chemical equation. This understanding was very important for future lessons about chemical reactions. The teacher asked the students to symbolically represent the dissolution of copper(II) chloride. Students worked in pairs to generate their equations and shared them on the board. The teacher facilitated a discussion in which the class came to a consensus about the symbolic representations that were representative of most students' ideas. Student pairs went into the laboratory and dissolved solid copper(II) chloride in water. They observed the water turning blue, the solid disappearing, and the conductivity meter indicating that ions were present. The students were then asked to develop a particulate representation of copper(II) chloride in the form of drawings before and after it dissolved in water. Student pairs compared their drawings with other pairs and discussed similarities and differences among their representations. The teacher listened in on student discussions and prompted groups to talk about how their drawings fell short of their macroscopic observations. The teacher also asked students to consider the role of water in the process. At this point, students used their macroscopic observations and particulate drawings to revise their symbolic equation of the dissolution process. They revised their particulate*

Table 2 Representations in chemistry instruction (RICI) protocol

Indicator	Pre-inquiry (Level 1)	Developing inquiry (Level 2)	Proficient inquiry (Level 3)	Exemplary inquiry (Level 4)
R1: User description	Teacher incorporated representations into lesson.	Students used teacher-provided representations to a small extent.	Students used teacher-provided representations frequently.	Students developed and tested their own representations.
R2: Conceptual understanding	Teacher connected one type of representation (macroscopic, symbolic, or particulate) to chemical concept.	Teacher directs students in using more than one type of representation to describe chemical concept.	Students translated one representation of a chemical concept into another representation type (e.g., macroscopic to symbolic, or symbolic to particulate).	Students generated or selected appropriate representations of a concept to make explanations, predictions, or justifications.
R3: Quality of discourse around representations	There was no discourse around the representations.	Teacher typically controlled and directed the explanation of chemical concepts using the representation. Occasional student input focused on description.	Students used representations to explain chemical concepts but did not debate about scope and limitations of representations.	Students engaged in classroom discussion which included explanation and debate about the scope and limitations of the representations.
R4: Integration of macroscopic, symbolic, and particulate concepts	Lesson focused on one view of chemical phenomena (macroscopic, symbolic, or particulate).	Lesson focused on two views of chemical phenomena but did not explicitly connect multiple domains of chemical knowledge.	Lesson explicitly connected macroscopic and symbolic representational domains.	Lesson connected the particulate representational domain with at least one other domain (macroscopic and/or symbolic).

representations after watching an animation of sodium chloride dissolving in water. The assessment for the lab was to describe sodium chloride dissolving in water using words, symbols and drawings, and explain the role of water in dissolving a salt.

The final RICI protocol was used to rate the remaining 11 (of the 17) videotaped lessons (from 11 different teachers participating in the TIMU study), with two researchers rating all videos independently to analyze inter-rater reliability. The Cohen's kappa ( $\kappa$ ) score was 0.71, which indicates substantial agreement (Landis and Koch, 1977). The EQUIP, whose structure was the model for the RICI, was analyzed for validity and reliability in much the same way (Marshall *et al.*, 2009). The researchers reconciled any differences in the scores from the 11 lessons and those results are reported in this paper, along with the six lessons initially evaluated during the protocol development process.

### Data analysis using the EQUIP and RTOP

Two of the researchers used the EQUIP to evaluate the same 17 lessons used for validation of the RICI. The researchers negotiated independent EQUIP ratings to achieve agreement on the evaluations. The same 17 lessons were analyzed independently by three trained researchers using the RTOP, with similar inter-rater negotiation sessions.

## Results

### Comparison of TIMU data using three observation protocols

The RTOP, EQUIP, and RICI protocol scores from each teacher are reported in Table 3. The RTOP scores ranged from 45 to 69, out of a possible 100 points. The EQUIP scores ranged from

Table 3 Scores from RTOP, EQUIP and RICI observation protocols

Teacher	RTOP score	EQUIP score	Representations in chemistry instruction (RICI) scores			
			User description	Concept understanding	Discourse	Integration
Kelly	45	1+	3	1	2	1
Carl	46	2	3	2	2	2
Meredith	46	2	3	3	2	3
Ben	47	2	3	2	2	3
Tony	47	2	3	2	2	3
Aaron	48	2	3	2	2	2
Belinda	48	3	3	2	2	3
Mason	48	2	3	2	2	3
Melanie	48	2	3	2	2	3
Stacy	48	2	3	1	2	1
Ken	49	2	3	2	2	2
Millie	49	2	3	2	2	2
Sandy	49	2+	3	2	2	2
Tori	49	2+	3	3	2	3
Mallory	51	2	3	2	2	2
Amelia	67	3+	3	2	3	2
Carson	69	4-	3	2	3	2

mostly 1's with some 2's (1+) to mostly 4's but some 2's and 3's (4-). The highest score a lesson can receive with EQUIP is a 4 (exemplary inquiry) and the lowest score is a 1 (pre-inquiry). We reported the RTOP and EQUIP scores to show that lessons observed may have the same inquiry or reformed teaching score overall, but score differently on the RICI, depending on the teacher's use of representations. To examine the features of the new instrument, we reported the results from all four indicators of the RICI separately. Each indicator on the RICI is rated from 1 (pre-inquiry) to 4 (exemplary inquiry), with 3 (proficient inquiry) being the target for most lessons (see Table 3).

## Discussion

### Evaluation of chemistry instruction using the RICI protocol

Our literature review emphasized that a unique and important feature of chemistry instruction is the use of representations at Johnstone's (1991) three levels of chemical thought (macroscopic, symbolic and particulate) to foster deep conceptual understanding. To evaluate how teachers incorporate representations into a chemistry lesson, the RICI protocol was developed using Johnstone's Triangle as a framework. Our results from observing teacher-labeled student-centered activities with the RICI protocol show that, although all the lessons we observed required students to actively use teacher-provided representations (Indicator R1 at level of proficient inquiry), the inquiry levels of the other three indicators varied by the teacher's lesson. Six teachers demonstrated developing inquiry for the other three indicators (3-2-2-2). Seven teachers scored at the proficient inquiry level for two of the indicators and at the developing inquiry level for the other two indicators, in two different ways: (1) five teachers were at the proficient inquiry level in how they integrated representations into the lesson, while at the developing inquiry for conceptual understanding and quality of discourse (3-2-2-3); and (2) two teachers were at the proficient inquiry level in discourse around the representations, while at the developing inquiry level for depth of conceptual understanding and integration of representations (3-2-3-2). Two teachers demonstrated proficient inquiry for all indicators except quality of discourse (3-3-2-3), and two teachers, while at the developing inquiry level for discourse, scored only at the pre-inquiry level for both conceptual understanding and integration of representations in the lesson (3-1-2-1). These varying scores for similar lessons (mostly cookbook-style laboratory investigations) show that the RICI can detect differences in instructional use of representations within lessons of a similar RTOP score. Although all lessons included discourse around the representations, very few of the lessons we observed encouraged students to either explain underlying chemical concepts using representations or discuss the limitations of the representations they were using, therefore most scores for quality of discourse around the representations (Indicator R3) were at the developing inquiry level. We obtained a similar outcome for the use of representations to encourage deeper conceptual understanding (Indicator R2). The 17 lessons we observed focused on chemistry content included in the local science standards for high school

chemistry. Teachers employed laboratory investigations or worksheets assigned to small collaborative groups that verified chemical concepts already discussed in previous class sessions. Most lessons we observed featured representations used by the students with obvious teacher control in order to describe chemical phenomena (such as chemical symbols describing an observed chemical reaction). Only two lessons observed encouraged students to translate a macroscopic representation into a symbolic representation without teacher direction (proficient inquiry level). The RICI was able to assess specific features of each chemistry lesson using descriptive indicators that performed independently from each other.

Using multiple representations in instruction does not necessarily equate to improved inquiry; however, the RICI is situated in an inquiry framework. That being said, a lesson could get a high score on Indicator R4 in a teacher centered classroom (for instance, a teacher-delivered lecture that linked macroscopic, particulate and symbolic representations for students, who receive the information passively), while scoring low on Indicators R1 (who used the representations: teacher), R2 (conceptual depth of the representations: description) and R3 (discourse quality around use of the representations: none or teacher controlled). Use of representations could be considered only a content issue, but the RICI was developed to support the idea that content learning is more effective in an inquiry-based pedagogy.

### Comparison of RICI protocol evaluations to RTOP and EQUIP evaluations

The observational features of the EQUIP, RTOP, and RICI showed that the teachers' lessons were primarily in a phase of developing inquiry. This aligns with the nature of the Target Inquiry program at this point in the program—teachers in the program are interested in “doing more inquiry,” but have had very little experience thus far in meaningfully changing their practice through development of curricular materials or even examining what parts of their practice need the most change to be considered inquiry-based. At the extremes of the protocol scores, only one teacher received a pre-inquiry lesson rating and only one teacher received an exemplary inquiry rating on the EQUIP, while two teachers received pre-inquiry ratings on the RICI protocol, for the indicators R2: connecting representations to conceptual understanding, and R4: integration of representational domains in the lesson. No teacher received an exemplary rating for any indicator on the RICI. Likewise, RTOP scores for 15 of the 17 teacher-labeled student-centered lessons ranged from 45 to 51 out of 100 points. This is consistent with MacIsaac and Falconer's (2002) opinion that lessons scoring between 45 and 55 on the RTOP are more reformed than a traditional lecture with student questions, and typical of a partially reformed high school course with some group work but with most investigation and discourse still under the teacher's control. The narrow range of scores indicated that most of the teachers' lessons and implementation of those lessons were similar in quality. Because most teachers were doing either a “cookbook-style lab” with their students or

had small groups of students working together to complete a worksheet, both of which focused on application of chemistry algorithms or phenomena description, rather than fostering construction of theory, student-generated scientific explanations, or student-led investigations, this narrow range of scores on the RIC, RTOP and EQUIP is a reasonable assessment of the quality of the chemistry instruction observed. As a comparison example, teachers presenting material in a lecture with students participating passively would have scored much lower on the RTOP, probably in the 20–25 point range (MacIsaac and Falconer, 2002). In contrast, two teachers' lessons observed did feature a student-generated investigation of a teacher-posed question, and subsequently scored higher on the RTOP (67 and 69 respectively) and higher on the EQUIP (high 3+ and low 4, respectively).

The lessons' scores on the protocols were also analyzed for performance on specific types of indicators of inquiry-based instruction. This analysis showed that the teachers' lessons scored relatively higher on RTOP items related to lesson design and implementation than on items related to more complex aspects of inquiry-based teaching such as content (propositional or procedural knowledge) or classroom culture. A similar trend was observed with EQUIP scores. On indicators related to instructional factors such as active student learning and teachers as facilitators, observation scores were higher in comparison to indicators related to discourse and assessment. On the RIC, all the lessons observed encouraged students to employ teacher-provided representations frequently throughout the lesson (R1), which is identified as proficient inquiry. Most teachers reached only a developing inquiry level in terms of using representations to merely describe a chemical concept (R2), which may not foster deep conceptual understanding of a phenomenon. Moreover, most teachers controlled the classroom discourse around representations used in the lesson (developing inquiry), rather than encouraging the students to use the representations to explain chemical concepts in their own words (proficient inquiry) (R3). Integrating macroscopic, particulate, and symbolic representational domains was the most disparate indicator (R4). While two of the lessons observed used only one type of representation (macroscopic or symbolic) in the lesson (pre-inquiry), the remaining 15 lessons either focused on more than one type (developing inquiry) or explicitly connected more than one type of representation (proficient inquiry). This analysis of the observational features of the protocols chosen or developed in this study highlights the ability of the protocols to demonstrate the aspects of inquiry-based instruction that are either more easily attained (student engagement) or more difficult to incorporate (activities and discourse that support student-generated models and explanations, and use of particulate representations) for teachers who may be interested in using inquiry-based techniques, but have not developed the resources or expertise to enact high quality inquiry-based instruction.

In comparing the scores that lessons received on the RTOP with the scores on the RIC, the most highly reformed lessons (RTOP scores of 67 and 69) did not score much differently on the RIC from those that implemented a more-teacher centered lab or activity and scored in the mid to high 40's on the RTOP.

The RIC is meant to reliably evaluate how teachers implement representations in an inquiry-based lesson using four indicators, and the RTOP only captures the general use of representations in two items out of 25. RTOP Item 9 in the propositional knowledge construct evaluates the use of representations towards theory development; and RTOP Item 11 in the procedural knowledge construct evaluates the use of a variety of representations. For these two high RTOP lessons, scores on RTOP Item 9 (2 on a 0 to 4 scale) and Item 11 (3 on the 0 to 4 scale), indicate some varied representations were used but with no encouragement toward theory development. Moreover, the RTOP was not designed to evaluate instruction by item or construct. The RIC, however, was designed to identify and describe the following from these two nearly identical lessons: (R1) who was using the representations (students using teacher provided representations), (R2) the depth of conceptual understanding (aimed at description of a chemical concept), (R3) the talk among students about the representations (focused on their own explanations of the concept), and (R4) the integration of representations (did not explicitly connect the macroscopic and symbolic representations). Therefore, the RIC provided a much more complete picture of chemistry-specific instruction with respect to use of representations, while the RTOP and EQUIP were able to evaluate the quality of instruction in other ways. The RTOP was able to quantify the quality of reformed-based teaching, and the EQUIP provided qualitative descriptors that were both reliable and valid for evaluating the lessons of the TIMU teachers in their first year of a 2.5 year professional development program.

## Conclusion

It is imperative that chemistry education researchers have the means to measure the quality of inquiry-based instruction in classrooms. Because of the complex, multifaceted nature of inquiry instruction that should frame chemistry teaching, it has been very challenging to develop an observation protocol that assesses the chemistry instruction in a valid and reliable manner (Marshall *et al.*, 2009). However, our work here provides a description of the development of a chemistry-specific observation protocol, RIC, which could act as a valuable addendum to the EQUIP or RTOP for secondary chemistry teachers (see Table 2). The instrument development descriptions and the new protocol from this study lay the groundwork for further research into inquiry based teaching for secondary chemistry classrooms.

## Implications for research

Similar to the developers of the EQUIP, we were stymied by the lack of a specific observational protocol for our particular type of research on inquiry-based teaching in high school chemistry. The RIC protocol was designed to be used as a type of rubric, along with an inquiry-based or constructivist-based instructional observation protocol (such as the EQUIP or RTOP) for observing high school classrooms where the content being taught is predominantly chemistry or biochemistry and when the lesson is intended to support some degree of inquiry-based

teaching. The RICi captured qualities of chemistry teacher instruction that the RTOP or EQUIP was not developed to do, namely, (a) using Johnstone's (1991) levels of representation in chemistry instruction, especially the particulate level, that is necessary for deep conceptual understanding in chemistry; and (b) supporting students' ability to translate one representation into another representation level depending on conceptual views (Wu *et al.*, 2001); (c) fostering student discourse about concepts underlying the representations used in a lesson (Grosslight *et al.*, 1991); and (d) making explicit connections between multiple representation levels (Wu and Shah, 2004). The RICi was explicitly designed to be used as an observational protocol in an inquiry-based instructional setting. It may be a limitation of the RICi that it was not designed to be used to evaluate chemistry lectures or other teacher-centered instructional methods.

With the protocol's initial interrater reliability measure indicating substantial agreement between trained, independent observers of our study's teacher participants, we have assessed the reliability of the data generated by the RICi protocol with research-based descriptive indicators and adequate user training similar to the EQUIP and RTOP to maintain a high degree of consistency from observer to observer. With its basis in chemistry education literature, the RICi protocol, combined with other appropriate observational protocols, can also help researchers more accurately measure the quality of chemistry instruction than using previously published protocols alone.

### Implications for practice

Observation instruments are beneficial tools for researchers looking to study the quality of instruction being led in classrooms, but they also provide useful information for classroom teachers. Furthermore, observation instruments can illustrate features of quality practice by providing concrete descriptions, especially in protocols like the EQUIP and the RICi which have descriptive indicators for each individual item. Because its descriptors are grounded in research-based recommendations for high quality instruction, the RICi protocol provides meaningful data from its descriptive indicators for chemistry teachers to improve their practice, and ultimately strengthen student-learning outcomes, similar to the EQUIP (Joe *et al.*, 2013). The scores provided from the instrument not only show a teacher how a lesson ranked in terms of representations used for instruction, but can also guide improvement of the lesson by considering the qualitative descriptors associated with higher levels of inquiry for each indicator.

### Future plans

Use of the RICi protocol, along with the RTOP and EQUIP, by trained observers for the TIMU project is ongoing in order further validate the RICi protocol and to measure teacher quality changes in the use of representations in chemistry instruction over a long-term professional development program. We will continue to refine the training for new observers. We also plan to use the RICi protocol as a way to provide feedback for teachers wishing to improve their inquiry practice in chemistry teaching in long-term professional development.

## Acknowledgements

We sincerely appreciate the TIMU teachers' participation in the program. We thank the graduate student members of the Yeziarski Research Group for their contributions to the data collection tasks in this project: Justin Carmel, Jordan Harshman, and especially Sara Nielsen for RTOP evaluation work. We also thank the Bretz Research Group at Miami University. We acknowledge the Miami University Department of Chemistry and Biochemistry for hosting the summer research experience for undergraduates (REU) and the National Science Foundation for funding the REU (NSF #DUE-1004875) and TIMU programs (NSF #DRL-1118749).

## References

- Bodner G. M. and Domin D. S., (2000), Mental models: The role of representations in problem solving in chemistry, *Univ. Chem. Educ.*, **4**(1), 24–30.
- Bodner G., Klobuchar M. and Geelan D., (2001), The many forms of constructivism, *J. Chem. Educ.*, **78**(8), 1107.
- Bridle C. A. and Yeziarski E. J., (2011), Evidence for the effectiveness of inquiry-based, particulate-level instruction on conceptions of the particulate nature of matter, *J. Chem. Educ.*, **89**(2), 192–198.
- Brooks J. G. and Brooks M. G., (1999), *In search of understanding: The case for constructivist classrooms*, Alexandria, VA: Association for Supervision and Curriculum Development.
- Bunce D. and Gabel D., (1991), Improving chemistry achievement through emphasis on the particulate nature of matter. A paper presented at the *Annual National Association of Research in Science Teaching Conference*, vol. 64, Fontana, WI.
- Carolan J., Prain V. and Waldrip B., (2008), Using representations for teaching and learning in science, *Teach. Sci.*, **54**(1), 18–23.
- Chandrasegaran A. L., Treagust D. F. and Mocerino M., (2008), An evaluation of a teaching intervention to promote students' ability to use multiple levels of representation when describing and explaining chemical reactions, *Res. Sci. Educ.*, **38**(2), 237–248.
- Chittleborough G. and Treagust D., (2008), Correct interpretation of chemical diagrams requires transforming from one level of representation to another, *Res. Sci. Educ.*, **38**(4), 463–482.
- College Board (2013), *AP chemistry course and exam description (revised edition)*, New York, NY: College Board. Retrieved from <http://media.collegeboard.com/digitalServices/pdf/ap/ap-chemistry-course-and-exam-description.pdf>.
- Copolo C. E. and Hounshell P. B., (1995), Using three-dimensional models to teach molecular structures in high school chemistry, *J. Sci. Educ. Technol.*, **4**(4), 295–305.
- DiSessa A. A., (2004), Metarepresentation: Native competence and targets for instruction, *Cogn. Instr.*, **22**(3), 293–331.
- Gabel D., (1993), Use of the particle nature of matter in developing conceptual understanding, *J. Chem. Educ.*, **70**(3), 193–194.
- Gabel D., (1999), Improving teaching and learning through chemistry education research: A look to the future, *J. Chem. Educ.*, **76**(4), 548–554.

- Gabel D. L., Samuel K. V. and Hunn D., (1987), Understanding the particulate nature of matter, *J. Chem. Educ.*, **64**(8), 695–697.
- Giere R. N. and Moffatt B., (2003), Distributed cognition: Where the cognitive and the social merge, *Soc. Stud. Sci.*, **33**(2), 301–310.
- Grosslight L., Unger C., Jay E. and Smith C. L., (1991), Understanding models and their use in science: Conceptions of middle and high school students and experts, *J. Res. Sci. Teach.*, **28**(9), 799–822.
- Harrison A. G. and Treagust D. F., (1998), Modelling in science lessons: Are there better ways to learn with models? *Sch. Sci. Math.*, **98**(8), 420–429.
- Joe J. N., Tocci C. M., Holtzman S. L. and Williams J. C., (2013), *Foundations of observation*, Princeton, NJ: Educational Testing Service.
- Johnstone A. H., (1991), Why is science difficult to learn? Things are seldom what they seem, *J. Comput. Assist. Learn.*, **7**, 75–83.
- Johnstone A. H., (2000), Teaching of chemistry: Logical or psychological? *Chem. Educ.*, **1**(1), 9–15.
- Kane T. J. and Staiger D. O., (2012), *Gathering feedback for teaching: Combining high-quality observations with student surveys and achievement gains*, MET Project: Bill & Melinda Gates Foundation.
- Kozma R., (2003), The material features of multiple representations and their cognitive and social affordances for science understanding, *Learn. Instr.*, **13**, 205–226.
- Kozma R. and Russell J., (1997), Multimedia and understanding: Expert and novice responses to different representations of chemical phenomena, *J. Res. Sci. Teach.*, **34**(9), 949–968.
- Landis J. R. and Koch G. G., (1977), The measurement of observer agreement for categorical data, *Biometrics*, **33**, 159–174.
- MacIsaac D. and Falconer K., (2002), Reforming physics instruction via RTOP, *Phys. Teach.*, **40**(8), 479–485.
- Marshall J. C., Smart J. and Horton R. M., (2009), The design and validation of EQUIP: An instrument to assess inquiry-based instruction, *Int. J. Sci. Math. Educ.*, **8**(2), 299–321.
- Marshall J. C., Smart J., Lotter C. and Sirbu C., (2004), Comparative analysis of two inquiry observational protocols: Striving to better understand the quality of teacher-facilitated inquiry-based instruction, *Sch. Sci. Math.*, **111**(6), 306–315.
- National Research Council (ed.), (1996), *National science education standards*, Washington, D.C.: National Academy Press.
- National Research Council (2007), *Rising above the gathering storm: Energizing and employing America for a brighter economic future*, Washington, DC: National Academy of Sciences, National Academy of Engineering, Institute of Medicine.
- National Research Council (2011), *A framework for K-12 science education: Practices, crosscutting concepts, and core ideas*, Washington, DC: National Academies Press.
- Novick S. and Nussbaum J., (1981), Pupils' understanding of the particulate nature of matter: A cross-age study, *Sci. Educ.*, **65**(2), 187–196.
- Nyachwaya J. M., Mohamed A. R., Roehrig G. H., Wood N. B., Kern A. L. and Schneider J. L., (2011), The development of an open-ended drawing tool: an alternative diagnostic tool for assessing students' understanding of the particulate nature of matter, *Chem. Educ. Res. Pract.*, **12**(2), 121–132.
- Piburn M., Sawada D., Turley J., Falconer K., Benford R., Bloom I. and Judson E., (2000), *Reformed teaching observation protocol (RTOP) reference manual*, Tempe, Arizona: Arizona Collaborative for Excellence in the Preparation of Teachers.
- Pickering M., (1990), Further studies on concept learning versus problem solving, *J. Chem. Educ.*, **67**, (3) 254–255.
- Sawada D., Piburn M., Judson E., Turley J., Falconer K., Benford R. and Bloom I., (2002), Measuring reform practices in science and mathematics classrooms: The reformed teaching observation protocol, *Sch. Sci. Math.*, **102**(6), 245–253.
- Sawrey B., (1990), Concept learning versus problem solving: Revisited, *J. Chem. Educ.*, **67**, (3), 253–254.
- Taber K. S., (2013), Revisiting the chemistry triplet: Drawing upon the nature of chemical knowledge and the psychology of learning to inform chemistry education, *Chem. Educ. Res. Pract.*, **14**, 156–168.
- Von Glasersfeld E., (1989), Cognition, construction of knowledge, and teaching, *Synthese*, **80**(1), 121–140.
- Weiss I. R., Pasley J. D., Smith P. S., Banilower E. R. and Heck D. J., (2003), *Looking inside the classroom: A study of K-12 mathematics and science education in the United States*, Chapel Hill, NC: Horizon Research Inc.
- Wu H.-K., Krajcik J. S. and Soloway E., (2001), Promoting understanding of chemical representations: Students' use of a visualization tool in the classroom, *J. Res. Sci. Teach.*, **38**(7), 821–842.
- Wu H.-K. and Shah P., (2004), Exploring visuospatial thinking in chemistry learning, *Sci. Educ.*, **88**, 465–492.
- Yeziarski E. J., (2014), Observation as a tool for investigating chemistry teaching and learning, in Bunce D. and Cole R. (ed.), *Tools of chemistry education research*, Washington, D.C.: American Chemical Society, DOI: 10.1021/bk-2014-1166.ch002.
- Yeziarski E. J. and Herrington D. G., (2011), Improving practice with Target Inquiry: High school chemistry teacher professional development that works, *Chem. Educ. Res. Pract.*, **12**(3), 344–354.
- Yeziarski E. J. and Herrington D. G., (2013), The road to scientific literacy: How Target Inquiry is improving instruction and student achievement in high school chemistry, in Koba S. and Wojnowski B. (ed.), *Exemplary science: Best practices in professional development*, Arlington, VA: National Science Teachers Association, pp. 111–130.