The ChemCollective Digital Library

David Yaron[*], Michael Karabinos[*], Karen Evans[±], Jordi Cuadros[‡], Jodi Davenport[ß], Gaea Leinhardt[†] and James G. Greeno[†]

[*]Department of Chemistry, Carnegie Mellon University, Pittsburgh PA 15213
[ß]Department of Psychology, Carnegie Mellon University, Pittsburgh, PA 15213
[†]Learning Research and Development Center, University of Pittsburgh, Pittsburgh PA 15213
[±]Department of Education, Mercyhurst College, Erie PA 16546
[‡]Institut Químic de Sarrià, Via Augusta 390, 08017 Barcelona

The ChemCollective (**http://www.chemcollective.org/**)[1]. is a digital library of activities for introductory college and high school chemistry. The collection includes virtual labs that allow students to design and carry out their own experiments, scenario based learning activities that invite students to apply their knowledge to real world situations, and interactive tutorials that support student problem solving. A central goal in creating this library is to allow community contributions that can interatively improve the materials. One way we have gone about meeting this goal has been to build in support for asynchronous contributions from chemical educators, learning technologists, and learning scientists.

1. Introduction

The ChemCollective is a digital library of online activities for general chemistry instruction that engages students in more authentic problem solving activities than those found in current textbook problems[2-7]. This is a focused digital library collection that is meant to function as a subportion of the larger national libraries embodied in the ChemDL and NSDL (National Science Digital Library). Our goal is to pursue an expanded vision of a library collection, namely, that of supporting the chemistry educational community in a way that allows us to work together to address some of the educational challenges involved in teaching and learning introductory chemistry.

A primary anticipated benefit is that such a collection can redress one of the major challenges facing educational development projects, that of disseminating educational innovations beyond the research group that developed them. Our premise was that the solution to this dissemination challenge lies not in merely developing new strategies for marketing and delivery. Rather, the "dissemination challenge" is a symptom of a flaw in current development practices that treat materials development and dissemination as a two-step process. This project uses digital library structures to shorten the timescale usually associated with the develop-assess-disseminate-modify cycle by engaging the community in the development process itself. The role of the digital library collection is then not only to make materials available to instructors, but also to empower them to adapt existing materials to their local needs and to create their own materials. A longer term goal is to gather information that can be used to understand how the materials impact student learning.

The materials in the collection fall into the following three broad categories:

- **Virtual labs** and java applets that bridge the procedural knowledge of the course with authentic chemistry by, for instance, allowing students to perform experiments that test the results of their calculations or to use their procedural knowledge to design and carry out experiments.
- **Scenario based learning activities** that embed such authentic chemistry activities in real world contexts that highlight the utility of chemistry to bigger problems in everyday science or the broader scientific enterprise. Examples include our Mixed Reception murder mystery activity and our Ozone activity for kinetics.
- **Tutorials** that combine instruction on key concepts with practice problems that are scaffolded through hints and feedback on student responses. This is a relatively recent addition to the collection prompted by instructor requests for materials that can bring students up to the point where they can engage in the more challenging activities of the above two types. Many of these tutorials were created for use in the Open Learning Initiative (OLI) chemistry course (**http://www.cmu.edu/oli/courses/chemistry/**)[8].

Our focus on problem solving activities is motivated by two main considerations. The first consideration relates to dissemination. Most instructors feel personal ownership of their lectures, and physical labs are difficult to modify due to practical and economic constraints. However, instructors typically assign textbook problems as homework. A collection of online homework that substitutes for part of these textbook assignments therefore provides a viable strategy for shifting undergraduate chemical education, a large system with considerable inertia, towards an improved instructional approach. The second consideration is that much of the learning in undergraduate courses occurs in problem solving activities that take place outside of the classroom. Improved problem solving materials can therefore significantly impact students' experiences in the course.

2. Support for collaboration between community members

Creation of effective online educational materials requires contributions from at least the following three types of expertise:

- **Learning technologist**: We will refer to this contribution as **programming**, although it includes any contribution (databasing, networking, etc.) that goes beyond the expertise required to use an office suite (such as Microsoft office/Frontpage) to author simple web materials.
- **Content expert (instructor)**: By this we mean expertise in both the subject matter and the classroom environment.
- **Learning scientist**: By this we mean expertise in pedagogy, cognition, and assessment.

Although these areas of expertise do not necessarily reside in different individuals, it is rare to find individuals possessing all three types of expertise. Projects such as VANTH (**http://www.vanth.org/**)[9] and Escot (**http://www.escot.org/**)[10] have pursued models of synchronous collaboration by forming development teams consisting of the above three types of experts. Our goal was to enable a collaborative process that is spatially remote and temporally asynchronous (loose coupling), but that continues to bring the above three areas of expertise to bear on materials development.

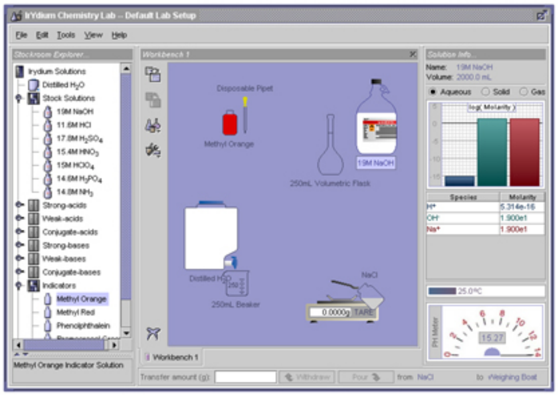
2.a. Learning technologist

The primary means through which we enable the contribution of learning technology is through an approach we refer to as "Configuration-as-Authoring", which successfully engages nonprogrammers in the construction of active educational content. In Configuration-as-Authoring, an educational technologists contributes a highly-configurable software program (virtual lab, simulation, multimedia presentation tool, etc.) to the community. Community members without that level of technological expertise may then create sophisticated learning activities, for instance by configuring the ChemCollective virtual lab for a particular type of experiment. The simulation writes a configuration file which students can then load to engage in the activity. Other examples of such an approach are Interactive Physics 2000 (**http://www.design-simulation.com/IP/**)[11] and the AgentSheets tool (**http://www.agentsheets.com/**)[12]. Sharing of Mathematica and Maple worksheets may be taken as a very advanced example, in which the software tool is very highly configurable.

2.a.i. Virtual labs

The ChemCollective virtual lab (**http://www.chemcollective.org/vlab**)[13] provides a simulation of solution chemistry which can be configured by specifying:

- **Chemical species and reactions**. This allows authors to add new chemical to the lab by specifying their thermodynamic properties (heats of formation and standard entropies) and the chemical reactions in which these species participate. These can include fictional materials, such as acids with randomly generated dissociated constants (for unknown identification activities) or idealized biological molecules such as proteins and drugs (for use in activities that are set in biological contexts).
- **Solutions**. The virtual lab stockroom can be configured to contain any desired solutions. This process is similar to setting out starting materials for a physical lab.



Species	Molarity
H+	5.314e-16
OH-	1.900e1
Na+	1.900e1

Figure 1. Screenshot of the Virtual lab

- **Viewers and instruments.** This allows the author to control access to instruments, such as the pH meter and thermometer, and viewers, such as the list of all chemical species in a solution and their concentrations. For instance, an unknown-acid activity would necessarily need to turn off the chemical species viewer. Access to the pH meter could be provided, or withheld if the author wants students to use a pH indicator such as phenolphthalein.
- **Transfer bar**. The virtual lab provides three means by which students can transfer chemical solutions between containers. The first is "precise transfer" in which students type in a specific volume. This is useful if the goal is to focus student attention on the chemistry without paying attention to experimental technique. The second is "realistic transfer" in which

the volume transferred depends on how long an onscreen button is depressed. This is calibrated such that the attainable accuracy reflects that possible in a real lab. This requires students to use reasonable glassware, such as a buret for a titration. The third is "significant figures transfer", which is similar to precise transfer except that students must enter the desired volume using the number of significant figures possible with the given glassware. This mode of transfer was suggested by a community member and assessment by that community member indicated that this mode was substantially more effective in teaching the concept of significant figures than any of that university's extensive previous attempts.

- **Activity Description**. An HTML description of the activity can be included, which the student can view from directly inside the virtual lab.

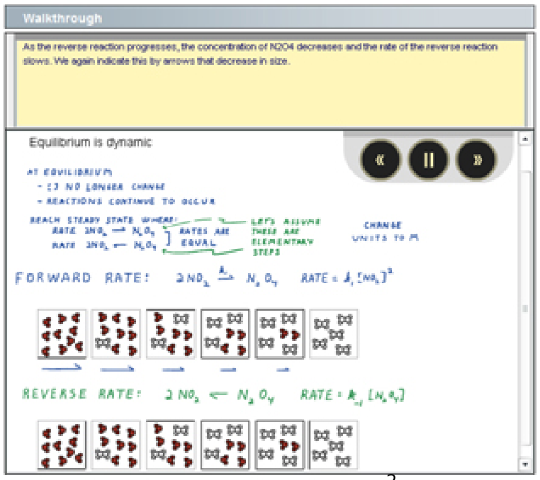
The above functionality provides authors with considerable flexibility in the design of virtual lab activities. The configuration is specified in an XML file. The XML file can be altered with any text editor, however, one must adhere to a fairly rigid set of formatting rules. To make it easier to configure the lab, we also created a virtual lab authoring tool that provides a graphical tool for configuring the virtual lab. This authoring tool saves the configuration to the required XML format. We have found, however, that the XML file is sufficiently easy to edit directly that most users either opt to do so, or simply send us all the required information and we create the configuration file.

2.a.ii. Tutorials

We have also been working on Configuration-as-Authoring approaches to creation of tutorials. Our tutorials consists of online explanations with embedded assessments. A number of tools exist for authoring and delivering assessments, including course management systems such as Blackboard[14] (which store assessments in XML files based on the IMS QTI specification, **http://www.imsglobal.org/question/**)[15]. To allow a more flexible approach to providing students with hints and feedback, we have used both javascript and assessment tools developed by the OLI project. Such tools are continuing to evolve both in terms of sophistication (see, for instance, the Cognitive Tutor Authoring Tools at **http://ctat.pact.cs.cmu.edu/**)[16] and ease of use (see, for instance, **http://www.blackboard.com/**)[14].

Similarly useful tools for creation of online explanations are, however, currently lacking. Video captures of lectures are one option, and with the success of projects such as OpenCourseWare (OCW) at MIT (**http://ocw.mit.edu/**)[17], such captures seem likely to proliferate on the web. While not without merit, such video captures have a number of potentially serious drawbacks including high bandwidth demands and a content format that is difficult. The difficulty of modifications makes it difficult for the community to participation in the evolution and improvement of the materials over time. At the other extreme of development time are materials with high production value, for instance, materials created using flash. Our stoichiometry tutorials were created directly in flash, through a time consuming and tedious development process. Such presentations are also difficult to modify, because the time that goes into the original production can be huge, and it is difficult to get the authors motivated to invest even more time into making large changes.

The EX^2 system we are currently developing is meant to provide a more convenient means to create and modify online instructional explanations. The system is designed to reflect the way most instructors deliver explanations during a lecture. During a lecture, the instructor writes items on the blackboard and discusses these items verbally. In EX^2, the items written on the blackboard appear instead in a scrollable computer window, and so become part of a permanent visible record of the lecture. We refer to these permanently-visible items as "objects". The spoken text associated with these items is referred to as "annotations" on the objects. EX^2 allows annotations be either audio or text. In either case, the annotation is present only while the associated object is in focus. The annotations are transient and do not become part of the permanent visual record, although they can be assessed at any time by using a mouse to click on the relevant object. An underlying assumption of EX^2 is that the transient character of annotations is a key distinguishing features of lectures. The items written on a blackboard during a lecture capture the overall flow of the explanation, summarizing the big picture of the argument. Transient annotations (the speaking part of a lecture) allow the instructor to give detailed motivations and justifications for each step without losing the overall flow of the discussion, since a concise record is kept visible on the blackboard. Textbooks and static web pages do not have


Walkthrough

As the reverse reaction progresses, the concentration of N2O4 decreases and the rate of the reverse reaction slows. We again indicate this by arrows that decrease in size.

Equilibrium is dynamic

Figure 2. Screenshot of the EX^2 Player

an equivalent means to separate big-picture flow of the explanation from supporting details. Our working hypothesis is that this distinction is one of the primary reasons many instructors and students find the lecture format more appealing than equivalent written documents.

The EX^2 system supports online explanations that couple permanently-visible objects with transient annotations. The author writes an object on a tablet computer, and annotates it with what they would normally say in the classroom (in audio, text, or both). The content can be view in multiple modes. In lecture mode (currently available), the objects appear in sequence, with annotations being visible (or audible) for only the most recent object. Review mode (not yet available) will show all objects at once and allow one to click to see or hear desired annotations. Review mode is analogous to walking into a lecture hall after class, touching an object on the blackboard, and hearing what the instructor said at the time they wrote that object down.

A central design goal of the EX^2 system is to aid not only creation and delivery of online explanations, but also modification of such explanations. Facile modification will allow the content to be refined and improved over time. The outcome of the authoring process is an XML file with a list of objects and associated annotations. The viewer for EX^2 content is written in Adobe Flash, meaning any content type supported by Flash can in principal be included in a lecture. The content is modifiable because one can rearrange the objects, insert new objects, alter the annotations, easily cut and paste existing content into alternative sequences, etc. Our current version of the system was used to create instructional explanations for our equilibrium tutorials. We currently have a first version of an EX^2 viewer, but the content is authored through a process requiring considerable expertise with the XML format. We are hoping to begin development of a graphical authoring tool that will make creation of EX^2 content nearly as efficient as video capture.

2.b. Content expert/instructor

The ChemCollective has been quite successful in engaging instructors as authors of new materials which they then contribute to the collection. Of the 117 virtual labs in the collection, 56 were contributed by 11 different user groups. Many of these contributions were created in collaboration with ChemCollective staff. Instructors submitted ideas for their activity, and these were implemented locally. In addition, use of the authoring tool described above is growing. The community contributions have substantially increased the diversity of the collection, since many of the topic areas and approaches are outside that which we would have developed on our own.

2.c. Learning scientist

Learning science contributed to both the design and evaluation of the collection and its contents. Much of the development of virtual laboratories, scenarios and tutorials for the ChemCollective was done with strong, but local, collaboration between developers and learning scientists. This effort was organized through biweekly meetings involving in depth discussions of design issues such as learning goals and level and types of feedback. This team also designed and carried out the assessment efforts (see next section) which involved extensive observation of student problem solving and analysis of artifacts from student assignments. These student observations fed back extensively into the modification of existing materials and the choice and design of new materials.

Learning science also contributed technical approaches to the collection of data for learning assessment. The virtual lab is now instrumented to save all student actions to a log file. We are working with the Pittsburgh Science of Learning Center (**http://www.learnlab.org/**)[18] on the challenging task of extracting useful information regarding use and learning from such log files.

3. Assessment

3.a. Learning assessment

Assessment of usability aspects of the virtual lab has been ongoing since the earliest versions. It takes about 5 minutes for a student to become sufficiently familiar with the software that their attention shifts almost exclusively to the chemistry goals. Our current instructors find the lab easy to implement in their classrooms, although this may reflect some bias in the sample of instructors who have used the lab and respond to our surveys. We also have student surveys from multiple sites that indicate students find the lab to be useful for their learning (typically 3.5 on a scale of 5), but feel the need for more direction while using the laboratory (also typically 3.5 on a scale of 5).

Throughout the development we have also routinely observed student problem solving (mostly at Carnegie Mellon, but also with students at Florida Atlantic University, University of British Columbia and Pittsburgh's Central Catholic High School). Robert Belford has also reported back extensively on student experiences at University of Arkansas at Little Rock.

Classroom observations of students performing online experiments led to a particularly interesting finding. Initially, our instructional goal for these online experiments was to embed the procedural knowledge of the course in a context that highlights its utility, such that students learn not only how to do a procedure but also when to do it. Our observations, however, shifted our instructional perspective to one of helping students move beyond shallow problem solving strategies. In particular, we noticed that students employ a potentially superficial strategy in which they analyze the language in a word problem to discover the given and requested quantities and then search for equations that connect the given to the requested quantitities. For instance, a calorimetry text problem may give a measured change in temperature (ΔT given) and ask for a heat (q requested). A student, having first written down the ΔT and q from the problem statement will then sift through the equations in the current textbook chapter and identify $q = m*Cp*\Delta T$ as a connecting equation. While this is a potentially useful skill, it does not give students practice with the fundamental concepts underlying calorimetry. An activity that requires deeper reflection is our virtual lab that requires students to design an experiment to measure the enthalpy of a reaction. The text-matching strategy fails in experimental design and the student must instead realize that the equation $q = m*Cp*\Delta T$ represents an experiment in which a temperature change is used to measure heat. Our observations show that students find the experimental design problem considerably more difficult than the text problem, and often have as a first question "what equation do I use?". These observations suggest that online experiments promote additional conceptual learning.

A more formal assessment effort occurred in the Spring 2004 semester at Carnegie Mellon, in which we collected extensive data regarding the effects of the use of Virtual Lab and scenario-based learning activities on students' understanding of basic chemistry concepts. The semester was divided into three 5-week segments, corresponding to the course hour exams. Each segment had used situated activities: (i) online materials used only during recitation periods (1 hour time limit on activities, with help from human tutors), (ii) traditional paper-and-pencil homework activities and (iii) online homework (no time limit, with access to human tutors through office hours and email). All work handed in by students was saved, including recitation sheets, homework assignments, quizzes and exams. Student attitude surveys were given before each of the three exams, and included items that measure student views of the activities and their confidence with the various course concepts. Unannounced pre-tests were given in recitation the week before scheduled exams in the second and third sessions to assess students' understanding of concepts. These reflect the learning that took place through the activities prior to studying for the exam. The virtual lab was also instrumented to collect a trace of student interaction, and this data is available for the entire course. The principle findings are[7]:

a. A significant portion of the learning takes place during self-directed study the last few days before the exams. This came from comparison of student performance on the practice exams (~30%) versus the actual exams giving about 5 days later (~80%).
b. Self-directed study and homework, are the most relevant learning opportunities, explaining the above finding. A structural equation model (see figure 1) was developed that could account for 48% of the variance in student performance on the final exam. In this structural equation, homework and self-directed study are the two main contributors and have equal weights.
c. Authentic problem-solving activities have an important mediating effect in learning. The structural equation shows the influence of homework on performance in the corresponding topic areas on the exams.
d. Study and carefully planned homework activities can overcome the initial differences in prior knowledge.
e. Although homework has a strong influence on exam performance, this effect is not present in the practice exams. Apparently, the benefits of the homework show up only after a period of self-study between the practice and actual exams. In this context, we note that the homework used here was intentionally designed to provide experiences and modes of practice (real-world contexts and use of virtual labs) that complement, but are distinct from, more traditional problem solving such as that on the exams. The influence, on more traditional problems, of the knowledge gained from the homework may be apparent only after students have done considerable self-study with traditional problems.

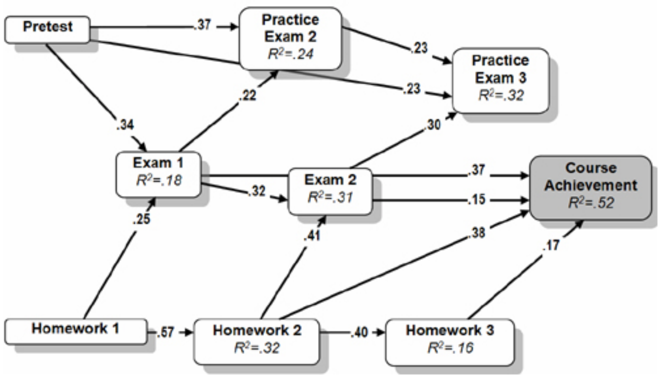


Figure 3. Structural equation model from Cuadros, Leinhardt and Yaron.[7]

Most recently, we conducted a well-controlled study of students enrolled in an online course (**http://www.cmu.edu/oli/courses/chemistry/**)[8, 19] that uses ChemCollective materials. The results indicate that students who engage in virtual lab activities show a solid and direct relationship between the number of events of engagement with the virtual lab and learning as measured on a post test and controlling for variances in SAT scores.

3.b. Collection assessment

We have chosen not to require users to identify themselves to gain access to the materials. While this may maximize use of the materials, it has the down side of making it challenging to monitor who is using the collection and how. To better understand our dissemination, we developed a strategy of attempting to estimate three factors: exposure to, use of and participation in the community. The strategy and results, up to 2005, are contained in a report entitled "The ChemCollective: Monitoring the Path from Seeing to Using to Contributing"[20]. The goal of the analysis was to estimate what percent of our target audience has been exposed to the ChemCollective materials (see), how many use those materials (use), and how many contribute materials (contribute). The results of the report, with updated numbers, are briefly summarized here.

> **Target Audience**: We consider the primary target audience of ChemCollective materials to be the teachers of introductory chemistry. Based on information from the U.S. Department of labor, we estimated 100,000 high school chemistry teachers and 9,000 college introductory chemistry instructors.
>
> **See**: Our estimate for exposure to the materials is based on number of CDROM's distributed at conferences, attendees at workshops and talks, readership of magazines and journals that have featured our materials, and number of unique visitors coming from search engines such as google and yahoo. This leads to an estimate of 7000 chemical educators who have been made aware of the materials.
>
> **Use**: Since the collection does not require users to register or log in, and the software can either be run from the web site, downloaded to the local computer, or run from a CD-ROM, we do not have complete usage statistics. From available statistics, we can however make a conservative estimate that 200 classrooms currently make extensive use of our materials.
>
> **Contribute**: Eleven instructors have contributed a total of 56 different activities to the collection. (These contributors work closely with our development team.) 8 groups have made contributed translations of the virtual lab interface along with over 70 activities (Spanish, Portuguese, French, Catalan, Galician, German, Russian, and Lithuanian) with 3 more languages in progress (Traditional Chinese, Polish and Check). We therefore have 22 current contributors. The web page invites instructors to contribute feedback on the use of the materials in their classroom and to participate in assessment studies. In the past four years, 40 instructors have provided feedback (although much of it is general and does not address specific educational issues or goals) and about 13 expressed interest in participating in studies.

The above conservative estimates for our target audience are that 7000 instructors have seen the collection, 200 use activities in their classroom, 22 have contributed activities, and 40 have contributed feedback. These results give insight into the number of users that go from seeing to using a collection (3%), from using to contributing activities (11%), and from using to contributing feedback (20%).

The following table is a brief overview of our web statistics:

	2004	**2005**	**2006**	**2007**	**Total**
ChemCollective Website Unique Visitors	101,397	106,429	123,400	161,481	492,707
Vlab (individual users)					
Access the applet to perform experiment	18,757	48,626	59,733	62,871	189,987
Download the virtual lab to local drive	4,329	6,425	15,678	17,556	43,988

4. Future plans

The ChemCollective remains an active collection, with continuing development of all portions of the collection. We are also working on integrating the collection with the ChemEd DL such that all materials can be discovered and accessed through that more general portal.

5. Acknowledgements

We gratefully acknowledge support for this work from the National Science Foundation (DUE-0443041 and DUE-0333720), the William and Flora Hewlett Foundation and the Howard Hughes Medical Institute.

6. References

1. Yaron, D. et al. The ChemCollective. *http://www.chemcollective.org/* (accessed Mar 2008).
2. Yaron, D.; Freeland, R.; Lange, D.; Milton, D. Using simulations to transform the nature of chemistry homework. Presented at *CONFCHEM: Use of Computer Simulations in General Chemistry*, [Online] **2000**, Summer.
3. Yaron, D.; Milton, J.; Karabinos, M. A Digital Library that Promotes Creation and Use of Modifiable Student Activities. In *Developing Digital Libraries for K-12 Education*; Mardis, M., Ed; ERIC IT Clearinghouse: Washington, DC, 2003.
4. Yaron, D.; Evans, K. L.; Karabinos, M. Scenes and Labs Supporting Online Chemistry. Presented at 83rd Annual AERA National Conference, Chicago, Il, April 2003.
5. Yaron, D.; Cuadros, J.; Karabinos, M.; Leinhardt, G.; Evans, K. L. Virtual Laboratories and Scenes to Support Chemistry Instruction. In *About Invention and Impact: Building Excellence in Undergraduate STEM (Science, Technology, Engineering, and Mathematics) Education, (Proceedings from the National Science Foundation Course, Curriculum and Laboratory Improvement (NSF-CCLI) program conference, 2004)*; Cunningham, S., George, Y. Eds; American Association for the Advancement of Science: Washington, DC, 2005.
6. Evans, K. L.; Leinhardt, G.; Karabinos, M.; Yaron, D. Chemistry in the field and chemistry in the classroom: A cognitive disconnect?. *J. Chem. Educ.* **2006**, 83, 655.
7. Cuadros, J.; Leinhardt, G.; Yaron, D. One firm spot: the role of homework as lever in acquiring conceptual and performance competence in college chemistry. *J. Chem. Educ.* **2007**, 84, 1047.
8. Open Learning Initiative Free Online Courses - Chemistry. *http://www.cmu.edu/oli/courses/chemistry/* (accessed Mar 2008).
9. Vanderbilt-Northwestern-Texas- Harvard/MIT Engineering Research Center. *http://www.vanth.org* (accessed Mar 2008).
10. Roschelle, J.; DiGiano, C.; Koutlis, M.; Repenning, A.; Phillips, J.; Suthers, D. Developing educational software components. *Computer* **1999**, 32:9, 50-8.
11. Design Simulation Technologies. Interactive Physics. *http://www.design-simulation.com/IP/* (accessed Mar 2008).
12. Cherry, G.; Ioannidou, A et al. Simulations for Lifelong Learning. In *Spotlight on the Future, NECC '99. 20th National Educational Computing Conference Proceedings*, Atlantic City, NJ.
13. Yaron, D. et al. The ChemCollective Virtual Laboratory. *http://www.chemcollective.org/vlab* (accessed Mar 2008).
14. Blackboard. *http://www.blackboard.com/us/* (accessed Mar 2008).
15. IMS Global Learning Consortium: IMS Question & Test Interoperability Specification. *http://www.imsglobal.org/question/* (accessed Mar 2008).
16. Cognitive Tutor Authoring Tools. *http://ctat.pact.cs.cmu.edu/* (accessed Mar 2008).
17. Free Online Course Materials | MIT OpenCourseWare. *http://ocw.mit.edu/* (accessed Mar 2008).
18. Pittsburgh Science of Learning Center. *http://www.learnlab.org/* (accessed Mar 2008).
19. Evans, K. L.; Leinhardt, G.; Yaron, D. Learning stoichiometry: A comparison of text and multimedia formats. *Chem. Educ. Res. Pract.* **2008**, 9, 208 - 218.
20. Yaron, D.; Cuadros, J.; Leinhardt, G.; Rehm, E.; Karabinos, M.; Paluka, T. The ChemCollective: Monitoring the Path from Seeing to Using to Contributing". *Internal project report*, 2005. Note: available at **http://www.chemcollective.org/pdf/papers/monitorpath.pdf**