Factors Mediating Learning and Application of Computational Modeling by Life Scientists

Aasakiran Madamanchi
Department of Agricultural & Biological Engineering,
Purdue University
West Lafayette, Indiana, USA
Email: amadaman@purdue.edu

Monica E. Cardella
School of Engineering
Purdue University
West Lafayette, Indiana, USA
Email: mcardell@purdue.edu

James A. Glazier
Department of Intelligent Systems Engineering
Biocomplexity Institute
Indiana University
Bloomington, Indiana, USA
Email: glazier@indiana.edu

David M. Umulis
Department of Agricultural & Biological Engineering,
Purdue University
West Lafayette, Indiana, USA
Email: dumulis@purdue.edu

***Abstract*—This Work-in-Progress paper in the Research Category uses a retrospective mixed-methods study to better understand the factors that mediate learning of computational modeling by life scientists. Key stakeholders, including leading scientists, universities and funding agencies, have promoted computational modeling to enable life sciences research and improve the translation of genetic and molecular biology high-throughput data into clinical results. Software platforms to facilitate computational modeling by biologists who lack advanced mathematical or programming skills have had some success, but none has achieved widespread use among life scientists. Because computational modeling is a core engineering skill of value to other STEM fields, it is critical for engineering and computer science educators to consider how we help students from across STEM disciplines learn computational modeling. Currently we lack sufficient research on how best to help life scientists learn computational modeling.**

To address this gap, in 2017, we observed a short-format summer course designed for life scientists to learn computational modeling. The course used a simulation environment designed to lower programming barriers. We used semi-structured interviews to understand students' experiences while taking the course and in applying computational modeling after the course. We conducted interviews with graduate students and post- doctoral researchers who had completed the course. We also interviewed students who took the course between 2010 and 2013. Among these past attendees, we selected equal numbers of interview subjects who had and had not successfully published journal articles that incorporated computational modeling. This Work-in-Progress paper applies social cognitive theory to analyze the motivations of life scientists who seek training in computational modeling and their attitudes towards computational modeling. Additionally, we identify important social and environmental variables that influence successful application of computational modeling after course completion. The findings from this study may therefore help us educate biomedical and biological engineering students more effectively.

Although this study focuses on life scientists, its findings can inform engineering and computer science education more broadly. Insights from this study may be especially useful in aiding incoming engineering and computer science students who do not have advanced mathematical or programming skills and in preparing undergraduate engineering students for collaborative work with life scientists.

Primary Topic— Approaches to Interdisciplinary Education. Secondary Topics: Engineering Education Research; Discipline Specific Issues: Bioengineering and/or Biomedical Engineering.

I. INTRODUCTION

While high-throughput -omic technologies have produced unprecedented quantities of biomedical data, these data have yet to translate into significantly better clinical outcomes. After decades of improvement, U.S. life expectancy has declined, and development costs for a new drug now exceed 1 billion dollars [1, 2]. A broad cross-section of stakeholders including luminaries in the field, leading universities and national funding agencies have called for 'convergence' between disciplines and between science and engineering approaches to improve life-science research productivity [3]. To date, most initiatives for promoting convergence have either recruited engineers into the life sciences or created selective interdisciplinary training programs at elite universities [4, 5]. Efforts to provide interdisciplinary engineering education to the large number of practicing life scientists and life science students have been more limited; often restricted to boot camps in programming or other computational skills [6]. An alternative approach has been the creation of software tools to facilitate adoption of computational methodologies developed in other disciplines by traditionally trained life scientists. In the realm of computational modeling, several groups have produced platforms that seek to help scientists with limited programming or quantitative backgrounds to develop mechanistic computational models of biological systems.

Intense interest in promoting convergence in the life sciences has not led to much rigorous scholarship on how best to promote interdisciplinary engineering approaches. Recent work by Feldon et al., reports that short-format courses have null effect on scholarly productivity in life sciences PhD students [7]. Research by economists and economic historians suggests that broad dissemination of training into the workforce promotes the translation of technological advances into productivity more than investment in selective research centers [8-9]. This Work-in-Progress paper in the research category of 'Frontiers in Education 2018', addresses a major gap in the literature by initiating a rigorous quantitative and qualitative investigation of the factors that mediate learning of computational

NSF 1720625, "Network for Computational Nanotechnology – Engineered nanoBIO node," (2017/22)
NIGMS R01 GM122424, "Competitive Renewal of Development and Improvement of the Tissue Simulation Toolkit," (2017/21)

modeling by life scientists who attend a short-format training course in the use of a computational modeling platform.

II. Theoretical Framework & Research Questions

A. *Theoretical Framework*

Social cognitive theory (SCT) posits that behavior, environmental, and individual factors mutually interact to determine human learning [10]. In education-related studies environmental factors often include social norms and expectations [11]. SCT learning theory applies to many areas of research. Education research grounded in SCT has included studies of the learning and use of new technology-intensive approaches in many contexts, including the adoption of e-learning systems, teacher training in technology use, and nursing education in the use new technologies [12-14].

B. *Research Questions*

To better understand how to improve the practice of teaching engineering methods to life scientists, we ask the following two research questions:

- What are the medium-term outcomes for attendees of a short-format training course on computational modeling within a specific accessible software platform?
- What are the attitudinal, social and structural factors that mediate learning of computational modeling among life scientists?

III. Methods

A. *Research Context*

We studied the outcomes and experiences of attendees of an annual short-format computational modeling course called the CompuCell3D User-Training Course. CompuCell3D is an NIH- and NSF-funded software tool designed to facilitate the development of computational models by life scientists [15, 16]. The CompuCell3D User-Training Course has been conducted annually for thirteen years.

B. Data Collection & Analysis

1) Bibliometric Analysis: Using pubmed searches we analyzed scholarly output of attendees who participated in the course from 2010-2013. We also analyzed a control group of non-attending life scientists from an interest mailing list.

2) Qualitative Interviews: We conducted individual semi-structured interviews of course attendees. We interviewed an equal number of 'successful' attendees who published computational modeling work after course attendance and 'unsuccessful' course attendees who did not publish computational modeling articles. Additionally, we interviewed attendees of courses held between 2013 and 2017 to gain a more recent perspective. We conducted interviews via Skype and audio-recorded them. We transcribed the recordings and coded them using the 'long table method.' We then compared emergent themes from this analysis with our research questions.

IV. Preliminary Findings

Here we share the preliminary results from our quantitative and qualitative studies.

A. *Quantitative Analysis*

There were 126 unique attendees of the CompuCell3D User Training Course between 2010 and 2013. Attendees primarily consisted of graduate students and postdoctoral researchers, with a smaller number of undergraduates and senior scientists. The attendees represent a broad range of US-based and international institutions.

In this cohort, 27 (21%) participated as authors on at least one peer-reviewed publication using computational modeling after course attendance. We did not conduct regression analysis to identify factors that mediate publication success among course attendees, due to the limited number of successful cases. However, in the overwhelming majority of successful cases, 21 out of 27 (78%), attendees came from institutions that sent more than one attendee to the course (either in the same year or in the year immediately preceding or following). The difference in success between the cohorts from institutions/labs with single and multiple course attendees, was significant ($p < .01$) via Chi Square testing. This comparison was independent of scientific success broadly defined, since attendees who eventually published a computational modeling paper, and those who did not, had roughly similar levels of total publications per attendee.

As a comparison we also analyzed the publication history of individuals subscribed to an email interest list for the CompuCell3D modeling platform. Of 37 individuals subscribed to the email list we were able to identify 29 as life scientists. We found that 4 of those 29 (14%) later published a journal article using computational modeling. Course attendees therefore seem to have published in computational modeling at a greater rate than a similarly interested group of non-attendees. However, we have not assessed this difference statistically, because we have not yet determined whether the email-list cohort and the course attendee cohort have comparable educational backgrounds and career stages.

B. Interview data

1) Theme I: Mentor commitment to computational modeling

A major theme in interviews with both 'successful' and 'unsuccessful' attendees was the importance of the faculty mentor or principal investigators' commitment to computational modeling. Attendees who published computational modeling articles frequently reported that their mentor/PI encouraged their modeling efforts. Interviewees described these mentor/PIs as viewing modeling work as being as important as 'wet lab' or experimental work. In contrast, the attendees who never published computational models suggested that a major impediment was that their mentor/PI viewed computational modeling as a potential waste of time or resources. Several interviewees from this cohort reported that their mentor/PIs discouraged them from doing modeling work during the day. One interviewee stated 'it became clear pretty quickly that modeling was something I needed to do on my own time, and that when my boss was around I needed to be doing experiments at the bench.'

2) Theme II: Opportunity cost

A theme that surfaced among those who did not publish computational modeling papers, was a view that computational modeling competed rather than complemented the use of other molecular biology technologies. These interviewees often compared computational modeling with technologies such as CRISPR or Flow Cytometry. They described computational modeling in the CompuCell3D platform as having a 'steep learning curve,' while they viewed other technologies as providing the opportunity for 'rapid data generation.' One interviewee described abandoning computational modeling in favor of developing expertise in a new wet lab technique:

> *"I was kind of bogged down and then [new molecular biology technique] came around and it was like OK let's do that. And that was just simpler. Like I knew how to make that work and we got a lot of data really quickly, and so we just jumped on it and kind of let this stuff go..."*

Conversely, interviewees who published computational modeling papers viewed modeling as a fundamental part of the scientific process. These interviewees considered adopting other new technologies as a mechanism to generate data for modeling rather than as a potential replacement for modeling. We are not certain whether this differing view was a cause of persistence in computational modeling or reflected the effect of later academic specialization in modeling.

3) Theme III: Motivation for computational modeling.

The vast majority of interviewees reported that they were primarily motivated by the potential for computational modeling to provide scientific insight. Very few interviewees expressed interest in developing coding or quantitative skills as a mechanism to gain skills for careers outside science. However, several interviewees from traditional experimental biology backgrounds suggested that a secondary motivation for their interest in computational modeling, was to raise the impact of their research. These interviewees viewed modeling expertise as a potential avenue for 'signaling' the scientific quality of their work. Interviewees who worked in labs that already participated in computational modeling seldom reported such secondary motivations. Instead interviewees from 'modeling' backgrounds seemed to view computational modeling as a routine part of science rather than as a way to stand out or differentiate themselves.

4) Theme IV: Computational Thinking not Coding is the Central Challenge of Computational Modeling

We were particularly interested to identify what attendees who did not publish computational modeling papers viewed as the most significant challenges associated with applying computational modeling. Interestingly, regardless of their level of programming or math background, interviewees rarely reported technical issues as sources of difficulty for computational modeling within the CompuCell3D modeling environment. Most interviewees reported that the CompuCell3D User-Training Course had provided sufficient instruction to allow them to replicate existing computational models.

However, interviewees who did not publish computational models often reported that their primary difficulty was in understanding how to translate their research into a computational model. One interviewee stated *"...I was making progress, but there was still this disconnect. Like, I could make stuff move on the screen but how to translate that into a hypothesis was tough."*

Several interviewees expressed confusion about how to break complex biological pathways or cellular behaviors into specific mechanisms that could be modeled. Others reported uncertainty about how coding specific biological actions at the molecular or cellular level could be manifest as higher-level behavior at the tissue level. These issues closely align with difficulty in 'problem decomposition,' and 'abstraction,' two key components of Computational Thinking [17, 18].

5) Theme V: 'Hidden Curriculum'

A theme that surfaced in interviews with attendees who did not publish a computational model was confusion about prevailing norms and practices in computational modeling. For these interviewees ambivalence over technical choices in model parameterization, or mathematical methods in part of their model proved a major obstacle. Several interviewees expressed frustration that information about how to decide between options was not available in the manuals, or other publications. This (missing) information seems to operate analogously to a 'hidden curriculum.' Interviewees who successfully published computational modeling papers often indicated that they received this information informally from their mentor/PI or more senior colleagues.

V. Discussion & Future Work

Our quantitative work indicates that a significant number (21%) of the attendees of the CompuCell3D User Training Course went on to publish a computational modeling paper. However, the overwhelming majority of these success cases (78%) came from research labs or universities that sent multiple attendees to the Course, perhaps indicating strong PI commitment to computational modeling. These results align closely with our preliminary qualitative findings that environmental or structural factors such as mentor/PI commitment to computational modeling and the informal transmission of modeling practices are critical for successful learning and application of computational modeling.

Our preliminary qualitative results provide important insight into the experiences of life-sciences trainees who attempt to learn computational modeling. Perhaps most interesting is that computational thinking rather than coding or mathematical background seems to be the key limiting factor in learning computational modeling. This suggests that, to be truly successful, even computational modeling tools which successfully lower the technical barriers to computational modeling (such as the CompuCell3D modeling environment), must be paired with resources to help life scientists with computational thinking more broadly. Similarly, the computational tools community must establish curated resources on the 'hidden curriculum' of computational modeling. For example, life scientists new to modeling must be connected with guides to emerging methodologies for qualitative data parameterization [19-22].

The developers of the CompuCell3D modeling environment have already made several changes to their training resources to reflect our preliminary findings. First, they have added a user forum to their website to provide a public mechanism for discussion of problems that arise while developing computational models. They hope that this user forum will make the 'hidden curriculum' of computational modeling more accessible to those outside computational modeling labs. Additionally, they hope that these user forums will provide a virtual 'community of practice' which will help attendees who are not in computational modeling labs form peer-to-peer support structures

Our findings do not address the findings of Feldon et al. [7], concerning the (lack of) efficacy of short courses in increasing overall scholarly output, but they do support their observation that long-term engagement with a subject is more important to success than a short week-long intervention. Our interpretation of the data in this study is that structural factors such as mentor/PI commitment and the presence of other course attendees in a research group are instrumental in facilitating the long-term engagement with computational modeling needed for meaningful learning. Our finding of relevance for both individual cognitive factors (computational thinking) and structural factors (PI commitment, and peer course attendance) is consistent with SCT.

Critical limitations to our study to consider when evaluating our findings include the self-selection of our interview participants. All of our interview participants first self-selected to attend the CompuCell3D User-Training Course in computational modeling, and then agreed to an interview. Consequently, our interview sample demonstrated significant interest in computational modeling. Our future work will investigate the attitudes and knowledge barriers of life scientists who have not made a prior investment in computational modeling. Our focus on a single modeling platform limited our sample size preventing regression analysis to identify factors that mediate success in learning computational modeling.

This Work-in-Progress paper represents the initial stage in our work to better understand how to support life scientists in gaining interdisciplinary education in 'convergence' approaches such as computational modeling. In future studies we hope to complement these findings by evaluating key effects of student background on learning and using interdisciplinary engineering approaches.

ACKNOWLEDGMENT

The authors thank Dr. James Sluka for his helpful comments.

REFERENCES

[1] K. D. Kochanek, S. L. Murphy, J. Q. Xu, E. Arias. Mortality in the United States, 2016. *NCHS Data Brief, no 293*. Hyattsville, MD: National Center for Health Statistics. 2017.

[2] J. W. Scannel, A. Blanckley, H. Boldon, and B. Warrington, "Diagnosing the decline in pharmaceutical R&D efficiency," *Nature Reviews Drug Discovery*, vol. 11, no. 3, pp. 191–200, Mar. 2012.

[3] National Research Council (U.S.)., Convergence: Facilitating Transdisciplinary Integration of Life Sciences, Physical Sciences, Engineering, and Beyond. Washington, D.C.: National Academies Press, 2014.

[4] A. Arnold and M. Greer, "Convergence: a transformative approach to advanced research at the intersection of life, physical sciences and engineering and enhanced university-industry partnerships," *Journal of Translational Science*, vol. 1, no. 3, 2016.

[5] P. Sharp and S. Hockfield, "Convergence: The future of health," *Science*, vol. 355, no. 6325, Sep. 2017.

[6] M. I. Stefan, J. L. Gutlerner, R. T. Born, and M. Springer, "The Quantitative Methods Boot Camp: Teaching Quantitative Thinking and Computing Skills to Graduate Students in the Life Sciences," *PLOS Computational Biology*, vol. 11, no. 4, 2015.

[7] D. F. Feldon, S. Jeong, J. Peugh, J. Roksa, C. Maahs-Fladung, A. Shenoy, and M. Oliva, "Null effects of boot camps and short-format training for PhD students in life sciences," *Proceedings of the National Academy of Sciences*, vol. 114, no. 37, pp. 9854–9858, 2017.

[8] M. Kelly, J. Mokyr, and C. Ó. Gráda, "Precocious Albion: A New Interpretation of the British Industrial Revolution," *Annual Review of Economics*, vol. 6, no. 1, pp. 363–389, 2014.

[9] E. Brynjolfsson, D. Rock, and C. Syverson, "Artificial Intelligence and the Modern Productivity Paradox: A Clash of Expectations and Statistics," 2017.

[10] Bandura, A., "Social Foundations of Thought and Action: A Social Cognitive Theory", Prentice Hall, Englewood Cliffs, NJ, 1986.

[11] R. W. Lent, M. J. Miller, P. E. Smith, B. A. Watford, K. Hui, R. H. Lim, "Social cognitive model of adjustment to engineering majors: Longitudinal test across gender and race/ethnicity", *Journal of Vocational Behavior*, vol. 86, pp. 77-85, 2015.

[12] Wang, S.-L. and P.-Y. Wu, "The Role of Feedback and Self-Efficacy on Web-Based Learning: The Social Cognitive Perspective," Computers & Education, 51, 2008, pp. 1589-1598.

[13] H. Burke and L. Mancuso, "Social Cognitive Theory, Metacognition, and Simulation Learning in Nursing Education," *Journal of Nursing Education*, vol. 51, no. 10, pp. 543–548, 2012.

[14] L. D. Rubenstein, L. M. Ridgley, G. L. Callan, S. Karami, and J. Ehlinger, "How teachers perceive factors that influence creativity development: Applying a Social Cognitive Theory perspective," *Teaching and Teacher Education*, vol. 70, pp. 100–110, 2018.

[15] M. H. Swat, S. D. Hester, A. I. Balter, R. W. Heiland, B. L. Zaitlen, and J. A. Glazier, "Multicell Simulations of Development and Disease Using the CompuCell3D Simulation Environment," *Methods in Molecular Biology Systems Biology*, pp. 361–428, 2009.

[16] M. H. Swat, G. L. Thomas, J. M. Belmonte, A. Shirinifard, D. Hmeljak, and J. A. Glazier, "Multi-Scale Modeling of Tissues Using CompuCell3D," *Methods in Cell Biology Computational Methods in Cell Biology*, pp. 325–366, 2012.

[17] S. Grover and R. Pea, "Computational Thinking in K–12," *Educational Researcher*, vol. 42, no. 1, pp. 38–43, 2013.

[18] J. M. Wing, "Computational thinking and thinking about computing," *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, vol. 366, no. 1881, pp. 3717–3725, 2008.

[19] M. Pargett and D. M. Umulis, "Quantitative model analysis with diverse biological data: Applications in developmental pattern formation," *Methods*, vol. 62, no. 1, pp. 56–67, 2013.

[20] M. Pargett, A. E. Rundell, G. T. Buzzard, and D. M. Umulis, "Model-Based Analysis for Qualitative Data: An Application in Drosophila Germline Stem Cell Regulation," *PLoS Computational Biology*, vol. 10, no. 3, 2014.

[21] J. Hengenius, M. Gribskov, A. Rundell, and D. Umulis, "Making models match measurements: Model optimization for morphogen patterning networks," *Seminars in Cell & Developmental Biology*, vol. 35, pp. 109–123, 2014.

[22] D. M. Umulis and H. G. Othmer, “The Role of Mathematical Models in Understanding Pattern Formation in Developmental Biology,” *Bulletin of Mathematical Biology*, vol. 77, no. 5, pp. 817–845, Apr. 2014.