

Chemistry Education Research and Practice

Accepted Manuscript



This article can be cited before page numbers have been issued, to do this please use: M. Park, X. Liu, E. Smith and N. Waight, *Chem. Educ. Res. Pract.*, 2017, DOI: 10.1039/C7RP00018A.



This is an Accepted Manuscript, which has been through the Royal Society of Chemistry peer review process and has been accepted for publication.

Accepted Manuscripts are published online shortly after acceptance, before technical editing, formatting and proof reading. Using this free service, authors can make their results available to the community, in citable form, before we publish the edited article. We will replace this Accepted Manuscript with the edited and formatted Advance Article as soon as it is available.

You can find more information about Accepted Manuscripts in the [author guidelines](#).

Please note that technical editing may introduce minor changes to the text and/or graphics, which may alter content. The journal's standard [Terms & Conditions](#) and the ethical guidelines, outlined in our [author and reviewer resource centre](#), still apply. In no event shall the Royal Society of Chemistry be held responsible for any errors or omissions in this Accepted Manuscript or any consequences arising from the use of any information it contains.

Chemistry Education Research and Practice

PAPER

The effect of computer models as formative assessment on student understanding of the nature of models

Mihwa Park,^{*a} Xiufeng Liu,^a Erica Smith^b and Noemi Waight^a

Received 00th January 20xx,
Accepted 00th January 20xx

DOI: 10.1039/x0xx00000x

www.rsc.org/

This study reports the effect of computer models as formative assessment on high school students' understanding of the nature of models. Nine high school teachers integrated computer models and associated formative assessments into their yearlong high school chemistry course. A pre-test and post-test of students' understanding of the nature of models using a published measurement instrument on the nature of models were conducted. A four-step hierarchical multiple regression and a two-level (level 1- student and level 2-teacher) hierarchical linear modeling were used to test the effect of the intervention on students' understanding of the nature of models. Our analysis revealed a significant effect of frequencies of using computer models for four of the five sub-scales related to the nature of models. The implications of these findings are that, as students have more experience using computer models in their classroom, they develop a better understanding of the nature of models. However, their understanding of models as multiple representations didn't show a significant improvement, possibly due to the lack of support from teachers, who in turn need both content and pedagogical supports within their teaching.

Introduction

In chemistry teaching, using, constructing, and understanding models and modelling are essential for student conceptual change and developing deeper understanding (Gilbert and Boulter, 2000; Wu *et al.*, 2001; Coll *et al.*, 2005; Liu and Hmelo-Silver, 2009; Adadan *et al.*, 2010). Models within science are conceptualized for (1) explaining an aspect of reality based on a set of assumptions, (2) being limited in every model's scope, which allows for explaining only a limited aspect of a complex reality, and (3) being one possibility, of many, in its power (i.e., ability) to explain a phenomenon (Snir *et al.*, 2003). Coll *et al.* (2005) pointed out that experts within the field of science utilize models as an attempt to explain the macroscopic nature of the phenomenon under study and the relationships between the macroscopic and molecular.

The term *model* has several different meanings; in general, it refers to objects or people that are worthy of emulation (Chamizo, 2013). In science, Gilbert and Boulter (2000) define a model as a representation of a phenomenon built with a specific purpose. Specific purposes in science are to simplify a phenomenon of scientific inquiries, to show compositions of an object, or to visualize abstractions or ideas as objects (Gilbert and Boulter, 2000). Schwarz *et al.* (2009) also define a scientific model as "a representation that abstracts and simplifies a system by focusing on key features to explain and predict scientific phenomena" (p. 633).

In this study, we are concerned with a specific type of model, computer models, as a representation of phenomena within macro, submicro, and symbolic domains of chemistry teaching. Many studies reported positive results from using computer models to

transform different representations (Dori and Hameiri, 2003) and to reduce cognitive load and visuospatial demands (Wu *et al.*, 2001), which are important to improve conceptual understanding in science (Cook, 2006). In a literature review of 61 empirical studies over the last decade, Smetana and Bell (2012) found that computer models could be as effective, and in many ways more effective, than traditional (e.g., lecture-based, textbook-based, and/or physical hands-on) instructional practices in student knowledge acquisition, developing process skills, and facilitating conceptual change. In chemistry, Barak and Dori (2005) concluded that computer models promote chemistry understanding by illustrating chemical concepts at the macroscopic, submicroscopic, symbolic, and chemical process levels. Ardac and Akaygun (2004)'s study provides further evidence of the positive effect of technology on developing students' understanding of chemistry by enabling students to visualize reactions occurring at a molecular level that would otherwise not be visible to students.

Even though using models has been shown to facilitate student learning in science, studies have also documented students' difficulties in understanding of models or limited use of models within science classrooms (Gilbert and Boulter, 2000; Snir *et al.*, 2003; Coll *et al.*, 2005; Schwarz *et al.*, 2009). Treagust *et al.* (2002) argue that students' understanding of the role of models should be assessed because their incorrect understanding of models may hinder students from scientific understanding in various science topics. In this respect, Treagust *et al.* conceptualize students' understanding of the nature of models to consist of five dimensions, and developed a test instrument, Students' Understanding of Models in Science (SUMS), to measure student understanding of models in science on the basis of the dimensions:

(1) the models as multiple representations (MR) meaning "students' acceptance of using a variety of representations simultaneously, and their understanding of the need for this variety" (p. 359).

^a University at Buffalo, Department of Learning and Instruction, Buffalo, NY, USA.
E-mail: mihwapar@buffalo.edu; Tel: +1 7166451007

^b Tulane University, New Orleans, LA, USA

Paper

(2) the models as exact replicas (ER) indicating “students’ perception of how close a model is to the real thing” (p. 359).

(3) the models as explanatory tools (ET) referring to “what a model does to help the student understand an idea” (p. 359).

(4) the uses of scientific models (USM) referring “students’ understanding of how models can be used in science, beyond their descriptive and explanatory purpose” (p. 359).

(5) the changing nature of models (CNM) indicating understanding of the dynamic changes of models related with scientific changes.

Another promising approach to chemistry teaching is formative assessment. Formative assessment involves the process of collecting data from students and using it to improve students’ learning (Wiliam, 2011). Black and Wiliam (1998) conclude from their review of more than 250 books and articles that formative assessment with feedback to students could have significant positive effects on student achievement.

While numerous studies have found that computer models and formative assessment can facilitate students’ understanding of scientific concepts and enhance classroom instruction, there have been no reported studies of the impact of them on student understanding of the models themselves. This study examined the impact of utilizing computer models as formative assessment tools in classroom instruction on students’ understanding of the models in science. This study addressed the following research question: To what extent do computer models and associated formative assessments facilitate student understanding of the nature of models?

Method

Connected Chemistry as Formative Assessment (CCFA)

CCFA is a pedagogy with matter, energy and scientific models as foundational ideas for high school chemistry. The CCFA includes two resources: computer models and computer model-based assessment. We intended to utilize formative assessment as a means for monitoring chemistry teaching and learning; and computer models as an approach to teaching and learning chemistry. Ten sets of computer models were developed for the following ten commonly taught chemistry topics in high school: atomic structure, periodic table, state of matter, solutions, gases, stoichiometry, chemical bonding, chemical equilibrium, redox, and acids-bases. Each set of computer models included at least one Flash model (Fig 1), and one NetLogo model (Wilensky, 1999) (Fig 2). The Flash models were developed to depict macroscopic representation while the NetLogo models were developed to provide submicroscopic and symbolic representations. In other words, Flash models illustrate what is happening with visual representations to provide an explanation of scientific phenomena at the macroscopic level, for example, changing color or temperature. On the other hand, NetLogo models demonstrate submicroscopic interactions between molecules to precisely depict how particles interact each other in a given situation. Both models allow students to manipulate variables so that they can test their hypothesis and explore scientific phenomena. The web-based platform for the models also enables students to interact with each model while providing dynamic, graphical representations.

We also developed ten sets of computer model-based formative assessments to assess student understanding of matter,

energy and models related to the computer models. Each question in the formative assessment addresses either matter, energy or computer models in each chemistry topic. Assessment questions were in an Ordered Multiple-Choice (OMC) format (Briggs *et al.*, 2006), with choices matching different levels of understanding of matter, energy and models. The initial version of the computer models and assessments was pilot-tested with one class of students (n=15-25) and reviewed by three experts with expertise in chemistry, psychometrics and science assessment during the academic year 2009-2010. The revised models and assessments were then field-tested in three high school chemistry classes during the 2010-2011 academic year. Based on data collected during the year, further revisions to the computer models and formative assessments were made.



Fig. 1 Chemical equilibrium Flash model.

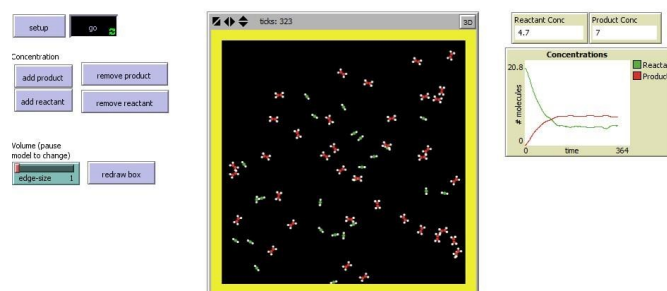


Fig. 2 Chemical equilibrium NetLogo model.

Sample

Nine high school chemistry teachers from nine high schools in the USA participated in the extended field-testing study. Before involving participants in the study, IRB approval was obtained. Participating teachers were selected based on an open call for volunteers. The call was sent to a state-wide science teacher listserv, and the interested teachers were contacted to provide additional information and requirements for participation. One requirement was that the teacher must teach at least two randomly formed sections of the NY Regents chemistry course; another requirement was that there must be a computer lab available for regular use by the experimental class during the entire school year; and the other requirement was that teachers must be willing to travel to the university to complete a full-day teacher professional development on Connected Chemistry as Formative Assessment. Through this procedure, 10 teachers were selected to participate in the project; however, we excluded one teacher from this study since that teacher dropped out for personal reasons. All of the participating teachers were certified in New York State to teach chemistry. The nine teachers had between 5 to 20 years of

experience teaching chemistry in high school. Written informed consent was obtained from participating students and their parents/guardians, and teachers before they entered the study.

During the 2011-2012 school year, the participant teachers incorporated the CCFA into one of their chemistry classes (designated as experimental class), and taught the other chemistry classes in traditional ways (designated as control classes). The recruited teachers received a 1-day training on the formative assessments and computer models prior to the start of the 2011-2012 school year. The training included rationale of formative assessment, nature of the computer models, and how computer models may be incorporated into lessons. A doctoral student in science education was also assigned to each teacher's classroom to support teacher integration of the models in teaching. The doctoral student visited an experimental class 2-3 times during the year; the purpose of the visit was to provide technical support and to check for fidelity of implementation. Note that the doctoral student was not involved in actual teaching. Table 1 presents years of experience teaching chemistry for participant teachers and the number of students who participated in this study.

Intervention

The experimental classes received instruction integrating the CCFA throughout a whole year chemistry course. For each unit, participant teachers were asked to integrate computer models into their unit instruction. For example, teachers introduced pertinent computer models to each unit topic for their lecture and content discussion. After using computer models, teachers asked students to explore specific feature of models and to manipulate variables to test their hypothesis. A student worksheet was also provided for teachers and students to help them use each computer model. At the end of each unit, teachers asked students to complete a computer model based formative assessment. In sum, participant teachers used Flash and NetLogo models in their instructions and monitored student understanding of matter, energy and models by implementing formative assessments to adjust their instruction to meet students' needs. Although the sequence of implementation and the incorporation of the computer models into instructional activities were determined individually by the participating teachers, they all followed a similar implementation plan. While working with the computer models, students were asked to complete worksheets that were designed to guide them through the models and assist them in developing an understanding of the different levels of chemistry understanding for that topic.

Table 1 Teachers' years in chemistry teaching and the number of participating students

Teacher	Years in Chemistry Teaching	N of students in Experimental class	N of students in Control class	Total N
Teacher 1	6	22	22	44
Teacher 2	18	23	50	73
Teacher 3	13	25	55	80
Teacher 4	8	11	35	46
Teacher 5	20	19	15	34
Teacher 6	18	14	23	37
Teacher 7	5	10	20	30
Teacher 8	8	22	78	100
Teacher 9	9	28	52	80
Total		174	350	524

Depending on the teacher, these worksheets were either done independently by students, in small groups, or as an entire classed by the teacher. After students had experienced working with the computer models for each chemistry topic, teachers administered the computer model-based formative assessment for each topic to experimental class students. The assessment was administered individually to students online. While taking the assessments, students were asked to open a second window to view the computer models simultaneously and could also refer to their worksheets for assistance if needed. The purpose of the assessment was to determine students' understanding on matter, energy and computer models. The assessments took between 25 and 40 minutes to complete.

Once students completed the online assessments, the researchers generated students' scores remotely, and provided computer printouts of these scores to teachers. Utilizing the students' score results of understanding of matter, energy and models, the participating teachers were then able to plan and implement differentiated instruction during the next unit.

In order to document teachers' implementation of the CCFA, they also recorded their use of the computer models in a log sheet (e.g., the frequencies of using each of the computer models). The range of frequency of using Flash and NetLogo models was four to 20 during the school year.

Teachers in the comparison classes used their traditional instructional approaches. We did not limit teachers to use computer models in the control classes because some teachers might already use computer models for various purposes such as introducing a topic or demonstrating a concept etc. The difference between experimental classes and comparison classes was that the experimental classes used CCFA systematically while comparison classes many have used computer models on an ad hoc base.

Measurement Instruments

Student understanding of models in science was measured using the Students' Understanding of Models in Science (SUMS) instrument (Treagust *et al.*, 2002). SUMS is a 27-item Likert-scale test instrument. Similar to its use in the study by Treagust *et al.* (2002), we used the instrument to assess how students in the study perceived the roles and uses of models in science. The instrument consists of five sub-scales: models as multiple representations (MR), models as exact replicas (ER), models as explanatory tools (ET), uses of scientific models (USM), and changing nature of models (CNM). Reliability coefficients, i.e., Cronbach's alpha, for the administrations of the scales were reported by Treagust *et al.* (2002) ranging from .71 to .84. The SUMS was given to both the experimental and control classes at the beginning and end of the academic school year. We also administered a measurement instrument, Progression of Understanding Matter (PUM) (Liu, 2007), to all students at the beginning of the school year. The PUM was developed to assess students' knowledge and understanding of basic chemistry concepts. We used student scores on the PUM as a control variable to account for the possibility that the effect of the CCFA may depend on students' prior knowledge on chemistry concepts.

Data Analysis

In this study, we used two methods in the analysis of data. First, hierarchical multiple regression analysis was used to explain the relationship between a dependent variable (outcome variable) and multiple independent variables (predictor variables). Student

Paper

Chemistry Education Research and Practice

pretest scores on the PUM were entered on the first step of the multiple regression as a control variable. Control and experimental classes (0: control class, 1: experimental class) were entered on the second step and frequencies of using Flash and NetLogo models were entered on the third step. Lastly, we entered years of chemistry teaching experience of participant teachers. IBM SPSS Statistics ver 23 was used to perform this hierarchical multiple regression analysis.

Second, predictor variables were submitted to a two-level (level 1- individual and level 2- teacher) hierarchical linear modelling (HLM) to simultaneously test the effect of student and teacher predictor variables. HLM7 (Raudenbush *et al.*, 2010) was used to conduct the hierarchical linear modelling analysis. The student-level variable was control/experimental class (0: control class, 1: experimental class) and frequencies of using Flash and NetLogo models. In addition, student pre-test scores on the PUM were included as a control variable in level 1. The teacher-level variable was years of chemistry teaching experience. Other teacher level variables such as teacher certification in chemistry, were not included because there was little to no variation among the nine teachers on these variables.

Outcome variables for both analyses were gain scores between pre and post-test of the five sub-scales in SUMS. The gain scores for each SUMS sub-scale between the pre and post-test were calculated by subtracting pre-test scores from post-test scores. The above outcome variables were tested separately. All variables used in the analysis were summarized in Table 2.

Table 2 Descriptions of Variables

Variable Name	Description
<i>Predictor Variables</i>	
PRESURVE	Student pre-test scores on Progression of Understanding Matter (PUM) test
CONTROLE	Control/Experimental class; Coded 0 = Control group, 1 = Experimental group
MODELFRE	Frequencies of using Flash and NetLogo models
CHEMT	years of chemistry teaching experience of teachers
<i>Outcome Variables</i>	
Gain MR	Gain scores of models as multiple representations scale
Gain ER	Gain scores of models as exact replicas scale
Gain ET	Gain scores of models as explanatory tools scale
Gain USM	Gain scores of the uses of scientific models scale
Gain CNM	Gain scores of the changing nature of models scale

Results and discussion

In the current study, Cronbach's alpha reliability coefficients for the scales were as follows: models as multiple representations (MR) $\alpha = 0.75$, models as exact replicas (ER) $\alpha = 0.75$, models as explanatory tools (ET) $\alpha = 0.80$, uses of scientific models (USM) $\alpha = 0.73$, and changing nature of models (CNM) $\alpha = 0.77$, which are comparable to what Treagust *et al.* (2002) reported.

Hierarchical Multiple Regression

View Article Online

DOI: 10.1039/C7RP00018A

In the hierarchical multiple regression analysis, we used the listwise deletion in which a case is dropped from an analysis if it has a missing value in at least one of the specified variables. As the result, 30% of students ($n = 159$) were omitted, which results in including 365 students for the analysis ($N = 365$). Outcome variables, Gain MR, Gain ER, Gain ET, Gain USM, and Gain CNM, were entered separately for the hierarchical multiple regression. From Table 3 to 5 show the results of the multiple regression analyses for each outcome variable. Regarding predictor variables, student pre-test scores on PUM (PRESURVE) were entered on the first step (Step 1s) of the multiple regression to see the relationship between student pre-test scores on PUM and each outcome variable. Control and experimental classes (CONTROLE) were entered on the second step, frequencies of using Flash and NetLogo models (MODELFRE) were entered on the third step, and lastly years of chemistry teaching experience of teachers (CHEMT) were entered on the fourth step.

We note that the initial models (Step 1s) entering only student pre-test scores on PUM were not significant for all outcome variables.

Table 3 presents results of hierarchical multiple regression model for Gain MR (gain multiple representations scale scores) as an outcome variable.

Table 3 Results of hierarchical multiple regression for Gain MR

Outcome	Predictor ^a	R ²	Test of Model	Test of Predictor
Gain MR	Step 1	.002	$F(1,363)=0.617$	
	PRESURVE			$\beta = -0.089$
	Step 2	.002	$F(2,362)=0.323$	
	PRESURVE			$\beta = -0.088$
	CONTROLE			$\beta = 0.153$
	Step 3	.002	$F(3,361)=0.242$	
	PRESURVE			$\beta = -0.079$
	CONTROLE			$\beta = 0.606$
	MODELFRE			$\beta = -0.037$
	Step 4	.004	$F(4,360)=0.347$	
	PRESURVE			$\beta = -0.063$
	CONTROLE			$\beta = 0.381$
	MODELFRE			$\beta = -0.011$
	CHEMT			$\beta = -0.074$

Note. ^aPredictor: PRESURVE = student pretest scores on PUM, CONTROLE= control/experimental class, MODELFRE= frequencies of using Flash model and NetLogo model, and CHEMT= years of chemistry teaching experience of teachers.

When including Gain MR as an outcome variable, all models were not significant and none of predictor variables significantly predicted the outcome variable (Gain MR).

When using Gain ER (Gain scores of models as exact replicas scale) and Gain USM (gain the uses of scientific models scale scores) as outcome variables, results from hierarchical multiple regression models were similar, so we present the results together in Table 4.

Table 4 Results of hierarchical multiple regression for Gain ER and Gain USM

Chemistry Education Research and Practice				
Outcome	Predictor	R ²	Test of Model	Test of Predictor
Gain ER	Step 1	.001	F(1,363)=0.383	
	PRESURVE			$\beta = -0.075$
	Step 2	.017	F(2,362)=3.070*	
	PRESURVE			$\beta = -0.067$
	CONTROLE			$\beta = 2.192^*$
	Step 3	.032	F(3,361)=4.012*	
	PRESURVE			$\beta = -0.145$
	CONTROLE			$\beta = -1.804$
	MODELFRE			$\beta = 0.326^*$
	Step 4	.033	F(4,360)=3.063*	
	PRESURVE			$\beta = -0.135$
	CONTROLE			$\beta = -1.947$
	MODELFRE			$\beta = 0.343^*$
	CHEMT			$\beta = -0.047$
	Step 1	.000	F(1,363)=0.150	
	PRESURVE			$\beta = -0.022$
Gain USM	Step 2	.019	F(2,362)=3.531*	
	PRESURVE			$\beta = -0.018$
	CONTROLE			$\beta = 1.133^*$
	Step 3	.029	F(3,361)=3.656*	
	PRESURVE			$\beta = -0.048$
	CONTROLE			$\beta = -0.405$
	MODELFRE			$\beta = 0.126^*$
	Step 4	.036	F(4,360)=3.319*	
	PRESURVE			$\beta = -0.034$
	CONTROLE			$\beta = -0.612$
	MODELFRE			$\beta = 0.149^*$
	CHEMT			$\beta = -0.068$
* $p < .05$				

In the case of Gain ER and Gain USM, all model effects were significant except the first model. In the second models for Gain ER and Gain USM, control and experimental classes (CONTROLE) emerged as a significant predictor, and in the third and fourth models, frequencies of using Flash and NetLogo models (MODELFRE) emerged as a significant predictor. These findings indicated that when the frequency of using Flash and NetLogo models increases, student scores on ER and USM tended to significantly increase.

Lastly, we present results from using Gain ET (gain explanatory tools scale scores) and Gain CNM (gain the changing nature of models scale scores) as outcome variables in Table 5.

Table 5 Results of hierarchical multiple regression for Gain ET and Gain CNM

Outcome	Predictor	R ²	Test of Model	Test of Predictor
Gain ET	Step 1	.000	F(1,363)=0.096	
	PRESURVE			$\beta = -0.27$
	Step 2	.011	F(2,362)=2.003	
	PRESURVE			$\beta = -0.022$
	CONTROLE			$\beta = 1.295^*$
	Step 3	.021	F(3,361)=2.579	
	PRESURVE			$\beta = -0.067$
	CONTROLE			$\beta = -0.997$
	MODELFRE			$\beta = 0.187$
	Step 4	.023	F(4,360)=2.147	

Gain CNM	Step 1	.000	F(1,363)=0.050	
	PRESURVE			$\beta = 0.013$
	Step 2	.000	F(2,362)=0.055	
	PRESURVE			$\beta = 0.013$
	CONTROLE			$\beta = 0.104$
	Step 3	.021	F(3,361)=2.537	
	PRESURVE			$\beta = -0.029$
	CONTROLE			$\beta = -2.019^*$
	MODELFRE			$\beta = 0.173^*$
	Step 4	.021	F(4,360)=1.944	
	PRESURVE			$\beta = -0.033$
	CONTROLE			$\beta = -1.961^*$
	MODELFRE			$\beta = 0.167^*$
	CHEMT			$\beta = 0.019$

* $p < .05$

In cases of Gain ET and Gain CNM as outcome variables, all models were not significant fits of the data. In the second model of Gain ET, control and experimental classes (CONTROLE) emerged as a significant predictor, and in the fourth model, frequencies of using Flash and NetLogo models (MODELFRE) emerged as a significant predictor. In the case of Gain CNM, frequencies of using Flash and NetLogo models (MODELFRE) were a significant predictor for the third and the fourth models, and control and experimental classes (CONTROLE) were also a significant predictor.

In sum, except Gain MR, frequencies of using Flash and NetLogo models (MODELFRE) emerged as a positively significant predictor for the outcome variables when other predictors were controlled. Years of chemistry teaching experience of teachers (CHEMT) contributed no variance in any outcome variables when student pre-test scores on PUM (PRESURVE), control and experimental classes (CONTROLE), and frequencies of using Flash and NetLogo models (MODELFRE) were taken into account.

Although the hierarchical multiple regression analysis revealed that as the frequency of using Flash and NetLogo models increased, students' scores on the models as exact replicas (ER), the models as explanatory tools (ET), the uses of scientific models (USM), and the changing nature of models (CNM) scales increased, we found that overall experimental classes' gain CNM scores were significantly lower than control classes' gain CNM scores when other predictors were controlled. We assumed that it might be due to the possibility that we didn't consider nested data structure in evaluating teacher effects. So we applied hierarchical linear modelling (HLM) to test relationships within and between grouped data, which tests simultaneously the effect of teacher level predictors and the effect of student level predictors on outcome variables.

Hierarchical Linear Modeling

A common feature of data in educational research is hierarchical in that observations are nested within units, for example, students are nested within classroom, which are nested in school, and so on. When data is hierarchically structured, HLM is used to examine the effect variables at the higher levels such as classroom or school

Paper

while simultaneously testing the effect of lower level variables such as student level variables on outcome variables (Raudenbush and Bryk, 2002). Simple linear regression techniques were insufficient for analysing the nested data due to their mistreatment of students' shared variance in a same classroom, i.e., ignorance of group differences (Woltman *et al.*, 2012); in contrast, HLM is capable to take into account the effect of predictor variables on outcome variables if the effect depends on that nesting nature (Raudenbush and Bryk, 2002; Woltman *et al.*, 2012). In other words, HLM takes individual and group level variables into account, then identifies the relationship between predictor and outcome variables. In the study, nine teachers were participated, so we assumed that students might show different results depending on teachers even if the intervention and pre-test scores on PUM were similar.

The Null Model of Teacher and Student Effects. As the convention for HLM, the unconditional model containing only an outcome variable and no independent variables was conducted first. This analysis provides useful information on the variability in an outcome variable by groups to see if there is a need to conduct HLM analysis (Woltman *et al.*, 2012).

First, using an outcome measure as the dependent variable (specifically: Gain MR scores, Gain ER scores, Gain ET scores, Gain USM scores, and Gain CNM scores, see table 2), the model is as follows:

Level-1 Model

$$\text{Dependent Variable}_{ij} = \theta_{0j} + r_{ij}$$

Level-2 Model

$$\theta_{0j} = \gamma_{00} + u_{0j}$$

Table 6 presents the results of the model for each of the dependent variables. Only when using Gain CNM scores (gain the changing nature of models scale scores) as an outcome variable, there was a statistically significant difference between teachers ($\chi^2(8) = 21.91, p = .005$). Between-teacher variance was 0.55 and the within-teacher variance was 14.24, thus the intra-class correlation was 3.7%. The intra-class correlation represents the percent of variance in the outcome variable that is between groups (Woltman *et al.*, 2012). So in the study, 3.7% of the variance in the gain CNM scores was at the group level. The statistically significant between-class variance indicates that average gain CNM score varied significantly across teachers. This supports the use of HLM for gain CNM scores.

Table 6 Unconditional HLM for partitioning variance in gain scores for each SUMS sub-scale

Type of variability	Variance	χ^2
<i>Gain MR Scores</i>		
Between-teacher	0.51	10.23
Within-teacher	60.09	
<i>Gain ER Scores</i>		
Between-teacher	0.03	6.05
Within-teacher	68.95	
<i>Gain ET Scores</i>		
Between-teacher	0.59	13.14

Chemistry Education Research and Practice

Within-teacher	34.51	
<i>Gain USM Score</i>		
Between-teacher	0.21	11.91
Within-teacher	15.25	
<i>Gain CNM Scores</i>		
Between-teacher	0.55	21.91*
Within-teacher	14.24	

* $p < .05$

Full Conditional Random Intercepts and Slopes Model on CNM. Next, full conditional random-intercept and random-slope models were applied by adding level-1 variables (student level variables); control/experimental class (CONTROLE), student pre test scores on PUM (PRESURVE), and frequencies of using Flash and NetLogo models (MODELFRE), and a level-2 variable (teacher level variable); years of chemistry teaching experience (CHEMT), into the each equation were tested. The models were used to identify any interaction effects between predictor variables in different levels (level-1 and level-2). In other words, this step was conducted to test for interaction between student level variables, i.e., CONTROLE, PRESURVE, and MODELFRE, and a teacher level variable, i.e., CHEMT.

When using gain CNM scores as an outcome variable, the model was as follows:

Level-1 Model

$$\text{GAINCNM}_{ij} = \theta_{0j} + \theta_{1j}(\text{CONTROLE}_{ij}) + \theta_{2j}(\text{PRESURVE}_{ij}) + \theta_{3j}(\text{MODELFRE}_{ij}) + r_{ij}$$

Level-2 Model

$$\theta_{0j} = \gamma_{00} + \gamma_{01}*(\text{CHEMT}_{ij}) + u_{0j}$$

$$\theta_{1j} = \gamma_{10} + \gamma_{11}*(\text{CHEMT}_{ij})$$

$$\theta_{2j} = \gamma_{20}$$

$$\theta_{3j} = \gamma_{30} + \gamma_{31}*(\text{CHEMT}_{ij}) + u_{3j}$$

Table 7 Full conditional HLM results to identify interaction between student and teacher level variables

Fixed Effect			
Independent Variables		Gamma Coefficient	Standard Error
Mean achievement	θ_0		
Intercept	γ_{00}	1.56*	0.51
CHEMT ^a	γ_{01}	0.12	0.12
CONTROLE ^b	θ_1		
Intercept	γ_{10}	-0.44	1.05
CHEMT	γ_{11}	-0.17	0.24
PRESURVE ^c	θ_2		
Intercept	γ_{20}	-0.04	0.06
MODELFRE ^d	θ_3		
Intercept	γ_{30}	-0.00	0.09
CHEMT	γ_{31}	0.01	0.02
Random Effect			
		Variance	χ^2
Intercept	u_0	0.65	19.82*
MODELFRE slope	u_3	0.01	26.01*

^aYears of chemistry teaching experience, ^bControl/Experimental class, ^cPre-test scores on PUM, ^dFrequencies of using Flash and NetLogo models. * $p < .05$.

Note. Gamma coefficients of CHEMT variable indicate as following: Student gain CNM scores were higher in classes with teachers who had more experience in teaching chemistry but the result was not statistically significant ($\gamma_{01} = 0.12$, $p > .05$), the cross-level interaction between control/experimental group (CONTROLE) and teacher experiences in teaching chemistry (CHEMT) was not statistically significant ($\gamma_{11} = -0.17$, $p > .05$), and the cross-level interaction between frequencies of using Flash and NetLogo models (MODELFRE) and teacher experiences in teaching chemistry (CHEMT) was not statistically significant ($\gamma_{31} = 0.01$, $p > .05$).

Table 7 presents the results of the above model. From the fixed effect section in Table 7, we can see that there was no statistically significant predictor for the differences in teacher level means (CHEMT) on gain CNM scores. HLM analysis results revealed that there was no cross-level interaction between frequencies of using Flash and NetLogo models and years of chemistry teaching experience ($\gamma_{31} = 0.01$, $p > .05$), meaning that years of chemistry teaching experience had no influence on the strength of the relationship between frequencies of using Flash and NetLogo models and student gain CNM scores. In addition, control and experimental classes didn't have a significant effect on student gain CNM scores ($\gamma_{10} = -0.44$, $p > .05$) when other predictors were controlled. The frequencies of using Flash and NetLogo models showed an insignificant effect on student gain CNM scores ($\gamma_{30} = 0.003$, $p > .05$) either. In sum, the fixed effect section in Table 7 indicates that the degree of teacher experience (CHEMT) had no influence on the strength of the relationship between predictor variables and gain CNM scores; all regression coefficients, γ , were not statistically significant, $p > .05$.

In the random effect section in Table 7, average gain CNM scores among teachers and among frequencies of using Flash and NetLogo models (MODELFRE) varied significantly ($\chi^2(6) = 19.82$, $p < .05$, and $\chi^2(6) = 26.01$, $p < .05$, respectively). This result implies that there still remained additional variability between teachers left unexplained.

Conclusions and implications

Analyses of students' understanding of models in science, as measured by the Students' Understanding of Models in Science (SUMS) instrument (Treagust *et al.*, 2002), revealed no significant differences between the control and experimental classes for four of the five sub-scales tested by the instrument when other predictors, i.e., student pretest score scores on PUM, frequencies of using Flash and NetLogo models, and years of chemistry teaching experience of teachers, were controlled. Hierarchical multiple regression analysis revealed that as teachers used more computer models in their instructions, students' scores on four scales which were the models as exact replicas (ER), the models as explanatory tools (ET), the uses of scientific models (USM), and the changing nature of models (CNM) scales tended to increase as well. Specifically, the more models used in class, the more student advanced their understanding of models. This finding indicated that the implementation of computer models as formative assessment had a significant intervention effect on students' understanding of nature of models.

HLM analysis results showed that overall students' gain scores were not significantly different among teachers except their gain the changing nature of models (CNM) scale scores. Although we

used the HLM analysis method to see the effect of teacher level and the effect of student level variables on student gain CNM scores, we didn't find any significant predictor to explain student gain CNM scores. Furthermore, HLM analysis on gain CNM scores revealed that a significant variation remained after controlling for the predictor variables, indicating that other variables not accounted may be important factors. This finding should be considered in light of the limitations of this study, which implies that there were other variables in teachers to predict students' gain CNM scores. Future work should include more teacher level variables such as types or amounts of their feedback to students.

This is one of the first studies to examine both computer models and formative assessments in chemistry teaching and learning in relation to students' understanding of nature of models. The findings could be the result of several factors, both at the student level and teacher level. At the student level, a factor that may have impacted the effectiveness of the CCFA intervention was the frequency of using computer models. Simply being exposed to computer models might not have a significant effect on students' understanding of the nature of models. Students who had greater exposure in the experimental classes showed more improved understanding of the nature of models in four of the five sub-scales in SUMS. Therefore, the amount of time they spent with the models did significantly impact their understanding of the nature of models. Previous studies found positive results from using computer models to promote chemistry learning (e.g., Ardac and Akaygun, 2004) and to develop their understanding of chemical representations (Wu *et al.*, 2001). The current study contributes to the current body of literature in that engaging students in using computer models also have a positive effect to students' understanding of the nature of models. As Snir *et al.* (2003) pointed out, many students have difficulty in relating scientific models to abstract scientific concepts due in part to a lack of support to engage students in developing correct understanding of models (Duschl *et al.*, 2007). With this regard, we expect that students' better understanding of scientific models will facilitate their understanding of scientific concept as well.

This study also found that students' understanding of models as multiple representations (MR) was not significantly improved after the yearlong intervention. This implies that providing students with different models in chemistry was not enough to improve their MR understanding; they need to be given additional support in understanding the complexity of those models in relation to the concepts they are being taught in class. The use of explicit instruction to promote student understanding of models has been cited by several studies on the effective use of models within classroom instruction (Snir *et al.*, 2003; Coll *et al.*, 2005; Gobert *et al.*, 2011). In particular, a study done by Gobert *et al.* (2011) on the relationship between high school students' understanding of nature of models and the use of model-based software, they found that explicit instruction that was provided for students within the Connected Chemistry curriculum was effective at enhancing student understanding of models. Waight and Gillmeister (2014) also found that student's lack of chemistry content knowledge and knowledge of scientific models made it difficult for students to make meaning of the different representations illustrated in the models in CCFA. This implies that the teacher's role may be more important in enhancing students' understanding of model as multiple representations. In other words, the ability of the participating teachers to understand and carry out the CCFA intervention could have led to the overall non-significant effect on developing student understanding of models as multiple

Paper

representations. This ability may have been impacted by not only the teacher's knowledge of chemistry but also their knowledge of models beyond the traditional ball-and-stick models, most often used within high school chemistry curriculum.

Further, the insignificant result of frequencies of using Flash and NetLogo models on the Gain CNM score from HLM analysis may have been due in part to the participating teachers lack of knowledge regarding the CCFA intervention and as a result it was not fully implemented into their classroom practices. Specifically, in our project, although teachers were provided with formative assessment test results regarding students' understanding of matter, energy and models as well as suggested differentiated labs and activities during the chemistry course, not all teachers consistently followed the suggestions. This indicates that feedback from teachers to students may vary considerably in quality and quantity. This difference in providing feedback might be an important factor in explaining the significant variation remaining after controlling for the predictor variables. Wiliam (2011) asserted the importance of providing feedback enabling to move learnings forward, suggesting that focused, specific, and scaffolded feedback is necessary to improve students' learning outcome. Our classroom observation revealed that there was a wide variation in the extent to which teachers used the data to inform their instruction from no impact on subsequent instruction to some forms of differentiated laboratory experiences for students. The difference in using information from formative assessments results into modifying their instructions to meet students' needs suggests a possible lack of teacher knowledge and skills in implementing formative assessments with feedback. The effectiveness of formative assessment depends on the content of learning activities and quality and quantity of feedback (Black and Wiliam, 1998). Additionally, in order for teachers to effectively implement the intervention into practice, they must have a good understanding of the scientific background on matter and energy underlying the computer models, and to relate the computer models to student learning difficulties on specific chemical concepts (e.g., chemical equilibrium) during the unit of instruction. During the field testing, we found that most teachers did not possess this type of knowledge, and thus providing teachers with more extended professional development will be critical to the implementation of the intervention (Waight and Gillmeister, 2013). In this study, we did not collect teacher level variables such as understanding of models and chemistry concepts, and the degree of modification of their instruction or provided feedback to students except for their years of chemistry teaching and teaching certificates. In light of this limitation, future studies should investigate more variables at the teacher level and their impacts on students' understanding of models.

We acknowledge a few potential uncontrolled threats to internal validity of findings of this study. Potential factors jeopardizing the validity of an experimental study (Campbell and Stanley, 1966) are: (1) History-extraneous event's effect on the dependent variable between pre and posttest; (2) Maturation-subject's natural growth; (3) Testing-effect of taking a pretest on posttest results; (4) Instrumentation-changes in the instrument or scores which may affect on the dependent variable; (5) Selection bias-differences between groups which are not comparable; (6) Statistical regression-due to the selection of subjects on the basis of extremely low or high scores; and (7) Experimental mortality-the loss of participants. In the current study, we used a pretest-posttest control-experimental group design in order to control several potential threats. More specifically, (1) history and (2)

maturation were addressed by using a control group, (3) testing was addressed by using a long time period between pretest and posttest (1 year), and (4) instrumentation was addressed by using a same published test with comparable reliability coefficients for scales (Campbell and Stanley, 1966; Parker, 1990; Tayler and Asmundson, 2008). However, we note that (5) selection bias, (6) statistical regression, and (7) experimental mortality might be potential threats to internal validity. Those three potential threats could be addressed by randomization of group membership, and providing rewards to groups to prevent attrition (Parker, 1990; Tayler and Asmundson, 2008), which the current study did not apply into our study design. The changes in students' understanding of models might be due to the possibility that the students in the experimental groups altered their behaviour because they were aware of being studied (Adair, 1984). Another possibility for the change might be because they were highly enthusiastic about the new computer simulations, however a novelty effect doesn't last long, rather it dissipates significantly in longer duration studies (Kulik *et al.*, 1983), which implies that this study might be less likely to be affected by the novelty effect.

In conclusion, the findings of this study reveal that the frequency of using different types of computer models had a positive influence on students' understanding of the nature of models. The findings also suggest additional efforts are required in regards to helping teachers develop their understanding of how to effectively use models and formative assessments in their instruction. This finding contributes to the literature by identifying several key factors for computer models as formative assessment to improve student understanding of models, including explicit instruction of models to students and extended teacher professional development with a focus on developing their understanding of models and how to use models to modify their instruction. It is not enough for teachers to be given access to new scientific models, further assistance to teachers in the implementation process is necessary in order for students to develop a better understanding of the nature of models in science and to be able to utilize models effectively to explain and predict phenomena. Moreover, as the importance of formative assessment has been emphasized to improve students' learning, it is necessary to enhance teachers' understanding and skills in incorporating formative assessment involving feedback to students into their instructions.

Acknowledgements

The materials reported in this paper are based upon the work supported by the National Science Foundation under Grant No. DRL-0918295.

Notes and references

- Adadan E., Trundle K.C. and Irving K.E., (2010), Exploring grade 11 students' conceptual pathways of the particulate nature of matter in the context of multirepresentational instruction. *J. Res. Sci. Teach.*, **47**(8), 1004–1035.
- Adair, J. G., (1984), The Hawthorne effect: a reconsideration of the methodological artifact. *J. Appl. Psychol.*, **69**(2), 334–345.
- Ardac D. and Akaygun S., (2004), Effectiveness of multimedia-based instruction that emphasizes molecular representation on

- students' understanding of chemical change, *J. Res. Sci. Teach.*, **41**(4), 317-337.
- Barak M. and Dori Y. J., (2005), Enhancing undergraduate students' chemistry understanding through project-based learning in an IT environment, *Sci. Educ.*, **89**(1), 117-139.
- Black P. J. and Wiliam D., (1998), Assessment and classroom learning, *Assess. Educ. Prin. Pol. Pract.*, **5**(1), 7-74.
- Briggs D. C., Alonzo A. C., Schwab C. and Wilson M., (2006), Diagnostic assessment with ordered multiple-choice items, *Educ. Asses.*, **11**(1), 33-63.
- Campbell, D. and Stanley, J. (1966). Experimental and quasi-experimental designs for research. Chicago: Rand McNally.
- Chamizo J. A., (2013), A new definition of models and modeling in chemistry's teaching. *Sci. & Educ.*, **22**(7), 1613-1632.
- Coll R. K., France B. and Taylor I., (2005), The role of models and analogies in science education: Implication from research, *Int. J. Sci. Educ.*, **27**(2), 183-198.
- Cook M. P., (2006), Visual representations in science education: The influence of prior knowledge and cognitive load theory on instructional design principles, *Sci. Educ.*, **90**(6), 1073-1091.
- Dori Y. J. and Hameiri M., (2003), Multidimensional analysis system for quantitative chemistry problems: Symbol, macro, micro, and process aspects, *J. Res. Sci. Teach.*, **40**(3), 278-302.
- Duschl R. A., Schweingruber H. A. and Shouse A. W., (Eds), (2007), *Taking science to school: Learning and teaching science in grades K-8*, Washington, DC, USA: National Academy Press.
- Gilbert J. K. and Boulter C. J., (Eds.), (2000), *Developing models in science education*, Dordrecht, Netherlands: Kluwer Academic Publishers.
- Gobert J., O'Dwyer L., Horwitz P., Buckley B. C., Tal Levy S. and Wilensky U., (2011), Examining the relationship between students' understanding of the nature of models and conceptual learning in biology, physics, and chemistry, *Int. J. Sci. Educ.*, **33**(5), 653-684.
- Kulik, J., Bangert, R., & Williams, G., (1983). Effects of computer-based teaching on secondary school students. *J. Educ. Psychol.*, **75**(1), 19-26.
- Liu X., (2007), Growth in students' understanding of matter during an academic year and from elementary through high school, *J. Chem. Educ.*, **84**(11), 1853-1856.
- Liu, L. and Hmelo-Silver C.E., (2009), Promoting complex systems learning through the use of conceptual representations in hypermedia. *J. Res. Sci. Teach.*, **46**(9), 1023-1040.
- Parker, R. M., (1990). Power, control, and validity in research. *J. Learn. Disabil.*, **23**(10), 613-620.
- Raudenbush S. W., Bryk A. S. and Congdon R., (2010), *HLM 7: Hierarchical linear and nonlinear modelling*, Skokie, IL, USA: Scientific Software International, Inc.
- Schwarz C. V., Reiser B. J., Davis E. A., Kenyon L., Achér A., Fortus D., Schwartz Y., Hug B. and Krajcik J., (2009), Developing a Learning Progression for Scientific Modeling: Making Scientific Modeling Accessible and Meaningful for Learners, *J. Res. Sci. Teach.*, **46**(6), 632-654.
- Smetana L. K. and Bell R. L., (2012), Computer simulations to support science instruction and learning: A critical review of the literature, *Int. J. Sci. Educ.*, **34**(9), 1337-1370.
- Snir J., Smith C. L. and Raz G., (2003), Linking phenomena with competing underlying models: A software tool for introducing students to the particulate model, *Sci. Educ.*, **87**(6), 794-830.
- Taylor, S. and Asmundson, G. J. G., (2008), Internal and external validity in clinical research. In D. McKay (Ed.), *Handbook of research methods in abnormal and clinical psychology* (pp. 23-34). Thousand Oaks, CA: Sage. DOI: 10.1039/C7RP00018A
- Treagust D. F., Chittleborough G. and Mamiala T. L., (2002), Students' understanding of roles of scientific models in learning science, *Int. J. Sci. Educ.*, **24**(4), 357-368.
- Waight N. and Gillmeister K., (2014), Teachers and students' conceptions of computer-based models in the context of high school chemistry: elicitation at the pre-intervention stage. *Res. Sci. Educ.*, **44**(2), 335-361.
- Wiliam D., (2011), *Embedded formative assessment*, Bloomington, IN, USA: Solution Tree Press.
- Wilensky, U., (1999), *NetLogo, Center for Connected Learning and Computer-Based Modeling*; Northwestern University: Evanston, IL, Retrieved from <http://ccl.northwestern.edu/netlogo> (accessed Dec 2016).
- Wolman H., Feldstain A., MacKay C. and Rocchi M., (2012), An introduction to hierarchical linear modelling, *Tutor Quant Methods Psychol.*, **8**(1), 52-69.
- Wu H., 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.