Kendra Seniow, Oregon State University
Kendra Seniow is an undergraduate student in Chemical Engineering, the University Honors College, and the International Studies Programs at Oregon State University. In pursuit of her University Honors College and International degrees, she is investigating student teams’ use of models during completion of the BioReactor Virtual Laboratory project, how similar educational principles can be applied to international development projects and how participation in both these authentic activities helps develop stronger engineers.

Eric Nefcy, Oregon State University
Erick Nefcy is a MS student in the School of Chemical, Biological and Environmental Engineering at Oregon State University. Through his undergraduate studies, he has held multiple internships at Intel Corporation. He is currently studying the growth of self forming barrier layers in copper thin films as well as investigating the student teams’ use of models during completion of the Chemical Vapor Deposition Virtual Laboratory project.

Christine Kelly, Oregon State University
Christine Kelly is an Associate Professor of Chemical Engineering at Oregon State University. She currently has research activity in biological production and use of ligninolytic enzymes in biomass deconstruction for bioproducts and biofuels manufacture. She also has interests in developing tools to promote the use of higher cognitive skills in engineering coursework. Dr. Kelly recently earned OSU’s Austin-Paul Engineering Faculty Award for mentoring students.

Milo Koretsky, Oregon State University
Milo Koretsky is an Associate Professor of Chemical Engineering at Oregon State University. He currently has research activity in areas related to thin film materials processing and engineering education. He is interested in integrating technology into effective educational practices and in promoting the use of higher level cognitive skills in engineering problem solving. Dr. Koretsky is a six-time Intel Faculty Fellow and has won awards for his work in engineering education at the university and national levels.
Representations of Student Model Development in Virtual Laboratories based on a Cognitive Apprenticeship Instructional Design

Abstract

The undergraduate laboratory plays a pivotal role in science and engineering curricula. However, traditional physical laboratories are resource intensive, and due to these constraints, do not always achieve their diverse set of intended learning outcomes. One way to overcome these limitations is to use alternative modes of delivery, such as virtual or remote laboratories. In a virtual laboratory, students do not interact with real equipment to obtain data, but rather with computer simulations of laboratory or industrial process equipment that produce results that can be obscured by pre-programmed statistical variation. In the most common approach, the virtual laboratory is used as an alternative mode and simulates a similar set of activities as in the corresponding physical laboratory. In a few cases, virtual laboratories have been used to create learning activities with no analog to the university instructional laboratory. The instructional and software design of the virtual laboratories described in this study falls into the latter case and is based on the situated context of a practicing engineer in industry. The cognitive apprenticeship approach used is structured around the task of having students determine the operating parameters for chemical processes for volume production through experimental design, interpretation and iteration. In this sense, the virtual laboratory project simulates what expert engineers do in practice, and ends up looking very different than the physical laboratory at the university.

In this paper, a method is presented to characterize student groups’ model development as they proceed through the task situated in the cyber-environment of the virtual laboratory. Data sources include laboratory journals, an initial design memorandum, the final written and oral reports, and experimental records available through the instructor interface. Classifications in the graphical representation of model development include: nature of the model component (quantitative, qualitative, empirical, statistical), utility of the model component (operationalized, modified, abandoned), correctness of the model component, action based on the model component (did it direct the values of input variables for a future run, was a run used to quantify model parameters, was the model qualitatively verified, etc.), and emotional responses to model verification or mismatch. Preliminary results from this model representation scheme are presented for two industrial scale virtual laboratories, one based on a transient biological system and one based on a steady-state chemical system. Different types of qualitative and quantitative models are evident in the students’ solutions and can be generally related to differences in the type of knowledge structures of the physical systems embodied by each of the virtual laboratories. Student groups also show distinct differences in ability to apply schematic and strategic knowledge, and strength in one knowledge type does not necessarily indicate strength in the other. This study is part of a larger project to compare and contrast the nature of learning elicited in the virtual laboratory experience.
Introduction

Engineers, students and professionals, invoke a wide variation in content and complexity of models as they complete open-ended, ill-structured tasks. In this context, a model is defined as the representation of a phenomenon, any intellectually interesting segregation of the world, used to build understanding and permit inquiry.\textsuperscript{1} This study examines student model development in two industrially situated virtual laboratories. By simulating the physical components of the laboratory, the emphasis of student learning shifts to outcomes that have been historically elusive in the academic environment: experimental design, analysis, redesign and iteration. Rather than emphasizing procedural or routine tasks, the virtual laboratory presents students an ill-structured and dynamic task. While engaging in these virtual laboratories, the students invoke and operationalize models. This paper describes a methodology to analyze model development and use during a virtual laboratory project.

Two virtual laboratories are studied: a Virtual Chemical Vapor Deposition (VCVD) Laboratory and a Virtual Bioreactor (VBioR) Laboratory. The VCVD Laboratory simulates an industrial-scaled vertical chemical vapor deposition reactor in which silicon nitride is deposited from dichlorosilane and ammonia gases at low pressure and high temperature. Students are tasked with achieving maximum thickness uniformity and minimum dichlorosilane utilization by adjusting operating parameters including gas feed rates, temperatures of five reactor zones, system pressure and duration of operation. The VBioR Laboratory is based on an industrial stirred-tank fed-batch bioreactor, and can be used for different functions, such as production of a recombinant protein, or degradation of waste, and run in either batch or fed-batch mode. Students aim to achieve maximum volumetric productivity by varying input parameters such as temperature, substrate concentrations, cultivation times, and feed flow rate. Random process and measurement variation is added to the output of both virtual laboratories. The details of the VCVD and VBioR Laboratories, their instructional design, and protocol analysis of student learning and student perception in the virtual laboratories have been previously published.\textsuperscript{2,3,4}

Though centered in different domains, both the Virtual CVD and Virtual Bioreactor have at their root reaction kinetics and material balances. The VCVD system can be described using a simulation with transient solid phase accumulation, and a pseudo-steady state gas phase. The VBioR is an inherently transient system, with cell growth, substrate consumption and product synthesis and degradation occurring throughout the cultivation. Both scenarios present an adequate challenge to students while eliciting the use of engineering principles and models and schematics of the simulated systems are indicated in Figure 1.

Shavelson has described scientific achievement as consisting of four types of knowledge: declarative (“knowing that”), procedural (“knowing how”), schematic (“knowing why”) and strategic (“knowing when, where and how our knowledge applies”).\textsuperscript{5} Though laboratory experiences are meant to draw upon and develop all four types of knowledge, often the physical laboratory at the university relies upon the declarative and procedural aspects of recall of facts and adherence to proper protocol. In the virtual laboratories, however, the physical component is removed and students are able to focus on developing schematic knowledge, by integrating concepts and building models, and strategic knowledge, by intelligently combining these models to tackle an ill-structured and open-ended task. Schematic knowledge includes principles,
schemas and mental models that explain the physical world. Strategic knowledge is demonstrated by determining how and what knowledge applies to a new situation and includes domain-specific conditional knowledge and strategies such as trouble shooting and problem-solving as well as monitoring.\textsuperscript{6,7,8} Effective systems thinking is related to accessing strategic knowledge and includes characteristics such as: seeing the big picture, understanding interconnections, understanding systems synergy, multiple perspectives, tolerance for ambiguity, and understanding analogies between systems.

![Schematic of the equipment simulated by the two virtual laboratories](image)

The virtual laboratories provide student teams dynamic access to data as they choose what runs and the measurements to make in a structure that requires iterative convergence on a solution, which specifically promote and develop students’ use of strategic knowledge. Success is intimately coupled not only to the ability to develop models to analyze and interpret this new information, but also to the ability to identify what information will be useful and how to move closer to an optimal solution. The ambiguous, ill-structured type of problems engineers face in the virtual laboratories and in practice require critical elements of schematic and strategic knowledge, and often involves model development and use to integrate schematic knowledge.

Characterization of the models that students construct, and the model progression is challenging, especially in ill-structured problems. One study has examined the models chemical engineering undergraduates invoke as they are engaged in a troubleshooting process in a simulated chemical plant in which the students played the role of plant operator.\textsuperscript{10} The method used was to compare student if-then responses to a faculty and two graduate student experts. In another study, protocol analysis was used to examine how instructional design experts used prior knowledge and experience to solve ill-structured problems.\textsuperscript{11} Results were correlated according to four expert characteristics. A third study involved developing a flow chart record of video-recorded
solutions of first year university physics students of context rich problem. That study examined how conversations of students’ everyday life experience developed into physics reasoning. The diversity of these approaches investigating model development and use illustrates how elusive and context dependent model characterization is. The goal of the method presented in this paper is to develop an explicit representation of the models the groups develop as they complete an ill-structured problem.

Overall, the instructional design of using the virtual laboratories yields great potential for enhancing students’ learning. One central outcome is the degree and extent to which students can integrate models into their experimental design and project solution. The objective is to have students recognize and use their engineering science background to more effectively pursue the experimental design process rather than being explicitly instructed to use a specific model. Students develop and use a wide spectrum of models in their efforts to create more effective experimental designs. These models range from quantitative predictive analytical models to qualitative descriptive physical models to empirically based statistical models. In addition, there is a range of sophistication in any one of these categories.

In order to capture the model construction and higher cognition and to characterize the schematic and strategic knowledge in the virtual laboratory project, we have developed Model Development and Usage Representations (Model Representations) as an analysis tool. The Model Representations are generated from student work products, such as journals/laboratory notebooks, written reports, and memorandums. They are a visual and chronological coding tool used to identify and characterize student knowledge structures and cognition as students perform the virtual laboratory project. The Model Representations can be used to identify the ways students use their schematic knowledge to build models and use their strategic knowledge to integrate these models into their project solutions. The methodology for developing Model representations is presented in this paper.

**Analysis Methodology**

Model Representations provide a visual representation of student modeling progression during completion of the virtual laboratory with regards to the types of model components employed (quantitative or qualitative, statistical or empirical), their degree of utilization (operationalized, abandoned or not engaged), their correctness, and the experimental runs to which they are relevant. This information is combined along a time line with experimental runs, emotional responses and instructor interaction to show context and form the complete Model Development and Usage Representation. All information is drawn from the instructor interface, which records all groups’ run parameters and results, and from student work products, which include an initial design strategy memorandum, update memorandum, oral presentation, final written report and experimental journal. Due to its chronological orientation, this method forms a linear representation; however, this depiction is not meant to imply that the learning process is direct and linear. A total of 27 Model Representations have been completed for the 2008 cohort in the capstone laboratory course at Oregon State University and examples are presented here.

Student journals serve as the primary source of information for coding since they are intended to contain all references, notes, results and calculations over the course of the project. Model
components are identified from the student journals chronologically and are then supplemented with information from other materials such as reports and memoranda which serve to confirm, explain or expand upon the journal content. Student researchers first individually dissect the work products to construct the preliminary Model Representation. Consensus is then obtained by a group of two students and two faculty. One faculty member, the domain expert in the appropriate field, examines the source material and evaluates the accuracy and context of the Model Representation. An Overview statement is then written which summarizes in a concise manner the group’s approach and integration of model components in their unique solution to this authentic, ill-structured problem. In order to assure reliability, the student and faculty who work in the other domain participate to assure consistency between the Model Representations of the two virtual laboratories.

_Model Components_

Models are identified either explicitly in student work or where students show evidence of a use of a deeper cognitive framework and reasoning in order to make decisions within the project. Once identified, the full mathematical expression or model descriptor is added to the Model Representation. Quantitative Model Components, those characterized by mathematical equations, are placed inside squares while Qualitative Model Components, those characterized by descriptive mechanisms and responses, reside in circles. Additionally, a model component can be designated as either statistical or empirical in nature by an ‘S’ or ‘E’ in the model component box. Figure 2 shows the different types of model components and an example of how they could appear together in the Model Representation.

![Figure 2. Primary and Secondary Model Components.](image)

Since notebooks provide a location for brainstorming as well as for results and discussion they often reflect those ideas which are never fully realized amongst those that are implemented. Additionally, student groups occasionally include in their memoranda, reports or presentations those ideas which they feel _should_ be relevant or correct, as indicated by, among other sources, cues during coursework or simply the apparent scientific sophistication of such components. Those model components discussed, noted or considered by students, but without evidence of
use or engagement in a meaningful way are considered “Not Engaged” and recorded in black text inside a black circle (if qualitative) or square (if quantitative). Any model component that was implemented and used in order to better understand the system and work towards final parameters and is retained throughout the project is considered "operationalized" and written in green. Model components which the group attempts to use, but which they then determine to be incorrect or not applicable to the project or final model, are recorded in red as “abandoned.”

Model components which occupy a central role in student reasoning and project foundation are designated Primary Model Components and placed in the center line of the time line while all other components are Secondary Model Components and drawn as additions by arrows to the center line. These decisions are made by the researchers’ judgments based on the evidence of the work products. Since a model component which is Not Engaged cannot form the foundation or central role of a group’s strategy, Not Engaged components are always secondary and connected to the center line by a dashed arrow. Problem Scoping is given as a unique component; its presence or absence, as well as depth and breadth, may influence the groups modeling approach. When evident in student work, it is indicated as a black box surrounding the words “Problem Scoping” and can include multiple associated literature references. Students have completed the virtual laboratory once they have obtained a set of recommended operating parameters, or “Final Recipe,” therefore, the Model Representation concludes with those words as well.

Experimental Run Markers

In addition to the model components, other aspects of student work are included to provide context. Experimental runs are documented along the time line by a symbol and run number. As shown in Figure 3(a), the symbol used indicates whether the nature of the experiment is "Parameter Defining" (data from this run was used to construct the model component, e.g., obtain a numerical value for a model parameter), "Model Directed" (run parameters were chosen based on model output), or "Qualitative Verification" (the run either confirms or contradicts model-based expectations). Experiments with no apparent relation to modeling are also included with a unique symbol so that all runs are accounted for in the Model Representation. Run numbers are noted alongside model components, such as in Figure 3(b), to show their relevance to either model development or implementation. Additionally, this notation indicates when student teams’ return to previously identified models and concepts and further develop them.

Figure 3. Run Marker Types and Notation. Part (a) lists four motives for conducting runs as they relate to modeling while part (b) provides an example of how runs and model components are integrated into the Model Representation.
**Other Components**

We believe the social interactions in the group play a critical role in determining the project direction. Quotations or summaries of student emotional responses are incorporated into the Model Representation by an outward-pointing, purple arrow. Instructor consultation, whether during meetings or by email, is quoted or summarized and is incorporated as an inward-pointing, blue arrow. When literature references are explicitly stated in student work they are listed (in black) near the relevant component of the Model Representation. It is helpful to distinguish the influence of the instructor or publications from student revelations. Models which are clearly incorrect are included and visible, but crossed out with red. Figure 4 summarizes these components.

![Figure 4. Additional Components of the Model Representation.](image)

**Overview Statement**

The Overview Statement is a holistic interpretation in words of the group’s solution that is intended to complement the sequential information provided visually in the graphical representation. It consists of the essence of a group’s approach and examples of transformations in the group’s understanding, cognitive conflicts or other significant cognitive or modeling events. A characterization of the group’s schematic knowledge and strategic knowledge as either low, neutral or high is included in the Overview Statement.

Model Representations will be used to identify the knowledge structures that students are demonstrating in their solutions of the Virtual Laboratory projects. The succinct yet thorough visual manner in which students’ models and strategies are presented permits the characterization of learning and cognition as well as the comparison of these traits between student teams. In addition, the manner in which domain content between the bioprocessing and silicon processing projects manifests in model development will be elucidated. The complete model representation component key is indicated in Figure 5.
Model Representations have been generated for the Virtual Laboratory projects of students of the senior capstone laboratory course at Oregon State University in Fall 2008. The class consisted of 55 students majoring in chemical engineering, 19 majoring in bioengineering and 6 majoring in environmental engineering, who worked in 27 teams of 2-3 students on either the Virtual CVD Laboratory (41 students) or the Virtual Bioreactor Laboratory (39 students).

Instructional Context

Over the course of the 10 week term students complete three laboratories, each requiring about three weeks. The first and third laboratories are physical laboratories (a heat exchanger and ion exchange) while the second is the virtual laboratory. Students form their own groups of three students (or the rare pair) and choose between the Virtual CVD Laboratory or the Virtual Bioreactor Laboratory, though the second option requires that at least two of the three group members be enrolled in currently or have taken a course on bioreactors (BioE 457). Environmental engineering and bioengineering students are required to complete BioE 457: Bioreactors and generally do so concurrently with the senior laboratory. Many Chemical Engineering students also enroll in this course to satisfy a Bioprocesses Option. An additional faculty member, a domain expert in either silicon processing or bioprocessing, respectively, serves as the lead instructor for each project by introducing the project, acting as the manager.
during update and presentation sessions and evaluating student performance. These same faculty members have served as the domain experts during the coding process presented here.

Student project deliverables include a design strategy memorandum, intermediate update memorandum, final written report, oral presentation and experimental journal. Prior to accessing the virtual laboratory students must submit the design strategy memorandum to the domain expert/manager which describes their strategy and the operating parameters and measurements for at least one run. At this point, the team has a twenty minute coaching session with the “expert” instructor. The following week the team submits an intermediate update memorandum which reflects on their current progress. Both of these memoranda and coaching sessions provide the opportunity for and encourage reflection and discussion by students on their own actions and understanding. Students finish the project by presenting their strategy and findings in both an oral presentation and final written report. These deliverables combined with student run data accessed by means of the instructor interface are the source of information for the generation of Model Development and Usage Representations which help to identify students’ learning and model use.

Sample Model Representations

Four complete Model Representations are included at the end of this paper: two VBioR teams (teams A and B) and two VCVD teams (teams C and D). A detailed explanation of the coding of Team A’s performance of the virtual bioreactor is provided in the following figures. These figures include portions of the Model Representation alongside the corresponding transcribed portions of the students’ experimental journal. Example journal entries have been rewritten to maintain student confidentiality while retaining original wording and formatting as much as possible.

Team A begins by finding bioreactor operating parameters reported in literature. They list these sources and information as shown in Figure 6. This Problem Scoping is not a modeling activity, but is still relevant to their strategy in the project since it can help guide students and is often taught as part of a correct scientific approach. Other teams vary in problem scoping techniques and depth, such as the number literature citations, shown in the journal and reports. Team A initially proposed the use of a Design of Experiments (DOE) to optimize operating parameters, but chose not to implement this idea due to its complexity.

![Figure 6. Problem Scoping](image)

The team lists data from multiple literature sources, though only a small portion is shown here. The originally plan to conduct runs based on a DOE, but decide against this idea due to its complexity.
Team A uses data from a previous run (5) in a material balance relationship they develop to refine the operating parameters of a subsequent run (6). Run 6 is, therefore, model-directed because the parameters come directly from the material balance quantitative model. Figure 7(a) shows the model equation as they wrote it in words while later on the same page (Figure 7(b)) they use that relationship to calculate the operating parameter of feed substrate concentration. This model component is green since it is fully operationalized yet crossed out since it is an incorrect model of bioreactor behavior. It can be seen in the update memorandum that the team consulted the instructor after developing the material balance equation, but prior to implementing it in run 6.

The team later notices that more product is formed while the cell concentration is constant. Though they introduce this idea prior to run 9, they do not confirm it confidently until examining run 9 results, as shown in figure 8. Run 10 confirms their findings to them. There is cell growth in the system, but it is masked by dilution by the feed stream (concentration remains constant, but total cell mass increases). Product formation is not actually limited to the period of no growth, so this component is labeled as Incorrect.
Upon observing that a lower substrate concentration leads to increased growth, Team A concludes that substrate inhibits cell growth. As illustrated in Figure 8(b), they incorporate this model to decide on parameters for the following run which confirms to them this mechanism.

Once Team A believes they have found the operating parameters which yield the most product, they repeat two additional runs with these same parameters in order to account for variation in the final reported volumetric productivity of the bioreactor. Though the journal entry does not explicitly state that they found the average and standard deviation, these values are reported in the Final Report.

Upon observing that a lower substrate concentration leads to increased growth, Team A concludes that substrate inhibits cell growth. As illustrated in Figure 8(b), they incorporate this model to decide on parameters for the following run which confirms to them this mechanism.

Figure 8. Qualitative Model and Verification. In part (a) the team believes that product is only formed when cells are not growing and directs runs 9 and 10 around this model. This is not the case, however, as indicated by the red cross-out. In part (b), the group bases the parameters of run 13 on a mechanism they suspect to be at play based on observation.

Once Team A believes they have found the operating parameters which yield the most product, they repeat two additional runs with these same parameters in order to account for variation in the final reported volumetric productivity of the bioreactor. Though the journal entry does not explicitly state that they found the average and standard deviation, these values are reported in the Final Report.

Figure 9. Statistical Analysis and Final Recipe. This journal entry indicates that runs 15 and 16 were conducted specifically to be able to calculate the average and standard deviation of the volumetric productivity of the bioreactor at their chosen conditions.
Complete Model Representations

Four complete Model Representations are shown, two for the VBioR (Figure 10[a], [b]) and two for the VCVD (Figure 10[c],[d]). The Model Representation of Team A highlights their reliance on predominately qualitative models of biological mechanisms and bioreactor behavior when a model was used during analysis. This tendency towards qualitative model components is not unusual for the Virtual Bioreactor since just as biology is often more qualitative in its science than either physics or chemistry, the virtual bioreactor tends to elicit the use of qualitative models while students draw upon more quantitative tools during the Virtual CVD. In many cases Team A does not fully utilize modeling to complete the task, instead testing run parameters at various levels while occasionally drawing on biological concepts in an attempt to explain behavior. These additional concepts are not necessarily fully integrated into an overall understanding. The Model Representation of Team A can be contrasted to that of Team B, shown in Figure 10(b). Team B included fewer biological models, but attempted to optimize operating parameters by organizing runs into formal or informal DOEs. The same comparison can be made between groups completing the Virtual CVD, such as in Figures 10(c) and 10(d) which show Teams C and D of the Virtual CVD, respectively.
Overview Statement:
This group initially showed schematic knowledge in their search for the optimum initial substrate concentration and temperature (for which they used $\mu_{max}$). Later, however, they just used a series of formal and informal DOEs in adjusting feed flow rate and concentration. Low schematic, low strategic.

Overview Statement:
The progression of a kinetics model component formed the central thread to this team's solution. They started by using a form of the first order rate law found in Wolf and Tauber. This was not explicitly recognized as such, and was replaced by a more complex, higher order rate equation (that may have been covered in their reactors class). When attempts to solve for the higher order rate parameters didn't work, it was simplified to a first order rate equation. The mass balance and Arrhenius relationship were integrated into this first order rate expression and anthropomorphically called "The Model". This was utilized with an empirical correction factor in this group's progression towards final parameters. High Schematic, High Strategic.
Model Representations capture students’ use of models and their associated learning in a clear and succinct manner. The Model Representation of the Virtual CVD Team C (Figure 10[c]) shows how this group progressed from a reaction rate equation found directly in the literature (Wolf and Tauber) to a more complicated relation and finally to a simpler equation that could be reconciled with their experimental data. Team C continued to integrate additional concepts, such as a mass balance, into their overall understanding of the Virtual CVD. Team C displayed high schematic knowledge by the model components that they chose and high strategic knowledge by the order and manner in which they used them. This knowledge integration contrasted to Team D, who attempted to utilize a scientific model (Arrhenius plot), but failed to do so correctly, highlighting their limited understanding of the model and of the physical phenomena which it describes. Though incomplete, this attempt to implement this model displayed a moderate, or “neutral,” level of strategic knowledge. The Model Representations facilitate the identification of student learning, comparison of the application of modeling components and characterization of schematic and strategic knowledge.

An important objective of the virtual laboratories is to challenge students’ tolerance for ambiguity by presenting them with an open-ended and ill-structured task. Students’ tolerance for ambiguity, or lack there of, is tied to their level of strategic knowledge and can be identified by the Model Representations. A discomfort with ambiguity can be identified by Emotional
Responses of frustration or by the rapid abandonment, rather than adjustment, of ideas and model components. Virtual CVD Team D displayed a low tolerance for ambiguity by their return to empirical methods after unsuccessful use of theoretical models. The identification of these cognitive conflicts, when actual results do not match expected results nor agree with established models of system behavior, is a valuable capability of the model representations.

Conclusion

It is expected that Model Representations will facilitate the understanding of how students learn which can then lead to improved instructional design of the CVD and Bioreactor Virtual Laboratories as well as software development of additional Virtual Laboratories. An understanding of student learning and cognitions will aid in the general integration of Virtual Laboratories into engineering curricula. By exposing students’ cognitions and model use, the Model Representations can provide a more thorough and robust tool than exams, course grades or student feedback to assess the abilities of students to solve ill-structured problems and to provide direction as to specifically what can be improved in instruction.

References