

This work is on a Creative Commons Attribution-NonCommercial-ShareAlike 3.0 Unported (CC BY-NC-SA 3.0) license, <https://creativecommons.org/licenses/by-nc-sa/3.0/>. Access to this work was provided by the University of Maryland, Baltimore County (UMBC) ScholarWorks@UMBC digital repository on the Maryland Shared Open Access (MD-SOAR) platform.

Please provide feedback. Please support the ScholarWorks@UMBC repository by emailing scholarworks-group@umbc.edu and telling us what having access to this work means to you and why it's important to you. Thank you.

Interactive Computer Simulations as Pedagogical Tools in Biology Labs

Karen Whitworth,^{†*} Sarah Leupen,[†] Christopher Rakes,[§] and Mauricio Bustos^{†*}

[†]Department of Biological Sciences and [§]Department of Education, University of Maryland Baltimore County, Baltimore, MD 21250

ABSTRACT

Student learning in biology may be impaired by instructional environments that emphasize technical methodology over analysis. We hypothesized that time gained by experimenting with accurate computer simulations could be used to engage students in analytical, creative learning. The effects of treatments that combined a week of simulated lab instruction with a week of standard lab instruction in different order (E-to-S and S-to-E) were examined using a controlled experimental design with random assignment of lab sections and hierarchical linear modeling analysis to account for possible clustering within sections. Data from a large sample of students ($N = 515$) revealed a significant increase (1.59 SD) in posttest scores for both treatment groups over the control. We posit as a plausible explanation the reinforcement of psychomotor learning due to strong engagement of cognitive processes facilitated by the computer simulation. This study supports a wider use of computer simulations as learning tools in laboratory courses.

INTRODUCTION

Efforts to increase quantitative literacy across the biology curriculum are often hampered by the difficulty of adding more instruction into an already packed schedule and by a paucity of financial and human resources needed to re-educate biology instructors in math and computer science or to staff multidisciplinary teaching teams with biologists, mathematicians, science, technology, engineering, and mathematics (STEM) education specialists, and computer scientists. As a result, students graduating from many biology programs lack sufficient quantitative skills (Gross, 2004; Hoy, 2004). This systemic problem particularly handicaps biology graduates compared with other STEM majors and hinders or even thwarts their success in life science careers that require quantitative skills. Contemporary biology curricula aim to educate students in three complementary areas: context, concept, and literacy (National Research Council, 2003). Traditional biology education focuses primarily on content and concept, while quantitative literacy in mathematics and computer science is addressed through courses offered by specialized departments (Brent, 2004; Gross, 2004; Hoy, 2004). Such separation of biological context and concept from quantitative literacy instruction transfers the burden of integrating knowledge from those areas squarely upon the students. Unfortunately, many otherwise excellent students who graduate from biology programs tend to have an overly contextualized understanding of basic concepts (Masatacusa *et al.*, 2011). Halpern and Hakel (2003) suggested that decontextualization needs to be actively promoted in biology teaching and recommended abstract representations of scientific concepts as effective decontextualization techniques. The use of simulations and models specifically has been recommended as a technique to enhance the transfer of conceptual knowledge to new contexts (Salomon and Perkins, 1996). Computer hardware and software advances have allowed instructors to model complex biological phenomena. While computer simulations based on mathematical modeling have been developed and used in physics and engineering, and to some degree in chemistry education, their use to enhance biology laboratory courses has

Kathryn E. Perez, *Monitoring Editor*

Submitted Sep 28, 2017; Revised Jun 15, 2018;

Accepted Jul 5, 2018

CBE Life Sci Educ September 1, 2018 17:ar46

DOI:10.1187/cbe.17-09-0208

[†]Present address: Department of Biology, Indiana University, Bloomington, IN 47405.

*Address correspondence to: Mauricio Bustos (bustos@umbc.edu).

© 2018 K. Whitworth *et al.* CBE—Life Sciences Education © 2018 The American Society for Cell Biology. This article is distributed by The American Society for Cell Biology under license from the author(s). It is available to the public under an Attribution–Noncommercial–Share Alike 3.0 Unported Creative Commons License (<http://creativecommons.org/licenses/by-nc-sa/3.0>).

“ASCB®” and “The American Society for Cell Biology®” are registered trademarks of The American Society for Cell Biology.

been slower to develop (Lemerle *et al.*, 2005). Methods for integrating computer simulations have been tried and found to either improve student achievement, student attitudes, or both (e.g., Hu *et al.*, 2012a,b), and a review of the relevant literature concludes that the positive learning effects of computer simulations are especially strong in the laboratory setting (Rutten *et al.*, 2012). Simulations driven by deterministic equations allow students to examine biological phenomena without the encumbrance of noise and experimental errors. Stochastic models, on the other hand, take advantage of mathematical precision while allowing for statistical variations in simulated outcomes, thereby providing students with more realistic simulated data (e.g., Linh and Ton, 2011; Lv and Wang, 2011; Chevalier and El-Samad, 2012).

The conceptual framework guiding this study was also founded in Woods's (2007) theory of interdisciplinary communicative competence, which defines interdisciplinary competence as an outgrowth of communication across academic knowledge domains and effective communication among academic cultures. In Woods's theory, the integration of multiple academic areas through a model of inquiry will provide better opportunities to develop more useful competencies and communication skills as well as more positive attitudes.

This study began with the creation of three computer simulations of an acid phosphatase-catalyzed chemical reaction using the scientific programming language MATLAB. The simulations accurately reproduced the results of experimental protocols that have been employed for many years in a standard general biology laboratory course, BIOL 300L. The deterministic simulation works by solving a system of ordinary differential equations, while the stochastic simulation implements the Gillespie algorithm (Gillespie, 1977). Students participating in the study were not instructed in the theory and algorithms behind either simulation, but they could change key simulation parameters by means of an interactive graphical user interface (GUI). Aside from the simulation, all other aspects of the laboratory exercise, including data analysis, report preparation, and assessments, were exactly the same for control and treatment groups. Results from pretest and posttest assessments and quizzes were collected from a large sample of students ($N = 515$). Hierarchical linear modeling (HLM) analysis of the data confirmed the lack of between-section variability and initial group equivalence. Furthermore, HLM revealed that treatment group posttest scores were higher (1.59 SD) than those of the control group, supporting the general conclusion that computer simulations increased students' learning of enzyme function.

MATERIALS AND METHODS

Enzyme Kinetics Experiments

The enzyme kinetics laboratory exercise is a part of the BIOL 300L General Biology Laboratory course that all students majoring in biology and biochemistry at University of Maryland Baltimore County (UMBC) are required to take. The exercise comprises 2 weeks of experiments on wheat germ acid phosphatase (Sigma-Aldrich) and a third week that is devoted to data analysis, discussion, and report preparation. During the first week, the exercise explores the effect of temperature and pH on a commercial acid phosphatase preparation that exhibits optimum activity at 37°C and pH 4.8. Supplemental Figure S1, A and B, shows representative results obtained by students

during week 1. During the second week, students learn to use Michaelis-Menten and Lineweaver-Burk plots to calculate kinetic parameters, V_{\max} and K_M , for acid phosphatase. They also investigate the effect of two reversible chemical inhibitors, phosphate ion and fluoride ion, on acid phosphatase V_{\max} and K_M (Supplemental Figure S1C). During both weeks, the students measure the activity (rate of reaction) of acid phosphatase on the artificial substrate *p*-nitrophenyl phosphate, which the enzyme hydrolyzes to the reaction product *p*-nitrophenol. The standard incubation period for this reaction at 37°C and pH 4.8 is 15 minutes. To calculate the rate of reaction v_0 in micromoles of product produced per minute, the students construct a standard curve of concentration versus absorbance for *p*-nitrophenol. This standard curve reflects the Beer-Lambert law of light absorption by a substance in solution. Further details of the experimental protocols used in weeks 1 and 2 are provided in the Michaelis-Menten.

Enzyme Kinetics Simulations and Graphical User Interface

Computer simulations of the acid phosphatase reaction system were first programmed in MATLAB using a modification of previously published work (Higham, 2008). Briefly, a system of seven mass-action kinetics equations representing a Michaelis-Menten enzyme (Michaelis-Menten) is solved numerically, either in a deterministic (DRRM) or a stochastic (SSAM) manner. The deterministic method is based on a differential equation-solver algorithm (MATLAB), while the stochastic method represents an application of the Gillespie algorithm (Gillespie, 1977). Temperature dependence is modeled with two logistic functions: a negative logistic function represents temperature-dependent enzyme degradation, while a positive logistic function represents temperature-dependent activation. The dependence on pH is represented as an equilibrium between the active enzyme and two inactive forms, EH^+ and EOH^- , that result from the reversible binding of H^+ and OH^- ions, respectively (Supplemental Material). Both methods are presented to the students as GUIs that are available from simlabs.umbc.edu, either as downloadable, stand-alone versions or online. Supplemental Figure S2 depicts an annotated screenshot of the stand-alone GUI for a generic enzyme without inhibitor used in week 2, which is also analogous to the online GUI. A user can select the simulation type and input values of elementary rate constants for enzyme and inhibitor (edit control boxes 1–7); reaction parameters (edit control boxes 9–14) such as starting concentrations of substrate, enzyme, and inhibitor (S_0 , E_0 , and I_0); reaction volume; and simulation and assay times. Values can be entered by hand or selected from a pull-down menu (control box 8) that offers a choice of four different sets of prefit enzyme systems for a generic enzyme, acid phosphatase (without inhibitor), acid phosphatase plus competitive inhibitor (phosphate ion), and acid phosphatase plus non-competitive inhibitor (fluoride ion). These parameter values are specific for commercial preparations and may change as the enzymes show decreased activity over time. Consequently, the default values of reaction constants listed in the Supplemental Material in Supplemental Table S1 should be recalibrated before use.

The simulations produce concentration profiles of substrate, enzyme, enzyme-substrate complex, and inhibitor complexes that are plotted at the bottom of the GUI. A rate of reaction v_0 is

TABLE 1. Technical and conceptual learning goals of the E and S protocols

Technical learning goals specific to the E protocol
1. Learn to use volumetric equipment, such as reagent dispenser bottles, graduated cylinders, and pipettes to mix reagent stocks of buffer, substrate, enzyme, and inhibitors
2. Learn to use strong alkali (NaOH) to stop the enzymatic reaction and increase the specific absorbance of the reaction product, <i>p</i> -nitrophenol.
3. Learn to control the pH of solution, and accurately time the duration of the assay (e.g., 15 minutes).
4. Learn to use a temperature-controlled water bath.
5. Learn the Beer-Lambert law of light absorbance and how to use a spectrophotometer to measure absorbance and the amount of a chemical (<i>p</i> -nitrophenol) dissolved in water.
6. Learn various laboratory psychomotor skills important for a profession in the life sciences.
Technical learning goals specific to the S protocol
1. Learn to use a graphic user interface (GUI) to control reaction parameters: pH, temperature, reaction time, and chemical concentrations (enzyme, substrate, inhibitors)
2. Learn to use a GUI to select preset simulation parameters: elementary rate constants for forward and reverse binding reactions for substrate and inhibitors; enzyme turnover rate.
3. Learn to read out the amount of product made predicted by the simulation and interpret concentration versus time plots for all chemical species included in the Michaelis-Menten model.
Conceptual learning goals common to both protocols
1. Learn basic concepts of experimental protocol design and execution.
2. Learn to calculate the activity of an enzyme from the amount of product made during a fixed period (assay time).
3. Learn to find the optimum pH and temperature of an enzyme-catalyzed reaction.
4. Learn to calculate the enzyme's kinetic parameters V_{\max} and K_M for a given substrate and how reversible chemical inhibitors can be used to manipulate those parameters.

calculated automatically for a specified assay time using a spline interpolation method. Individual concentration versus time plots of each chemical species can be selected for display using plot control check boxes labeled “S” (substrate), “E” (enzyme), “ES,” “P” (phosphate), “I” (inhibitor), “EI,” and “ESI.”

Education Research Protocol

A controlled design with random assignment of laboratory sections to experimental groups was used to test the effectiveness of computer simulations in an existing laboratory course, BIOL 300L General Biology Laboratory. The experimental protocol involving human subjects was reviewed and approved in advance by the UMBC Institutional Review Board. Hereafter, the capital letters “E” and “S” that appear in the group designations E-to-E, S-to-E, and E-to-S indicate the type of laboratory protocol used during weeks 1 and 2 (Supplemental Figure 2). The E (experimental) protocol was a standard enzyme kinetics protocol designed to teach students how to set up and incubate enzyme reaction assays and measure the concentration of the yellow-colored reaction product *p*-nitrophenol using a spectrophotometer. The S (simulation) protocol replaced the physical experiment with an accurate computer simulation of the enzyme-catalyzed reaction. Implementing the S protocol required additional instruction on how to use GUI controls to change the starting concentrations of substrate, enzyme, and inhibitor; how to read and record rate of reaction results; and how to interpret concentration versus time plots. During week 1, the S protocol used specialized GUIs and instructions on how to operate pH and temperature control sliders (Supplemental Material). Table 1 lists the technical learning goals specific to the E and S protocols and the conceptual learning goals that were common to both protocols.

Demographic data on gender, English language learning, academic seniority, and major were obtained from student

records and from individual responses to the psychometric survey. The sample of 515 students was 56% female and 44% male; 83% of respondents indicated English was their primary language, while 17% declared English was a second language. The sample was composed of juniors and seniors, mostly biology BA, biology BS, and biochemistry and molecular biology majors (90%), who are required to take the BIOL 300L course. The remaining 10% of students had declared majors in chemical engineering, computer science, psychology, and health administration and policy.

A random-number generator was used to assign laboratory sections, each consisting of 24 individuals, to the control and treatment groups. In Spring and Fall, the S-to-E and E-to-S groups comprised four sections each, while the control group E-to-E comprised two sections. In the Summer offering of the course, there were three sections, each assigned randomly to one experimental group. To maintain the same learning objectives for both E and S protocols, when students carried out the S protocol, we did not require them to investigate what would happen if they changed the rate constants of a simulation, but instead told them what would likely happen and allowed them to spend time trying different “what-if” scenarios.

Assessments

Three types of assessments—pretest, quizzes (1 and 2), and posttest—were used to evaluate learning outcomes. The assessment schedule for weeks 1–3 is summarized in Supplemental Figure S3. Immediate feedback on the exercises carried out during weeks 1 and 2 was obtained from two short quizzes that were given in weeks 2 and 3. Quizzes used in the Spring semester were modified for the Fall semester to follow recommendations from item response theory. Further details of Spring and Fall semester quizzes are provided in the Supplemental Material.

During the Spring and Summer semesters, posttest course outcome evaluations were based on eight final exam questions relevant to the enzyme kinetics exercise (Supplemental Material). Item response theory indicated weak discrimination power for two questions, prompting their subsequent modification for the Fall semester section. A total of 515 pretest, quiz, and posttest responses were collected during a three-semester period.

An 11-question, five-point Likert scale psychometric survey was administered to students of all groups at the end of the Spring and Fall semesters. The objective of the survey was to gain demographic information on gender and English language, as well as insight into students' perceptions and biases and preferences regarding the use of experiments, computers, and computer simulations in a laboratory environment (Supplemental Material).

Missing Data

Of 515 students who participated in this study, 154 (29.9%) were missing at least one value, and all variables had missing data, except the treatment condition indicator variable. Little's (1987) test of missing completely at random (MCAR) indicated that the missing data patterns were not MCAR, $\chi^2(30) = 68.99$, $p < 0.001$. Multiple imputation (Rubin, 1987), an expansion of multiple-regression imputation based on Bayesian inference, was used to impute values for the missing data. Multiple imputation is the preferred imputation method for data with missing patterns that are not MCAR (Gelman *et al.*, 2004). The imputation model regresses each missing data point on every other variable. The true value for the missing data point is considered the mean of a distribution, and sampling error will result in a potentially different value each time a multiple-regression

imputation is run. Logistic regression was used to impute missing values for indicator variables. The multiple imputation procedures were run using IBM SPSS Statistics v. 25. Five imputed data sets were produced, as recommended by Brick *et al.* (2005) and Garson (2009).

Data Analysis

Because treatment condition was assigned by section, two-level HLM (Raudenbush and Bryk, 2002) was used to examine possible clustering within sections. The first level (students) consisted of scores on the pretest, quiz 1, quiz 2, and posttest along with gender and English language indicator variables. The second level (section) consisted of the treatment condition and semester indicator variables. The sample consisted of 515 students in 23 sections, an average of 22 students per section. Sections were randomly assigned to treatment condition. Using Optimal Design software (Spybrook *et al.*, 2013), statistical power was estimated for an alpha level of 0.05. For an intra-class correlation (ICC) of 0.10 (the minimum needed to justify the use of HLM; Byrne, 2012), a statistical power of 0.80 was estimated for a minimum effect size of 0.47. All estimates were computed with full maximum likelihood. All HLM analyses were computed with an HLM of 7.03 (Raudenbush *et al.*, 2013).

RESULTS

Five imputed data sets were computed from the original. The imputed data sets were comparable to the original data set with no statistically significant *t* tests between the imputed data sets and the original (Table 2). The overall research goal was to examine the impact of the computer simulation treatment on the posttest. Descriptive statistics by treatment group are therefore provided (Table 3). All three groups appeared to show growth in

TABLE 2. Sample size, means, and SDs for each imputed data set

	Original	Imputed data set				
		1	2	3	4	5
Pretest						
Sample size (<i>N</i>)	513	515	515	515	515	515
Mean	0.69	0.69	0.69	0.70	0.69	0.69
SD	0.30	0.30	0.30	0.30	0.29	0.29
<i>t</i> Ratio from original	—	0.03	0.02	0.05	0.02	0.03
Quiz 1						
Sample size (<i>N</i>)	497	515	515	515	515	515
Mean	0.68	0.68	0.68	0.68	0.68	0.67
SD	0.21	0.21	0.21	0.21	0.21	0.21
<i>t</i> Ratio from original	—	0.04	0.01	0.01	0.01	0.11
Quiz 2						
Sample size (<i>N</i>)	492	515	515	515	515	515
Mean	0.72	0.72	0.72	0.72	0.72	0.72
SD	0.26	0.26	0.26	0.27	0.26	0.27
<i>t</i> Ratio from original	—	0.03	0.00	0.02	0.08	0.15
Posttest						
Sample size (<i>N</i>)	513	515	515	515	515	515
Mean	0.98	0.98	0.98	0.98	0.98	0.98
SD	0.26	0.27	0.27	0.26	0.26	0.26
<i>t</i> Ratio from original	—	0.03	0.01	0.02	0.02	0.03

TABLE 3. Descriptive statistics by treatment groups^a

Variable	E-to-E (control)			S-to-E (treatment week 1)			E-to-S (treatment week 2)		
	N	Mean	SD	N	Mean	SD	N	Mean	SD
Pretest	284	0.686	0.256	111	0.705	0.337	118	0.702	0.340
Quiz 1	270	0.700	0.206	109	0.651	0.208	118	0.644	0.222
Quiz 2	270	0.711	0.276	107	0.729	0.247	115	0.724	0.252
Posttest	284	0.886	0.227	111	1.066	0.268	118	1.113	0.261

^aDescriptive statistics are based on the original data set (un-imputed). The E-to-E group consisted of 13 sections. S-to-E and E-to-S groups consisted of five sections each.

their means from pretest to posttest. To determine the degree to which that growth was statistically significant and differed across groups, we used HLM to account for the clustering of students within sections. The HLM analyses consisted of five models:

1. Unconditional model (no predictors at the section or student levels)
2. Analysis of covariance (ANCOVA) model to account for student-level effects on the posttest
3. Random effects model to assess the degree of between-section variability in the student-level covariates
4. Treatment effects model to determine the effect of being in one of the treatment groups on the posttest.
5. Treatment comparison model to determine differences in the effect of being in treatment week 1 (S-to-E group) or treatment week 2 (E-to-S group) on the posttest.

Unconditional Model

The unconditional model (Eqs. 1 and 2) was computed to determine the overall amount of variance at the section level (Level 2) and to determine the amount of variance between sections using the ICC.

$$\text{Student Level 1 Model: } \text{POST}_{ij} = \beta_{0j} + r_{ij} \quad (1)$$

$$\text{Section Level 2 Model: } \beta_{0j} = \gamma_{00} + u_{0j} \quad (2)$$

where

POST_{ij} = the posttest score for student i in section j

γ_{00} = the intercept term representing the grand mean on the posttest

u_{0j} = the unique effect for each section j on the posttest grand mean (i.e., variance term)

β_{0j} = the section means on the posttest

r_{ij} = the unique effect for each student i in section j on the posttest grand mean (i.e., variance term)

The values for the unconditional model were $\gamma_{00} = 0.972$, $p < 0.001$; $u_0 = 0.024$, $\chi^2 (df = 22) = 292.76$, $p < 0.001$; $r = 0.046$. The ICC was 0.342, meaning that ~34% of the variance was at the section level. Byrne (2012) recommended a minimum ICC of 0.10, so the continued use of HLM was deemed to be justified. The reliability estimate for the group means (β_{0j}) was 0.920, providing further evidence that the group means varied substantially across sections. Robust SEs were nearly identical to the ordinary SE estimates, indicating that the assumption of a normal distribution was unlikely to have been violated. Using the unconditional model as a baseline, the student model was developed to explain the impact of as many student characteristics as possible that may have been confounded by section effects (Ma *et al.*, 2008).

ANCOVA Model

The student model analysis began with an ANCOVA model, in which student-level predictors of the posttest were added to the model as fixed effects, that is, with no variance term at the section level. Using backward regression to develop the student model, all student-level variables were added to the unconditional model (Eqs. 3–9).

$$\begin{aligned} \text{Student Level 1 Model: } \text{POST}_{ij} = & \beta_{0j} + \beta_{1j} * (\text{FEMALE}_{ij}) \\ & + \beta_{2j} * (\text{ELL}_{ij}) + \beta_{3j} * (\text{PRE}_{ij}) + \beta_{4j} * (\text{QUIZ1}_{ij}) \\ & + \beta_{5j} * (\text{QUIZ2}_{ij}) + r_{ij} \end{aligned} \quad (3)$$

$$\text{Section Level 2 Model: } \beta_{0j} = \gamma_{00} + u_{0j} \quad (4)$$

$$\beta_{1j} = \gamma_{10} \quad (5)$$

$$\beta_{2j} = \gamma_{20} \quad (6)$$

$$\beta_{3j} = \gamma_{30} \quad (7)$$

$$\beta_{4j} = \gamma_{40} \quad (8)$$

$$\beta_{5j} = \gamma_{50} \quad (9)$$

where

POST_{ij} = the *posttest* score for student i in section j

γ_{00} = the intercept term representing the grand mean on the posttest

γ_{10} = the overall female slope

γ_{20} = the overall English language learner (ELL) slope

γ_{30} = the overall pretest slope

γ_{40} = the overall quiz 1 slope

γ_{50} = the overall quiz 2 slope

u_{0j} = the unique effect for each section j on the posttest grand mean (i.e., variance term)

β_{0j} = the section means on the posttest

β_{1j} = the section female slopes

β_{2j} = the section ELL slopes

β_{3j} = the section pretest slopes

β_{4j} = the section quiz 1 slopes

β_{5j} = the section quiz 2 slopes

r_{ij} = the unique effect for each student i in section j on the posttest grand mean (i.e., variance term)

Because the slopes were fixed effects, the section slopes (β_{1j} – β_{5j}) were held constant across sections to their overall

TABLE 4. Fixed and random effects of the ANCOVA model

Fixed effects	Coefficient	SE	df	t Ratio
Posttest mean, γ_{00}	0.978	0.033	22	30.02***
Female slope, γ_{10}	-0.007	0.019	320	-0.35
ELL slope, γ_{20}	-0.011	0.024	23	-0.45
Pretest slope, γ_{30}	0.158	0.036	487	4.41***
Quiz 1 slope, γ_{40}	0.141	0.042	487	3.37***
Quiz 2 slope, γ_{50}	0.109	0.040	487	2.74***
Random effects	Variance component	df	χ^2	p Value
Posttest mean, u_0	0.024	22	328.66	<0.001
Level 1, R	0.041			

*** $p < 0.001$.

respective slopes (γ_{10} - γ_{50}). This model does not test differences between sections; it measures the overall impact of the predictors on posttest. The female and ELL variables were dichotomous, so the slopes represent the impact on the posttest of being female or an English language learner. The variables pretest, quiz 1, and quiz 2 were group mean centered in this and all subsequent models, so their slopes (coefficients in Table 4) represent the impact on the posttest of being higher or lower than the section average on the predictor variables.

The reliability estimate for the section means (β_{0j}) was 0.929, indicating that even after adding student-level predictors to the model, the group means varied substantially across sections. Pretest, quiz 1, and quiz 2 scores were significant predictors of the posttest. The slopes for the female and ELL indicator variables were statistically nonsignificant (Table 4). These two variables were therefore removed from the model for parsimony, resulting in Eqs. 10–14. Coefficients and tests of statistical significance (Table 4) were similar to those of the full ANCOVA model.

$$\text{Student Level 1 Model: } \text{POST}_{ij} = \beta_{0j} + \beta_{1j} * (\text{PRE}_{ij}) + \beta_{2j} * (\text{QUIZ1}_{ij}) + \beta_{3j} * (\text{QUIZ2}_{ij}) + r_{ij} \quad (10)$$

$$\text{Section Level 2 Model: } \beta_{0j} = \gamma_{00} + u_{0j} \quad (11)$$

$$\beta_{1j} = \gamma_{10} \quad (12)$$

$$\beta_{2j} = \gamma_{20} \quad (13)$$

$$\beta_{3j} = \gamma_{30} \quad (14)$$

Robust SEs were nearly identical to the ordinary SE estimates, indicating that the assumption of a normal distribution was unlikely to have been violated.

Random Effects Model

The next model (Eqs. 15–19) added random effects to pretest, quiz 1, and quiz 2 slopes to determine the degree to which the impact of the predictors on posttest varied between sections.

$$\text{Student Level 1 Model: } \text{POST}_{ij} = \beta_{0j} + \beta_{1j} * (\text{PRE}_{ij}) + \beta_{2j} * (\text{QUIZ1}_{ij}) + \beta_{3j} * (\text{QUIZ2}_{ij}) + r_{ij} \quad (15)$$

$$\text{Section Level 2 Model: } \beta_{0j} = \gamma_{00} + u_{0j} \quad (16)$$

$$\beta_{1j} = \gamma_{10} + u_{1j} \quad (17)$$

$$\beta_{2j} = \gamma_{20} + u_{2j} \quad (18)$$

$$\beta_{3j} = \gamma_{30} + u_{3j} \quad (19)$$

where

- POST_{ij} = the posttest score for student *i* in section *j*
- γ_{00} = the intercept term representing the grand mean on the posttest
- γ_{10} = the overall pretest slope
- γ_{20} = the overall quiz 1 slope
- γ_{30} = the overall quiz 2 slope
- u_{0j} = the unique effect for each section *j* on the posttest grand mean (i.e., variance term)
- u_{1j} = the unique effect for each section *j* on the pretest slope (i.e., variance term)
- u_{2j} = the unique effect for each section *j* on the quiz 1 slope (i.e., variance term)
- u_{3j} = the unique effect for each section *j* on the quiz 2 slope (i.e., variance term)
- β_{0j} = the section means on the posttest
- β_{1j} = the section pretest slopes
- β_{2j} = the section quiz 1 slopes
- β_{3j} = the section quiz 2 slopes
- r_{ij} = the unique effect for each student *i* in section *j* on the posttest grand mean (i.e., variance term)

All fixed effects (γ 's) were statistically significant (Table 5). Pretest, quiz 1, and quiz 2 were therefore retained as predictors of posttest. Robust SEs were nearly identical to the ordinary SE estimates, indicating that the assumption of a normal distribution was unlikely to have been violated.

The reliability estimate for the group means (β_{0j}) was 0.932, indicating that section variability was large for the posttest. The reliability estimates for the section slopes for pretest (β_{1j}), quiz 1 (β_{2j}), and quiz 2 (β_{3j}) were 0.249, 0.144, and 0.146, respectively. These reliability estimates indicated that section variability was

TABLE 5. Fixed and random effects for random effects model

Fixed effects	Coefficient	SE	df	t Ratio
Posttest mean, γ_{00}	0.972	0.034	22	28.88***
Pretest slope, γ_{10}	0.170	0.038	22	4.44***
Quiz 1 slope, γ_{20}	0.144	0.052	22	2.79*
Quiz 2 slope, γ_{30}	0.104	0.040	22	2.61*
Random effects	Variance component	df	χ^2	p Value
Posttest mean, u_0	0.024	22	340.38	<0.001
Pretest slope, u_1	0.009	22	26.02	0.250
Quiz 1 slope, u_2	0.009	22	21.18	>0.500
Quiz 2 slope, u_3	0.005	22	28.33	0.165
Level 1, r	0.040			

* $p < 0.05$.*** $p < 0.01$.

small for all three predictor variables. Their random effects were also not statistically significant (u_1 , u_2 , and u_3 in Table 5). The lack of between-section variability in the pretest provides evidence of initial group equivalence. Their variance terms (u coefficients) were therefore removed in subsequent models for parsimony. Their removal made the revised ANCOVA model (Eqs. 10–14; Table 6) the comparison model for subsequent models.

Impact of Treatment Condition

The next model added the treatment condition to determine its impact on the posttest mean (Eqs. 20–24). The treatment variable combined sections in week 1 (S-to-E) and week 2 (E-to-S) treatment groups into a single group to determine the overall effect of being in a treatment group.

$$\text{Student Level 1 Model: } \text{POST}_{ij} = \beta_{0j} + \beta_{1j} * (\text{PRE}_{ij}) + \beta_{2j} * (\text{QUIZ1}_{ij}) + \beta_{3j} * (\text{QUIZ2}_{ij}) + r_{ij} \quad (20)$$

Section Level 2 Model:

$$\beta_{0j} = \gamma_{00} + \gamma_{01} * (\text{Treatment}_j) + u_{0j} \quad (21)$$

$$\beta_{1j} = \gamma_{10} \quad (22)$$

$$\beta_{2j} = \gamma_{20} \quad (23)$$

$$\beta_{3j} = \gamma_{30} \quad (24)$$

where

POST_{ij} = the posttest score for student i in section j

γ_{00} = the intercept term representing the mean of the control group on the posttest

γ_{01} = the slope of treatment on the control group posttest mean

γ_{10} = the overall pretest slope

γ_{20} = the overall quiz 1 slope

γ_{30} = the overall quiz 2 slope

u_{0j} = the unique effect for each section j on the posttest (i.e., variance term)

β_{0j} = the section means on the posttest

β_{1j} = the section pretest slopes

β_{2j} = the section quiz 1 slopes

β_{3j} = the section quiz 2 slopes

r_{ij} = the unique effect for each student i in section j on the posttest grand mean (i.e., variance term)

Because the treatment variable was dichotomous, its slope (γ_{01}) represents the value added to the posttest control group mean for being in the treatment group, which was statistically significant (Table 7). The reliability estimate for the section means (β_{0j}) was 0.884, indicating that section variability remained large for the posttest after accounting for the effect of the treatment.

The treatment model resulted in a 41.8% reduction in variance in the posttest from the revised ANCOVA model

TABLE 6. Fixed and random effects for revised ANCOVA model

Fixed effects	Coefficient	SE	df	t Ratio
Posttest mean, γ_{00}	0.972	0.033	22	28.89***
Pretest slope, γ_{10}	0.160	0.032	489	4.96***
Quiz 1 slope, γ_{20}	0.142	0.048	489	2.99**
Quiz 2 slope, γ_{30}	0.109	0.037	489	2.94**
Random effects	Variance component	df	χ^2	p Value
Posttest mean, u_0	0.024	22	328.72	<0.001
Level 1, R	0.041			

** $p < 0.01$.*** $p < 0.001$.

TABLE 7. Fixed and random effects for treatment model

Fixed effects	Coefficient	SE	df	t Ratio
Posttest mean, γ_{00}	0.884	0.035	21	25.22***
Treatment slope, γ_{01}	0.203	0.053	21	3.82***
Pretest slope, γ_{10}	0.160	0.032	489	4.96***
Quiz 1 slope, γ_{20}	0.142	0.048	489	2.99**
Quiz 2 slope, γ_{30}	0.109	0.037	489	2.94**
Random effects	Variance component	df	χ^2	p Value
Posttest mean, u_0	0.014	21	199.69	<0.001
Level 1, r	0.041			

** $p < 0.01$.

*** $p < 0.001$.

(Eqs. 10–14; Table 6). The t value for γ_{01} (fixed effect for treatment) corresponded to an effect size of Cohen’s $d = 1.59$ (Lipsey and Wilson, 2001), meaning that the treatment group scored on average 1.59 SD higher than the control group.

Comparison of Treatment Conditions

Some treatment sections occurred during week 1 (S-to-E group) and others in week 2 (E-to-S group). The treatment comparison model (Eqs. 25–29) examined the effect of the order of the treatment.

$$\text{Student Level 1 Model: } \text{POST}_{ij} = \beta_{0j} + \beta_{1j} * (\text{PRE}_{ij}) + \beta_{2j} * (\text{QUIZ1}_{ij}) + \beta_{3j} * (\text{QUIZ2}_{ij}) + r_{ij} \quad (25)$$

$$\text{Section Level 2 Model: } \beta_{0j} = \gamma_{00} + \gamma_{01} * (\text{Treatment Week 1}_j) + \gamma_{02} * (\text{Treatment Week 2}_j) + u_{0j} \quad (26)$$

$$\beta_{1j} = \gamma_{10} \quad (27)$$

$$\beta_{2j} = \gamma_{20} \quad (28)$$

$$\beta_{3j} = \gamma_{30} \quad (29)$$

where

- POST_{ij} = the posttest score for student i in section j
- γ_{00} = the intercept term representing the mean of the control group on the posttest
- γ_{01} = the slope of treatment week 1 (S-to-E group) on the control group posttest mean
- γ_{02} = the slope of treatment week 2 (E-to-S group) on the control group posttest mean
- γ_{10} = the overall pretest slope
- γ_{20} = the overall quiz 1 slope
- γ_{30} = the overall quiz 2 slope
- u_{0j} = the unique effect for each section j on the posttest (i.e., variance term)
- β_{0j} = the section means on the posttest
- β_{1j} = the section pretest slopes
- β_{2j} = the section quiz 1 slopes
- β_{3j} = the section quiz 2 slopes
- r_{ij} = the unique effect for each student i in section j on the posttest (i.e., variance term)

The slopes for both treatment groups were statistically significant (Table 8). The coefficients for each group were used to compute an effect size difference. The t value for γ_{01} (fixed effect for treatment in week 1) corresponded to an effect size of Cohen’s $d = 1.11$ (Lipsey and Wilson, 2001). The t value for γ_{02} (fixed effect for treatment in week 2) corresponded to an effect size of Cohen’s $d = 1.46$ (Lipsey and Wilson, 2001). The effect size difference was small, 0.35, which is less than half an SD.

TABLE 8. Fixed and random effects for treatment comparison model

Fixed effects	Coefficient	SE	df	t Ratio
Posttest mean, γ_{00}	0.884	0.035	20	25.47***
Treatment week 1 slope, γ_{01}	0.175	0.066	20	2.66*
Treatment week 2 slope, γ_{02}	0.230	0.066	20	3.50**
Pretest slope, γ_{10}	0.160	0.032	489	4.96***
Quiz 1 slope, γ_{20}	0.142	0.048	489	2.99**
Quiz 2 slope, γ_{30}	0.109	0.037	489	2.94**
Random effects	Variance component	df	χ^2	p Value
Posttest mean, u_0	0.014	20	196.61	<0.001
Level 1, r	0.041			

* $p < 0.05$.

** $p < 0.01$.

*** $p < 0.001$.

DISCUSSION

The present study showed that combining a hands-on laboratory curriculum with a computer simulation of the same experiment (herein described as “treatment”) increased posttest scores on average 1.59 SD relative to a control group, allowing for the conclusion that a computer simulation improved conceptual understanding of the reaction type under study. A treatment comparison model designed to test a possible effect of the order of the treatment (S-to-E vs. E-to-S) uncovered an effect size difference of 0.35 in favor of the E-to-S treatment (treatment week 2). However, that difference was less than half an SD, and its meaning remains uncertain until a further, larger study can be carried out. In the future, it would be interesting to explore the possibility that undertaking a physical experiment to develop psychomotor skills in advance of a simulation may afford students the knowledge structures, or schemata, to which they could “attach” the concepts contained in the simulation (Ambrose *et al.*, 2010). The takeaway of our analysis is clear: combining a physical instructional environment with a simulation that recapitulates an important part of that environment can enhance learning. This concept will inform others’ use of simulations in laboratory courses.

Training students to become skilled in technical tasks has practical value, but it also detracts from teaching them how to think about the science behind the experiments. Generating accurate enzyme kinetics data with computer simulations reduced the amount of time and effort that students had to devote to technical chores during the class period. A perception survey (Supplemental Material) confirmed the students’ general acceptance of the simulation exercises, which will reassure instructors of the appropriateness of using computer simulations in comparable courses.

Although our data demonstrated that using a quantitatively accurate simulation of an actual laboratory experiment can enhance conceptual understanding, we believe that we will see a stronger effect on learning when we employ the simulation more robustly—that is, as a tool for students to actively explore the relationship between the variables (enzyme concentration, substrate concentration, inhibitor concentration, reaction rate constants, etc.) and the outcome of the experiment, rather than just doing a single run of the simulation according to one set of conditions, as was done here. Evidence shows that two of the strongest factors influencing learning are repeatedly testing oneself or being tested on the concept (Glover, 1989; Karpicke and Roediger, 2007) and receiving immediate feedback on the result (McKendree, 1990; Hattie and Timperley, 2007). A simulation is in fact a perfect tool to achieve both types of cognitive stimulation, as students can change the input variables to test an idea or hypothesis they have—or even in the absence of an idea or hypothesis—and immediately see the result. Effectively, students can perform hundreds of experiments using a simulation in the same time it will take them to do one or a few physical experiments. Having students experiment with the simulation and extract the conceptual principles from their results is probably the most robust use of a simulation, and it is the method we plan to employ in the future with this simulation and others.

Fundamentally, we want students to understand not just acid phosphatase function, but enzyme function in general, and to understand not just the kinetics of enzyme-catalyzed

reactions, but rate of change in general as a component of biophysical systems. That is, we want students to be able to decontextualize the concepts they learn from one example to the extent that they can transfer them across situations. Learning that is too bound to context has been described as having an “irrelevant concreteness” (Sloutsky *et al.*, 2005) that limits development and scope of conceptual understanding. By employing an accurate simulation, we can accomplish more decontextualization and transferability to other systems and other courses (Salomon and Perkins, 1996).

ACKNOWLEDGMENTS

We acknowledge the generous financial support for this study from the Office of the Dean, College of Natural and Mathematical Sciences, UMBC, and the Hrabowski Academic Innovation Fund. We thank the Department of Biological Sciences for its material support toward the purchase and use of 10 laptop computers. An online version of the enzyme kinetics simulation is available on the project’s website at simlabs.umbc.edu. Copies of the compiled MATLAB code may be requested via mail or email addressed to the corresponding author.

REFERENCES

- Ambrose, S., Bridges, M., DiPietro, M., Lovett, M., & Norman, M. (2010). *How learning works: 7 research-based principles for smart teaching* (chapter 2). San Francisco: Wiley.
- Brent, R. (2004). Intuition and numeracy. *Cell Biology Education*, 3, 88–90. doi: 10.1187/cbe.04-03-0041
- Brick, J. M., Jones, M. E., Kalton, G., & Valliant, R. (2005). Variance estimation with hot deck imputation: A simulation study of three methods. *Survey Methodology*, 31, 151–159.
- Byrne, B. M. (2012). *Structural equation modeling with MPlus: Basic concepts, applications, and programming*. Mahwah, NJ: Erlbaum.
- Chevalier, M. W., & El-Samad, H. (2012). Towards a minimal stochastic model for a large class of diffusion–reactions on biological membranes. *Journal of Chemical Physics*, 137, 084103. doi: 10.1063/1.4746692
- Garson, G. D. (2009). *Data imputation for missing values* [Online]. Raleigh: North Carolina State University. Retrieved February 14, 2018, from <https://faculty.chass.ncsu.edu/garson/PA765/missing.htm>
- Gelman, A., Carlin, J. B., Stern, H. S., & Rubin, D. B. (2004). *Bayesian data analysis* (2nd ed.). Boca Raton, FL: Chapman & Hall/CRC.
- Gillespie, D. (1977). Exact stochastic simulation of coupled chemical reactions. *Journal of Physical Chemistry*, 81, 2340–2361.
- Glover, J. A. (1989). The “testing phenomenon”: Not gone but nearly forgotten. *Journal of Educational Psychology*, 81, 392–399. doi: 10.1037/0022-0663.81.3.392
- Gross, L. J. (2004). Interdisciplinarity and the undergraduate biology curriculum: Finding a balance. *Cell Biology Education*, 3, 85–87. doi: 10.1187/cbe.04-03-0040
- Halpern, D. F., & Hakel, M. D. (2003). Applying the science of learning to the university and beyond. *Change*, 35, 36–41.
- Hattie, J., & Timperley, H. (2007). The power of feedback. *Review of Educational Research*, 77, 81–112.
- Higham, D. J. (2008). Modeling and simulating chemical reactions. *Society for Industrial and Applied Mathematics Review*, 50, 347–360. doi: 10.1137/060666457
- Hoy, R. (2004). New math for biology is the old new math. *Cell Biology Education*, 3, 90–92. doi: 10.1187/cbe.04-03-0042
- Hu, D., Li, M., Zhou, R., & Sun, Y. (2012a). Design and optimization of photo bioreactor for O₂ regulation and control by system dynamics and computer simulation. *Bioresource Technology*, 104, 608–615. doi: 10.1016/j.biortech.2011.11.049
- Hu, D., Zhou, R., Sun, Y., Tong, L., Li, M., & Zhang, H. (2012b). Construction of closed integrative system for gases robust stabilization employing

- microalgae peculiarity and computer experiment. *Ecological Engineering*, 44, 78–87. doi: 10.1016/j.ecoleng.2012.04.001
- Karpicke, J. D., & Roediger, H. L. (2007). Repeated retrieval during learning is the key to long-term retention. *Journal of Memory and Language*, 57, 151–162. doi: 10.1016/j.jml.2006.09.004
- Lemerle, C., Di Ventura, B., & Serrano, L. (2005). Space as the final frontier in stochastic simulations of biological systems. *FEBS Letters*, 579, 1789–1794. doi: 10.1016/j.febslet.2005.02.009
- Linh, N. T. H., & Ton, T. V. (2011). Dynamics of a stochastic ratio-dependent predator–prey model. *Analysis and Applications*, 9, 329–344. doi: 10.1142/S0219530511001868
- Lipsey, M. W., & Wilson, D. B. (2001). *Practical meta-analysis*. Thousand Oaks, CA: Sage.
- Little, R. J. A. (1987). A test of missing completely at random for multivariate data with missing values. *Journal of the American Statistical Association*, 83, 1198–1202.
- Lv, J., & Wang, K. (2011). Asymptotic properties of a stochastic predator–prey system with Holling II functional response. *Communications in Nonlinear Science & Numerical Simulation*, 16, 4037–4048. doi: 10.1016/j.cnsns.2011.01.015
- Ma, X., Ma, L., & Bradley, K. D. (2008). Using multilevel modeling to investigate school effects. In O’Connell, A. A., & McCoach, D. B. (Eds.), *Multilevel modeling of educational data* (pp. 59–110). Charlotte, NC: Information Age Publishing.
- Masatacusa, E. J., Snyder, W. J., & Hoyt, B. (2011). *Effective instruction for STEM disciplines: From learning theory to college teaching*. San Francisco: Wiley.
- McKendree, J. (1990). Effective feedback content for tutoring complex skills. *Human–Computer Interaction*, 5(4), 381–413.
- National Research Council. (2003). *BIO2010: Transforming undergraduate education for future research biologists*. Washington, DC: National Academies Press. Retrieved February 14, 2018, from www.nap.edu/openbook.php?isbn=0309085357
- Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear models: Applications and data analysis methods* (2nd ed.). Thousand Oaks, CA: Sage.
- Raudenbush, S., Bryk, T., & Congdon, R. (2013). *HLM 7 Hierarchical linear and nonlinear modeling*. Skokie, IL: Scientific Software International. Retrieved February 14, 2018, from www.ssicentral.com
- Rubin, D. B. (1987). *Multiple imputation for nonresponse in surveys*. Hoboken, NJ: Wiley.
- Rutten, N., van Joolingen, W., & van der Veen, J. (2012). The learning effects of computer simulations in science education. *Computers & Education*, 58(1), 136–153.
- Salomon, G., & Perkins, D. N. (1996). Learning in wonderland: What computers really offer education. In Kerr, S. (Ed.), *Technology and the future of education* (pp. 111–130). NSSE Yearbook. Chicago: University of Chicago Press. Retrieved February 14, 2018, from www.edu.haifa.ac.il/personal/gsalomon/nsse%5B1%5D.pdf
- Sloutsky, V. M., Kaminski, J. A., & Heckler, A. F. (2005). The advantage of simple symbols for learning and transfer. *Psychonomic Bulletin & Review*, 12, 508–513. doi: 10.1.1.120.318
- Spybrook, J., Bloom, H., Congdon, R., Hill, C., Liu, X., Martinez, A., & Raudenbush, S. (2013). *Optimal design plus empirical evidence*. Retrieved February 14, 2018, from <https://hlmssoft.net/od/>
- Woods, C. (2007). Researching and developing interdisciplinary teaching: Towards a conceptual framework for classroom communication. *Higher Education*, 54(6), 853–866.

HIGHLIGHTS:

The pedagogical effectiveness of a computer simulation of an enzyme kinetics experiment was evaluated. Hierarchical linear modeling analysis of posttest assessments showed that learning was improved significantly by employing a combination of physical and simulated experiments. The data support the wider use of computer simulations as learning tools in laboratory courses.