JOURNAL OF CHEMICAL EDUCATION
Article
pubs.acs.org/jchemeduc

Development of the Connected Chemistry as Formative Assessment Pedagogy for High School Chemistry Teaching

Mihwa Park,* Xiufeng Liu, and Noemi Waight

Department of Learning and Instruction, Graduate School of Education, University at Buffalo, Buffalo, New York 14260, United States

S *Supporting Information*

ABSTRACT: This paper describes the development of Connected Chemistry as Formative Assessment (CCFA) pedagogy, which integrates three promising teaching and learning approaches, computer models, formative assessments, and learning progressions, to promote student understanding in chemistry. CCFA supports student learning in making connections among the three domains of chemistry: the macroscopic; the submicroscopic; and the representational. There were 10 sets of computer models and computer-model-based formative assessment tests developed for 10 chemistry topics to enhance student understanding of matter and energy, and models. This article reports the development process of CCFA and evidence supporting the reliability and validity of measures of the formative assessment tests in CCFA based on the Rasch measurement application.

KEYWORDS: *High School/Introductory Chemistry, Chemical Education Research, Computer-Based Learning, Inquiry-Based/Discovery Learning, Testing/Assessment*

FEATURE: Chemical Education Research

INTRODUCTION

While it is expected for students to make connections between the macroscopic (tangible and visible phenomena), the submicroscopic (particles), and the representational (pictorial representations, chemical symbols, equations, etc.) domains in chemistry,[1] students often have difficulty in understanding how these domains are related.[2−4] In recent years, many studies on how to improve student learning and understanding in chemistry have been conducted. In attempting to maximize teacher effectiveness in meeting this student need, we integrated three promising approaches from previous studies to facilitate student learning in chemistry. First, using computerized models for chemistry teaching and learning can facilitate development of understanding the various chemical concepts from elementary through university levels.[5−12] Second, formative assessment is a well-known approach to improving student learning involving activities undertaken by teachers and students to assess themselves in order to provide feedback for modifying the activities in which they are engaged.[13] Research has established both a strong theoretical foundation and empirical evidence to support the use of formative assessment to improve science achievement.[14,15] Third, learning progression theory has emerged as another promising approach for improving science teaching and learning. Learning progression has been recommended as a foundation for organizing science curricula and designing effective science instruction,[2] and was applied in the development of the Next Generation Science Standards.[16,17]

The effectiveness of the three aforementioned approaches (computer modeling, formative assessment, learning progression theory) has been well-documented in previous studies; however, there is a lack of studies that have focused on the use of these three approaches simultaneously.[18] In this paper, we introduce the Connected Chemistry as Formative Assessment (CCFA) pedagogy to integrate the three approaches in chemistry teaching. Following the introduction to the CCFA pedagogy, we report the development process of computer-model-based formative assessment tests.

Three Components of Connected Chemistry as Formative Assessment Pedagogy

In this study, computer models followed the ideas of Connected Chemistry (CC).[19] CC provides a scaffolded learning environment based on NetLogo computer models to enhance student understanding of the connections of three spheres of knowledge:

1. Conceptual understanding of how particle interactions emerge into a system's global behaviors
2. Symbolic expression of the system's behaviors
3. Physical experiences of the phenomenon

In our study, we developed NetLogo models to provide submicroscopic and symbolic representations, illustrating submicroscopic interactions between molecules to help students conceptualize a phenomenon. We also developed another type of computer model called Flash models to primarily address macroscopic and symbolic representations. Flash models demonstrate what is happening with visual representations to provide an explanation of phenomena at the macroscopic level.

In this study, we sought to develop and integrate computer models as a formative assessment tool to support students in making connections among the macroscopic, the submicroscopic, and the representational domains of chemistry. The CCFA pedagogy includes the following topics.

- **Incorporating computer models into instructional activities**: During each unit of instruction, chemistry

Received: April 23, 2016
Revised: December 17, 2016

DOI: 10.1021/acs.jchemed.6b00299
J. Chem. Educ. XXXX, XXX, XXX−XXX

teachers will incorporate the CCFA computer models into a variety of teaching and learning activities, such as introducing a topic, demonstrating a concept, initiating a lab, facilitating a postlab discussion, and conducting a modeling project.

- **Monitoring student understanding of matter−energy and models continuously throughout the chemistry course**: During the middle of each instructional unit, students will complete a CCFA formative assessment test, and their understanding of matter−energy and models will be identified on the basis of a pre-established Rasch scale conversion chart.
- **Implementing differentiated instruction to meet the needs of different students' ability levels of understanding in matter−energy and models**: On the basis of the formative assessment test results, chemistry teachers will be able to provide students with pertinent learning activities during the rest of the instructional unit.

Thus, we intended to develop CCFA computer models as both teaching and learning resources and formative assessment environments.

In the following sections, we present three major components of CCFA pedagogy: (i) formative assessment to monitor chemistry teaching and learning; (ii) computer models to provide pertinent teaching and learning resources; and (iii) learning progressions of two cross-cutting concepts (i.e., matter−energy, and models) as a conceptual framework for the development of assessment and computer models.

Formative Assessment. Assessment refers to all activities that teachers and students undertake to obtain evidence "...to be used as feedback to modify teaching and learning activities. Such assessment becomes formative when the evidence is used to adapt the teaching to meet student needs" (ref 20, p 140). Formative assessment should be able to diagnose what students need in order to improve their learning. Many studies have shown the impact of formative assessment on student achievement in science.[13−15,21] The purpose of formative assessment is to obtain information from students about their progress and learning needs.[22] As ongoing adjustment to the instructional plan is essential to achieve desired instructional objectives,[22] formative assessment is a necessary component of teaching scientific concepts and has proved effective in increasing student scores on external examinations.[14,23] For example, on the basis of their review of more than 250 books and articles, Black and William (1998) concluded that formative assessment combined with appropriate feedback to students could have significant positive effects on student achievement.[13] Similarly, Gallagher (2007) reported a significant increase in student proficiency on a middle school state exam over four years through the use of formative assessment.[14] As such, formative assessment was incorporated as an essential component in CCFA, leading to the development of computer-model-based formative assessments of chemical reasoning.

Models and Modeling. It has been said that "chemistry is a visual science" (ref 24, p 465). As such, introducing students to macroscopic, submicroscopic, and symbolic representations is important for improving understanding of abstract chemistry concepts.[5,24−26] Studies have reported benefits in robust understanding of macroscopic, submicroscopic, and symbolic representations from utilizing models for chemistry teaching.[5,27]

Treagust et al. indicated the importance of scientific models in science teaching: models are used as learning and teaching aids, models represent abstract scientific concepts, and scientists' consensus models of scientific theories are even taught as fact.[26] However, they also pointed out the problem that scientific models have been used superficially in the classroom, which resulted in a failure to employ them as tools for prediction and correlation in scientific phenomena. Chittleborough and Treagust emphasized that scientific models can help students develop their mental models, so it is necessary to know how students understand scientific models to facilitate their learning of science.[28] Treagust and colleagues conceptualized students' understanding of scientific models into five scales:[26]

1. The models as multiple representations (MR) indicating "students' acceptance of using a variety of representations simultaneously, and their understanding of the need for this variety" (ref 26, p 359)
2. The models as exact replicas (ER) referring "students' perception of how close a model is to the real thing" (ref 26, p 359)
3. The models as explanatory tools (ET) referring to "what a model does to help the student understand an idea" (ref 26, p 359), which represents models' roles showing how something works
4. The uses of scientific models (USM) indicating "students' understanding of how models can be used in science, beyond their descriptive and explanatory purpose" (ref 26, p 359)
5. The changing nature of models (CNM) addressing students' understanding of the dynamic changes of models associated with scientific changes[26]

The computer models for CCFA were designed to address the three learning domains of chemistry involving macroscopic phenomenon, symbolic representation, and submicroscopic conception.[1] We developed two different types of computer models, Flash and NetLogo.[29] The Flash models were developed to address primarily macroscopic and symbolic representations, while the NetLogo models were intended to provide primarily submicroscopic and symbolic representations, although both Flash and NetLogo models potentially provide all three types of representations.

Learning Progression. Learning progression is defined as "description of the successively more sophisticated ways of thinking about a topic that can follow one another as children learn about and investigate a topic over a broad span of time" (ref 2, p 219). Initially, we began with conceptualizing the chemical reasoning to consist of three dimensions: understanding of matter, understanding of energy, and understanding of computer models, with each dimension representing one progress variable, characterized by shifts in thinking from descriptive to explanatory, from qualitative to quantitative, and from macroscopic to submicroscopic.[18] As chemistry is a discipline that explores the structure and change of the material (matter) world, considers energy to be a primary cause for particulate structures and changes of matter, and uses models as common tools to reason about matter and energy, expertise in chemistry must simultaneously involve competence in matter, energy, and models.

A similar argument for the existence of a foundational competence in understanding of chemical concepts has been reported in previous literature. For example, Claesgens et al. proposed[30] the perspective of chemists (PC) as a foundational student understanding of chemistry concepts. PC consists of three dimensions: matter, change, and energy.

DOI: 10.1021/acs.jchemed.6b00299
J. Chem. Educ. XXXX, XXX, XXX−XXX

Table 1. Levels of Understanding of Matter−Energy and Models

Level	Level Descriptions for Each Progress Variable: Matter	Energy	Models
1	Describe how matter exists in different states and has various types	Identify different energy forms	Use models for literal illustration of a single phenomenon
2	Explain chemical properties using structure theories of matter (atomic structure and bonding theories)	Explain that energy can be transferred from one form to another	Use models to depict a range of phenomena
3	Explain physical properties by considering the whole system that is characterized by interactions among particles (e.g., intermolecular forces)	Apply the idea that energy transfer tends to increase disorder in matter but the total energy remains constant	Use models as tools to make predictions and test hypotheses

After the first round of field testing, we found that student scores for matter and for energy were highly correlated ($r > 0.8$, $p < 0.05$), and a two-dimensional Rasch model fit the data better than a three-dimensional Rasch model. Combining the matter and energy concepts is not a new idea. Previous studies have combined two core concepts in their investigations of students' learning in science. For example, Lin and Hu (2003) combined energy flow and matter cycling into one topic, and investigated students' understanding of energy flow and matter cycling in various biology topics.[31] Quinn (2014) asserted the need to identify matter with energy to attain a coherent understanding of energy in K−12 science education.[32] She also pointed out a problem in teaching the energy concept as a nonactual entity, although coherent views of energy will be developed when energy is first taught at the submicroscopic level, and then taught in the macroscopic level in the classroom. This assertion indicates that energy and matter understanding are interrelated, so separating energy as an independent dimension from matter might be a problem in K−12 energy education. While we chose combining matter and energy into one dimension, we do acknowledge that in some contexts matter and energy may be better conceptualized as separate dimensions (e.g., Claesgens et al.[30]).

In CCFA, we focused on matter−energy and models, and conceptualized that they together formed a general construct called chemical reasoning. The reason why we considered the models as a dimension was due to our emphasis on computer models in chemistry as both content and context of high school chemistry teaching and learning. We hypothesize that the chemical reasoning will enhance students' understanding of chemistry concepts. We also hypothesize that there is a linear attribute underlying each of the progress variables. The linear attribute is characterized by systematic thinking from descriptive to explanatory, from qualitative to quantitative, and from macroscopic to submicroscopic.[18] In our previous paper, we discussed how the systematic approach in matter, energy, and models could be described.[19] For example, systematic thinking in matter can be achieved by answering these sequential questions: How does matter exist in different states? How does its structure change? How do its properties change chemically? How do its physical properties change? Example questions to characterize systematic thinking in energy could be as follows: What kinds of energy are involved? How is energy transferred? How is the total amount of energy changed? Lastly, systematic thinking in models can be approached by answering these sequential questions: What does the model show? What does the model explain? For what purpose can the model be used?

Finally, we hypothesized that the variables of matter, energy, and models involve three performance levels based on the above hypothesis of students' development of chemical reasoning. Table 1 presents the progress variables for each level of understanding in matter, energy, and models.

The models progress variable was based on the work of Schwarz et al.[33] Because matter−energy and models are related to two cross-cutting concepts (i.e., energy and matter, and systems and system models) in the conceptual framework for the Next Generation of Science Standards[16,17] and are foundational ideas of chemistry, we hypothesize that as students develop their understanding of these fundamental concepts through formative assessment and its associated learning activities, their understanding of chemical concepts can be anticipated to improve.

RESEARCH QUESTIONS

This study intends to answer the following question: What is the evidence of reliability and validity for the computer-model-based formative assessment tests to measure high school students' chemical reasoning following the CCFA pedagogy?

METHOD

Development of Computer-Model-Based Formative Assessment Tests

We developed formative assessment tests based on the three components of CCFA described above. The development of the computer-model-based assessment tests followed the construct modeling approach: (i) starting with a clearly defined construct operationalized by progress variables; (ii) deriving assessment tasks from the variables; and (iii) examining the fit between the variables and collected data from the pilot and field testing.[34]

In this study, the target construct of measurement was chemical reasoning in high school chemistry. Progress variables on matter−energy and models consisted of three levels as described in the Learning Progression section. Students' behaviors on each of the progress variables were based on their responses to computer-model-based assessment questions. That is, students first interacted with computer models, and then answered questions related to specific aspects of the models. This approach was based on ideas of Connected Chemistry (CC).[19,35] Computer models and their associated assessment tests were developed for the following 10 commonly taught topics in a high school chemistry course: atomic structure, periodic table, states of matter, solutions, gases, stoichiometry, chemical bonding, chemical equilibrium, redox, and acids and bases. In other words, each formative assessment includes at least one Flash model of a relevant chemical phenomenon, one NetLogo model, and a set of assessment questions. Each question in the formative assessment test addresses either matter, energy, or the computer model in each chemistry topic. Assessment questions are in the format of Ordered

DOI: 10.1021/acs.jchemed.6b00299
J. Chem. Educ. XXXX, XXX, XXX−XXX

Table 2. Number of Participant Students for Each Formative Assessment Test

Topic	School 1[a]	School 2[b]	School 3[a]	School 4[c]	School 5[c]	School 6[c]	School 7[b]	School 8[a]	School 9[c]	Total
Acids and bases	22	23	24	9	17	10	5	14	0	124
Atomic structure	20	22	24	10	18	13	10	17	20	154
Chemical bonding	22	23	23	10	16	8	6	16	0	124
Chemical equilibrium	20	22	23	9	12	0	9	16	0	111
Ideal gases	21	23	23	9	17	0	10	13	0	116
Periodic table	19	18	21	7	16	0	10	12	0	103
Redox	19	22	23	10	15	8	10	15	0	122
Solutions	0	23	23	10	16	12	3	17	0	104
States of matter	23	24	24	10	16	13	7	18	17	152
Stoichiometry	20	23	13	10	19	0	9	16	10	120
Total										1230

[a]Suburban area schools. [b]Rural area schools. [c]Urban area schools.

Multiple-Choice (OMC),[36] with choices of an OMC question matching different levels of the progress variable.

Pilot Test

The initial version of the computer models and assessment tests was pilot-tested during the 2009−2010 academic year. Every test was piloted with a different class (n = 15−25). Liu and colleagues reported this procedure in detail.[19] During the pilot test, students were asked to comment on the assessment questions to determine if there was any confusion or if a question was unclear for what it was asking. We conducted interviews with selected students by asking them to think aloud about the questions. After revising questions in light of the student feedback, the revised assessments were sent to experts for review.

Three experts with different expertise, a college chemistry professor with a Ph.D. in chemistry, a psychometrician with a Ph.D. in psychometrics and extensive experience in developing OMC questions, and a science educator with a Ph.D. in science education and extensive experience in chemistry education, reviewed the 10 test sets. The expert reviewers were asked to suggest specific ways in which improvements to the assessment questions could be made.

First Field Test

After expert reviews, the tests were revised accordingly, and were then administered to three different schools during the academic year 2010−2011. One school was located in an urban area (n = 23), another school was in a suburban area (n = 26), and the other school was in a rural area (n = 22). Throughout the school year, students completed all 10 computer-model-based formative assessment tests. We conducted interviews with selected students to provide insight into the thought process of students responding to the questions. The findings regarding the reliability and validity of measures from the formative assessments from this testing was reported in a previous publication.[18] Using the two-dimensional Rasch model, we found that all items fit well to the model.[18] Although items fit well in the two-dimensional Rasch model, the analysis results showed that several items should be modified and developed more to assess the wide range of students' understanding in matter−energy and models. We noted that the reliabilities for the formative assessment tests ranged from 0.129 to 0.667 in the first field testing. From analyzing interview data, we found that students did not respond to the questions randomly; rather, they provided reasons why they selected the option in a question even though the OMC was not familiar to the students.[18] We then made further revisions to the assessments, and completed the final version of the computer models and assessment tests.

Extended Field Test

In the present study, we aimed to provide concrete evidence supporting the reliability and validity from extended field-testing of the final version of the computer-model-based formative assessment tests. An extended field testing was conducted during the academic year 2011−2012. There were 10 chemistry teachers from 10 schools who initially participated in the extended field-testing, and 1 teacher dropped out during the fall semester for personal reasons. In this study, we included the nine remaining teachers from nine different high schools. The nine high schools were located in suburban, urban, and rural areas: Schools 1, 3, and 8 were from suburban areas; Schools 4, 5, 6, and 9 were from urban areas; and Schools 2 and 7 were from rural areas (Table 2).

The participant teachers had between 5 and 20 years of chemistry teaching experience. The teachers incorporated computer models into their instructional activities as they planned their lessons. Consequently, the sequence of incorporation of the computer models into instructional activities varied depending on the teacher; however, all students had experience working with computer models by the end of each unit. After students had experience with the computer models, teachers administered computer-model-based assessments to students. Students took the test online individually and were given up to 40 min to complete it. Prior to collecting data, all participant students and teachers signed an informed consent form to participate in this study. In total, 1560 high school students participated in the extended field test. Among them, 330 students did not complete the formative assessments and were excluded from the analysis. Table 2 presents the number of students who completed the formative assessment tests.

During the year, each of the nine teachers incorporated the formative assessments into one of their chemistry classes. The complete computer models and assessments are available for public use online.[37]

Table 3 shows the number of questions in the final version of the formative assessment tests for each chemistry topic.

Here are three examples of final questions: Sample Question 1 is about matter, Sample Question 2 is about energy, and Sample Question 3 is about models (Box 1). Model 1a and Model 1c are both flash models (Figure 1). Model 1a illustrates the color change of the solution system of $CoCl_4^{-2}$. In Model 1a, students can add water or cobalt chloride to the test solution (cobalt chloride solution), and see color changes.

DOI: 10.1021/acs.jchemed.6b00299
J. Chem. Educ. XXXX, XXX, XXX−XXX

Table 3. Number of Questions of Formative Assessment for Each Chemistry Topic

Topic	Matter + Energy[a]	Models	Total
Acids and bases	10	11	21
Atomic structure	7	4	11
Chemical bonding	12	7	19
Chemical equilibrium	9	7	16
Ideal gases	16	12	28
Periodic table	3	6	9
Redox	11	8	19
Solutions	16	8	24
States of matter	16	4	20
Stoichiometry	12	10	22

[a]Matter + Energy indicates that the number of matter questions and the number of energy questions were combined.

Box 1. Final Assessment Questions for Chemical Equilibrium

Sample Question 1. Which of the following statements best explains the color change involving volume increase/reduction in Model 1c?

A. Increasing or reducing volume changes the direction of chemical equilibrium. (Level 2)
B. Increasing or reducing volume changes physical properties of gases. (Level 1)
C. Increasing or reducing volume changes intermolecular forces. (Level 3)

Sample Question 2. Which of the following statements best explains what happens to energy when chloride ions are added in Model 1a?

A. Energy is transferred from chloride ions to reactant molecules. (Level 2)
B. The direction of the reaction changes in order to maintain the energy of the system. (Level 3)
C. The amount of potential energy for reactants and products changes. (Level 1)

Sample Question 3. Which of the following statements best describes what is simulated in the addition of water and chloride ions to the cobalt chloride solution in Model 1a?

A. Chemical equilibrium between $CoCl_4^{-2}$ (aq) and $Co(H_2O)_6^{+2}$ (aq). (Level 2)
B. How the effect of adding reactants and products on chemical equilibrium may be demonstrated visually. (Level 3)
C. How adding water and chloride ions changes the color of the solution. (Level 1)

Model 1c demonstrates color change for the chemical reaction of NO_2 and N_2O_4. Model 1c provides an environment to explore color changes by adding NO_2 or N_2O_4 gases into the equilibrium gases. Students can also control the volume of the system to see the volume and pressure change effects.

In order to place students' understanding of matter−energy and models on the 10 formative assessments onto the same scale such that student learning progression on the two dimensions can be directly compared across topics, linking questions selected from the New York Regents chemistry exams were purposefully inserted into the 10 formative assessment tests. Table 4 shows this linking test design. For example, students who took the State of Matter formative assessment test were also asked to answer 6 State of Matter Regents (SMR) test questions and 6 Atomic Structure Regents (ASR) test questions, and students who took the Atomic Structure formative assessment test were asked to answer the same 12 Regents questions included in the State of Matter formative assessment test and 6 Periodic Table Regents (PR) test questions. Students who took the Periodic Table formative assessment test were given 18 Regents test questions: 6 questions were ASR test questions, 6 questions were PR questions, and 6 questions were Chemical Bonding Regents (CBR) questions. Likewise, all test sets connected with other test sets by including common questions from New York Regents chemistry exams as linking items to place students' abilities in three dimensions, i.e., matter−energy, models, and NY Regents exam, onto the same Rasch scale from the multidimensional Rasch modeling application.[38] The main purpose of the three-dimensional Rasch modeling application was to see the relationship between students' formative assessment scores and the Regents test scores.

Data Analysis

We used a multidimensional Rasch modeling analysis method to examine item and test technical qualities of the assessments. OMC question responses indicated that student understanding levels might be inconsistent across questions.[39,40] To address this issue, we chose to analyze data using the partial credit Rasch model,[41] which takes the following form:

$$\ln\left(\frac{P_{nik}}{1 - P_{nik}}\right) = B_n - D_{ik} \quad (1)$$

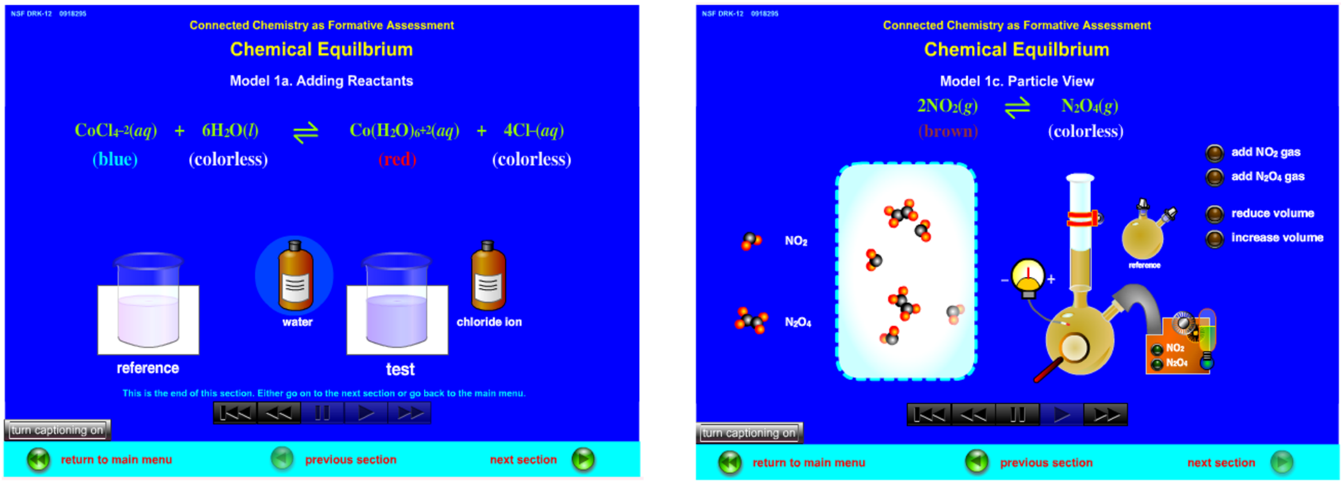
Here, the following abbreviations apply: P_{nik} is the probability for student n with an ability B_n responding at level k instead of level $k - 1$ of item i successfully, and D_{ik} is the difficulty of level k of item i. The partial credit Rasch model yields its own response structure for each item,[42] which enables us to examine possibly inconsistent understanding levels on the progress variables interacting with topics. We also took into consideration multidimensional constructs, namely, chemical reasoning consisting of two unidimensional attributes (matter−energy and models) in our analysis procedure by using multidimensional Rasch models, specifically the Multidimensional Random Coefficients Multinomial Logit (MRCML) model.[43] The MRCML model estimates item and ability parameters based on clusters of items (dimensions), and calculates the correlations between dimensions, which enables us to directly compare item difficulties and student abilities on two dimensions placed on the same scale. ConQuest computer software was used for this Rasch modeling analysis.[44]

RESULTS AND DISCUSSION

Reliability and Validity Evidence of the Computer-Model-Based Formative Assessment Tests

Reliability. Table 5 presents Expected A Posteriori/Plausible Value (EAP/PV) reliability coefficients for the 10 formative assessment tests. EAP/PV reliability is an estimate for test reliability obtained by dividing the variance of the individual expected a posteriori ability estimates by the estimated total variance of the latent ability.[45] EAP/PV reliability can be interpreted like Cronbach's α. From Table 5, we see that the EAP/PV reliability ranges from 0.317 to 0.785. The purpose of our assessments is formative, and formative assessments typically have lower reliabilities than those of summative

DOI: 10.1021/acs.jchemed.6b00299
J. Chem. Educ. XXXX, XXX, XXX−XXX



Model 1a. Adding reactants

Model 1c. Particle view

Figure 1. Flash models for chemical equilibrium.

Table 4. Linking Design of the Formative Assessment

	Regent Test Items Corresponding to CCFA Test Items									
CCFA Topic	SMR[a]	ASR[b]	PR[c]	CBR[d]	STR[e]	GR[f]	CER[g]	SR[h]	ABR[i]	RR[j]
States of matter	6	6								
Atomic structure	6	6	6							
Periodic table		6	6	6						
Chemical bonding			6	6	6					
Stoichiometry				6	6	6				
Gases					6	6	6			
Chemical equilibrium						6	6	6		
Solutions							6	6	6	
Acids and bases								6	6	6
Redox									6	6

[a]SMR = State of Matter Regent test question. [b]ARS = Atomic Structure Regent test question. [c]PR = Periodic Table Regent test question. [d]CBR = Chemical Bonding Regent test question. [e]STR = Stoichiometry Regent test question. [f]GR = Gases Regent test question. [g]CER = Chemical Equilibrium Regent test question. [h]SR = Solutions Regent test question. [i]ABR = Acids and Bases Regent test question. [j]RR = Redox Regent test question.

assessments.[46] During the learning unit when students take a formative test, their understanding of matter−energy and models is still evolving, which may result in low reliability coefficients.

Rasch Validation. The Rasch model produces four fit statistics, i.e., unweighted (using unweighted variance) mean square residual (MNSQ) and weighted (using weighted variance) MNSQ, and unweighted and weighted *t* values, which indicate how well each question fits within the underlying construct. MNSQ is a simple squared residual based on the difference between the observed response patterns and the predicted response patterns. A commonly used criterion for acceptable fit is that MNSQ ranges from 0.7 to 1.3 and *t* value ranges from −2.0 to +2.0. We found that all item fit statistics fell within the acceptable ranges, implying that all items fit well with the two-dimensional partial credit Rasch model across all 10 assessments. Table 6 presents item fit statistics for 28 items in the ideal gases formative assessment test. We can see that all MNSQs were within the range 0.7−1.3, and all standardized MNSQs, i.e., *t* values, were within the range −2.0 to +2.0. This result suggests that all questions fit the two-dimensional partial credit Rasch model well.

In addition to item fit statistics, we examined Wright maps for all assessments to provide more evidence to support validity of measures from assessment tests. In Wright maps, a match between student abilities and item difficulties is necessary in order to keep measurement errors of both student ability and item difficulty measures small. Figure 2 shows the joint Wright map on the two dimensions for the stoichiometry assessment test. The Wright map shows the distribution of person ability

Table 5. Expected A Posteriori/Plausible Value Reliability Results for the 10 Formative Assessment Tests

Test	Matter−Energy Scale[a]	Models Scale[a]
Acids and bases	0.317	0.270
Atomic structure	0.258	0.221
Chemical bonding	0.785	0.376
Chemical equilibrium	0.416	0.353
Ideal gases	0.526	0.514
Periodic table	N/A	0.234
Redox	0.508	0.472
Solutions	0.475	0.365
States of matter	0.425	0.371
Stoichiometry	0.536	0.506

[a]The proportion of student ability estimate variance over the total population ability estimate variance when the results from the measurement focus on the population instead of individual students.

DOI: 10.1021/acs.jchemed.6b00299
J. Chem. Educ. XXXX, XXX, XXX−XXX

Table 6. Fit Statistics for Items on Ideal Gases

		Unweighted Variance		Weighted Variance	
Item[a]	Item Difficulty	Mean Square Residual Values	*t* Values	Mean Square Residual Values	*t* Values
1	0.270	0.98	−0.1	0.98	−0.1
2	0.021	1.00	0.1	1.00	0.0
3	−0.212	0.99	0.0	0.99	−0.1
4	0.361	0.98	−0.1	0.98	−0.2
5	0.375	1.00	0.1	1.01	0.1
6	−0.211	1.00	0.1	1.00	0.0
7	−0.046	0.96	−0.3	0.96	−0.4
8	−0.495	1.01	0.1	1.01	0.1
9	0.377	1.00	0.1	1.00	0.0
10	−0.069	1.04	0.3	1.03	0.4
11	−0.028	0.97	−0.2	0.98	−0.2
12	−0.094	0.98	−0.1	0.98	−0.2
13	0.007	1.02	0.2	1.02	0.2
14	0.339	1.00	0.0	1.00	0.0
15	0.231	1.00	0.0	1.01	0.1
16	−0.193	1.01	0.1	1.01	0.1
17	−0.128	0.99	−0.1	0.99	−0.1
18	0.571	0.90	−0.7	0.94	−0.5
19	0.381	0.92	−0.6	0.96	−0.4
20	0.258	0.98	−0.1	0.98	−0.3
21	−0.130	1.03	0.3	1.04	0.5
22	0.068	1.03	0.3	1.03	0.5
23	−0.271	1.01	0.2	1.01	0.2
24	−0.534	1.00	0.1	1.01	0.1
25	0.238	1.09	0.7	1.08	0.8
26	−0.461	0.98	−0.1	0.98	−0.2
27	−0.611	0.99	−0.1	0.98	−0.2
28	−0.272	1.02	0.2	1.02	0.3

[a]See the Supporting Information for the test items.

(denoted by "X") on the left side of each vertical line and the distribution of item difficulty (denoted by item numbers) on the right side. The questions (person) were distributed from the most difficult (ability) one at the top to the least difficult (ability) one at the bottom.

From Figure 2 we see that overall the items covered the range of student abilities on the matter−energy (Dimension 1) and models (Dimension 2) dimensions. Similar findings were obtained from the other assessment tests. Note that although data from this study fit the Rasch model well and Wright maps showed linear progress in item difficulties, it does not necessarily suggest that student understanding of matter, energy, and models will progress linearly. Studies on learning progressions point out the complexity of learning, so students' linear progress is not always attainable; rather, it may be idiosyncratic.[40,47] In fact, we found that student understanding of matter, energy, and models during the academic year fluctuated from topic to topic while an overall improvement over time was found.

We also examined the relationship between matter−energy scores and models scores and the relationship between all 10 formative assessments scores and the NY Regents test scores. We hypothesize that understanding of matter and energy is positively associated with understanding of computer models. Table 7 presents correlation coefficients between students' abilities on the two dimensions for nine assessments (no questions were developed on matter−energy for the periodic table).

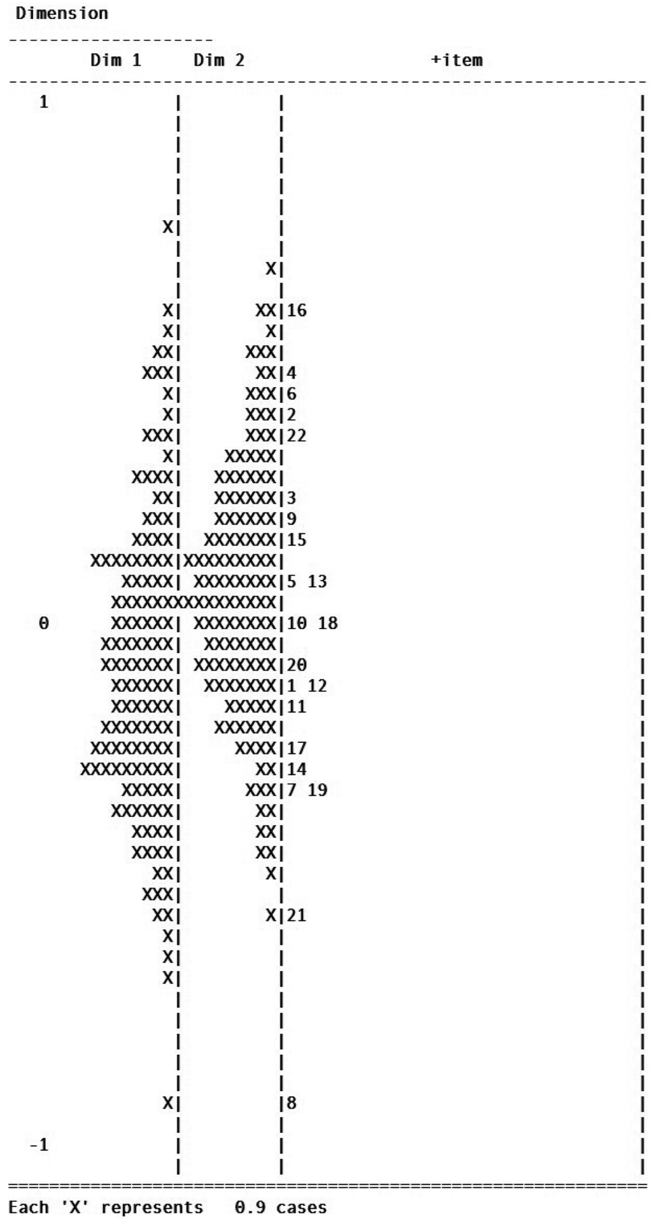


Figure 2. Wright map for the test on stoichiometry. Dimension 1 is the matter−energy dimension, and dimension 2 is the model dimension. Each "X" represents 0.9 respondent. Numbers on the right side indicate item numbers of the stoichiometry test.

Table 7. Results for Correlation between the Matter−Energy and Models Dimensions

Topic	Correlation Coefficients[a]
Acids and bases	0.724
Atomic structure	0.658
Chemical bonding	0.785
Chemical equilibrium	0.814
Ideal gases	0.782
Redox	0.890
Solutions	0.625
State of matter	0.746
Stoichiometry	0.889

[a]Evaluated at $p < 0.01$; $N = 1230$.

From Table 7, we see statistically significant and reasonably strong correlations between students' ability estimates on the

DOI: 10.1021/acs.jchemed.6b00299
J. Chem. Educ. XXXX, XXX, XXX−XXX

matter–energy dimension and on the models dimension across the nine assessments. This result supports the aforementioned hypothesis of a positive relationship between understanding of matter and energy and understanding of computer models, which are components of chemical reasoning in this study. We believe that the development of matter–energy and models understanding as chemical reasoning will enhance the development of students' understanding of chemical concepts as well. Further, we conducted a three-dimensional Rasch modeling analysis to find the relationship among students' scores on matter–energy, models, and the NY Regents tests. The results showed that there was a statistically significant correlation among the three dimensions. Table 8 presents the correlation coefficients.

Table 8. Results for Correlation among Matter–Energy, Models, and NY Regent Exams

Assessment	Matter–Energy[a]	Models[a]
Models	0.764	
NY Regents	0.189	0.277

[a]Evaluated at $p < 0.01$; $N = 1230$.

From Table 8, there was not only a statistically significant correlation between students' scores on matter–energy and that of models tests, but also statistically significant correlations between matter–energy and NY Regents test scores, and between models and NY Regents test scores.

In summary, the findings suggest valid measures from computer-model-based formative assessment tests in that all assessments fit well with the two-dimensional partial credit Rasch model, questions covered students' abilities well, and the relationship to NY Regents chemistry tests were significant. This result provides evidence supporting that the assessments successfully measured the intended construct, i.e., chemical reasoning in each topic.

In order to avoid conducting Rasch analysis every time the formative assessment test is used, we converted students' raw scores (i.e., total points) for matter–energy and modeling questions in each topic to Rasch scales scores. We suggest that teachers score student test results as they would do for a typical test, using the provided scoring rubrics for each topic, and then check the Rasch scale scores corresponding to the raw scores. The scoring conversion tables and scoring rubrics are available in the Supporting Information.

FUTURE STUDY

In the present study, we did not aim to develop a professional development training program for teachers in the use of CCFA. This suggests the need for further study on the effectiveness of teacher training programs in the implementation processes of the CCFA pedagogy to improving student learning in chemistry. In addition, OMC questions contain all correct options; each option addresses different levels of understanding, so they have a limitation in their potential to reveal student misconceptions on computer models or scientific concepts. Regarding this, interviewing students will provide more information on their perceptions and misconceptions for computer models and concepts in chemistry.

ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available on the ACS Publications website at DOI: 10.1021/acs.jchemed.6b00299.

Formative assessment tests, scoring rubrics, and scoring conversion charts (PDF, DOCX)

AUTHOR INFORMATION

Corresponding Author

*E-mail: mihwapar@buffalo.edu.

ORCID

Mihwa Park: 0000-0003-1517-3310

Notes

Any opinions, findings, and conclusions or recommendations expressed in the paper are those of the authors and do not necessarily reflect the views of the National Science Foundation.

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

This paper is based upon work supported by the National Science Foundation under Grant DRL-0918295.

REFERENCES

(1) Johnstone, A. H. The Development of Chemistry Teaching. *J. Chem. Educ.* **1993**, *70*, 701–705.

(2) *Taking Science to School: Learning and Teaching Science in Grades K–8*; Duschl, R. A.; Schweingruber, H. A.; Shouse, A. W., Eds.; The National Academies Press: Washington, DC, 2007; DOI: 10.17226/11625.

(3) Gabel, D. Improving Teaching and Learning through Chemistry Education Research: A Look to the Future. *J. Chem. Educ.* **1999**, *76* (4), 548–554.

(4) Kozma, R. B.; Russell, J. Multimedia and Understanding: Expert and Novice Responses to Different Representations of Chemical Phenomena. *J. Res. Sci. Teach.* **1997**, *34* (9), 949–968.

(5) Chang, H.-Y.; Quintana, C.; Krajcik, J. S. The Impact of Designing and Evaluating Molecular Animations on How Well Middle School Students Understand the Particulate Nature of Matter. *Sci. Educ.* **2010**, *94* (1), 73–94.

(6) Dori, Y. J.; Sasson, I. Chemical Understanding and Graphing Skills in an Honors Case-Based Computerized Chemistry Laboratory Environment: The Value of Bidirectional Visual and Textual Representations. *J. Res. Sci. Teach.* **2008**, *45* (2), 219–250.

(7) Doymus, K.; Karacop, A.; Simsek, U. Effects of Jigsaw and Animation Techniques on Students' Understanding of Concepts and Subjects in Electrochemistry. *Educ. Technol. Res. Dev.* **2010**, *58* (6), 671–691.

(8) Frailich, M.; Kesner, M.; Hofstein, A. Enhancing Students' Understanding of the Concept of Chemical Bonding by Using Activities Provided on an Interactive Website. *J. Res. Sci. Teach.* **2009**, *46* (3), 289–310.

(9) Ozmen, H. Effect of Animation Enhanced Conceptual Change Texts on 6th Grade Students' Understanding of the Particulate Nature of Matter and Transformation during Phase Changes. *Comput. Educ.* **2011**, *57*, 1114–1126.

(10) Plass, J. L.; Milne, C.; Homer, B. D.; Schwartz, R. N.; Hayward, E. O.; Jordan, T.; et al. Investigating the Effectiveness of Computer Simulations for Chemistry Learning. *J. Res. Sci. Teach.* **2012**, *49* (3), 394–419.

(11) Shen, J.; Jing, L.; Chang, H.; Namdar, B. Technology-Enhanced, Modeling-Based Instruction (TMBI) in Science Education. In *Handbook of Research on Educational Communication and Technology*, 4th ed.; Spector, J. M., Merrill, M. D., Elen, J., Bishop, M. J., Eds.; Springer: New York, 2011.

(12) Wu, H.-K.; Krajcik, J. S.; Soloway, E. Promoting Understanding of Chemical Representation: Students' Use of a Visualization Tool in the Classroom. *J. Res. Sci. Teach.* **2001**, *38* (7), 821–842.

DOI: 10.1021/acs.jchemed.6b00299
J. Chem. Educ. XXXX, XXX, XXX–XXX

(13) Black, P.; William, D. *Inside the Black Box: Raising Standards through Classroom Assessment*; King's College London School of Education: London, 1998.
(14) Gallagher, J. J. *Teaching Science for Understanding: A Practical Guide for Middle and High School Teachers*; Merrill: Upper Saddle River, NJ, 2007.
(15) National Research Council. *Assessment in Support of Instruction and Learning: Bridging the Gap between Large-Scale and Classroom Assessment—Workshop Report*; National Academic Press: Washington, DC, 2003.
(16) National Research Council. *A Framework for K−12 Science Education: Practices, Crosscutting Concepts, and Core Ideas*; National Academy of Sciences: Washington, DC, 2012.
(17) NGSS Lead States. *Next Generation Science Standards: For States, by States*; The National Academies Press: Washington, DC, 2013.
(18) Liu, X.; Waight, N.; Roberto Gregorius, G.; Smith, E.; Park, M. Developing Computer Model-Based Assessment of Chemical Reasoning: A Feasibility Study. *J. Comput. Math. Sci. Teach.* **2012**, *31* (3), 259−281.
(19) Stieff, M. Connected Chemistry—A Novel Modeling Environment for the Chemistry Classroom. *J. Chem. Educ.* **2005**, *82* (3), 489−493.
(20) Black, P.; Wiliam, D. Assessment and Classroom Learning. *Assess. Educ. Prin. Polym. Pract.* **1998**, *5* (1), 7−74.
(21) Bell, B. Classroom Assessment of Science Learning. In *Handbook of Research on Science Education*; Abell, S. K., Lederman, N. G., Eds.; Lawrence Erlbaum: Mahwah, NJ, 2007; pp 965−1006.
(22) Liu, X. *Essentials of Science Classroom Assessment*; SAGE: Thousand Oaks, CA, 2010.
(23) Black, P. Assessment by Teachers and the Improvement of Students' Learning. In *International Handbook of Science Education*; Fraser, B. J., Tobin, K. G., Eds.; Kluwer Academic Publishers: Dordrecht, The Netherlands, 1998; pp 811−822.
(24) Wu, H.-K.; Shah, P. Exploring Visuospatial Thinking in Chemistry Learning. *Sci. Educ.* **2004**, *88* (3), 465−492.
(25) Cook, M. P. Visual Representations in Science Education: The Influence of Prior Knowledge and Cognitive Load Theory on Instructional Design Principles. *Sci. Educ.* **2006**, *90* (6), 1073−1091.
(26) Treagust, D. F.; Chittleborough, G.; Mamiala, T. L. Students' Understanding of the Role of Scientific Models in Learning Science. *Int. J. Sci. Educ.* **2002**, *24* (4), 357−368.
(27) Adadan, E.; Trundle, K. C.; Irving, K. E. Exploring Grade 11 Students' Conceptual Pathways of the Particulate Nature of Matter in the Context of Multirepresentational Instruction. *J. Res. Sci. Teach.* **2010**, *47* (8), 1004−1035.
(28) Chittleborough, G.; Treagust, D. Why Models Are Advantageous to Learning Science. *Educacion Quimica* **2009**, *20* (1), 12−17.
(29) Wilensky, U. *NetLogo, Center for Connected Learning and Computer-Based Modeling*; Northwestern University: Evanston, IL, 1999. Retrieved from http://ccl.northwestern.edu/netlogo (accessed Dec 2016).
(30) Claesgens, J.; Scalise, K.; Wilson, M.; Stacy, A. Mapping Student Understanding in Chemistry: The Perspectives of Chemists. *Sci. Educ.* **2009**, *93* (1), 56−85.
(31) Lin, C.-Y.; Hu, R. Students' Understanding of Energy Flow and Matter Cycling in the Context of the Food Chain, Photosynthesis, and Respiration. *Int. J. Sci. Educ.* **2003**, *25* (12), 1529−1544.
(32) Quinn, H. R. A Physicist's Musings on Teaching about Energy. In *Teaching and Learning of Energy in K−12 Education*; Chen, R. F., Eisenkraft, A., Fortus, D., Krajcik, J., Neumann, K., Nordine, J. C., Scheff, A., Eds.; Springer: New York, 2014; pp 15−36.
(33) Schwarz, C. V.; Reiser, B. J.; Davis, E. A.; Kenyon, L.; Achér, A.; Fortus, D.; Shwartz, Y.; Hug, B.; Krajcik, J. Developing a Learning Progression for Scientific Modeling: Making Scientific Modeling Accessible and Meaningful for Learners. *J. Res. Sci. Teach.* **2009**, *46* (6), 632−654.
(34) Wilson, M. *Constructing Measures: An Item Response Modeling Approach*; Lawrence Erlbaum Associates Publisher: Mahwah, NJ, 2005.
(35) (a) Levy, S. T.; Wilensky, U. Crossing Levels and Representations: The Connected Chemistry (CC1) Curriculum. *J. Sci. Educ. Technol.* **2009**, *18* (3), 224−242. (b) Levy, S. T.; Wilensky, U. Students' Learning with the Connected Chemistry (CC1) Curriculum: Navigating the Complexities of the Particulate World. *J. Sci. Educ. Technol.* **2009**, *18* (3), 243−254.
(36) Briggs, D. C.; Alonzo, A. C.; Schwab, C.; Wilson, M. Diagnostic Assessment with Ordered Multiple-Choice Items. *Educ. Asses.* **2006**, *11* (1), 33−63.
(37) Department of Learning and Instruction, SUNY Buffalo. Connected Chemistry As Formative Assessment—Welcome. http://gseweb.gse.buffalo.edu/org/ConnectedChemistry/ (accessed Dec 2016).
(38) Kolen, M. J.; Brennan, R. L. *Test Equating, Scaling, and Linking*; Springer: New York, 2004.
(39) Alonzo, A.; Steedle, J. T. Developing and Assessing a Force and Motion Learning Progression. *Sci. Educ.* **2009**, *93* (3), 389−421.
(40) Steedle, J. T.; Shavelson, R. J. Supporting Valid Interpretations of Learning Progression Level Diagnoses. *J. Res. Sci. Teach.* **2009**, *46* (6), 699−715.
(41) Masters, G. N. A Rasch Model for Partial Credit Scoring. *Psychometrika.* **1982**, *47*, 149−174.
(42) Linacre, J. Comparing "Partial Credit Models" (PCM) and "Rating Scale Models. *Rasch Measurement Transactions* **2000**, *14* (3), 768.
(43) Adams, R. J.; Wilson, M.; Wang, W. The Multidimensional Random Coefficients Multinomial Logit Model. *Appl. Psych. Meas.* **1997**, *21*, 1−23.
(44) Wu, M. L.; Adams, R. J.; Wilson, M. *Conquest: Generalized Item Response Modelling Software—Manual*; Australian Council for Educational Research: Melbourne, Australia, 1997.
(45) Rauch, D. P.; Hartig, J. Multiple-Choice versus Open-Ended Response Formats of Reading Test Items: A Two-Dimensional IRT Analysis. *Psychol. Test. Assess. Model.* **2010**, *52* (4), 354−379.
(46) Eryilmaz, A. Development and Application of Three-Tier Heat and Temperature Test: Sample of Bachelor and Graduate Students. *Eurasian. J. Educ. Res.* **2010**, *40*, 53−76.
(47) Hammer, D.; Sikorski, T. R. Implications of Complexity for Research on Learning Progression. *Sci. Educ.* **2015**, *99* (3), 424−431.

DOI: 10.1021/acs.jchemed.6b00299
J. Chem. Educ. XXXX, XXX, XXX−XXX