

Article

Computer Literacy for Life Sciences: Helping the Digital-Era Biology Undergraduates Face Today's Research

Tomasz G. Smolinski

Department of Biology, Emory University, Atlanta, GA 30322

Submitted March 24, 2010; Revised May 17, 2010; Accepted June 23, 2010

Monitoring Editor: John Jungck

Computer literacy plays a critical role in today's life sciences research. Without the ability to use computers to efficiently manipulate and analyze large amounts of data resulting from biological experiments and simulations, many of the pressing questions in the life sciences could not be answered. Today's undergraduates, despite the ubiquity of computers in their lives, seem to be largely unfamiliar with how computers are being used to pursue and answer such questions. This article describes an innovative undergraduate-level course, titled Computer Literacy for Life Sciences, that aims to teach students the basics of a computerized scientific research pursuit. The purpose of the course is for students to develop a hands-on working experience in using standard computer software tools as well as computer techniques and methodologies used in life sciences research. This paper provides a detailed description of the didactical tools and assessment methods used in and outside of the classroom as well as a discussion of the lessons learned during the first installment of the course taught at Emory University in fall semester 2009.

INTRODUCTION

The ubiquity of computers in our lives is undisputed. For every modern-age college student it is virtually a must to own or have access to a computer for him or her to succeed in course work. Almost every assignment in today's classroom requires some computer work, be it researching literature, performing calculations, or writing a report. In addition, many young people take advantage of a variety of online communication resources, such as e-mail and instant messengers, or social websites such as MySpace, Facebook, and Twitter, and thus are familiar with many types of computer software. Given this image, one might think that the undergraduates of today are naturally ready to dive into the world of, presently highly computerized, scientific research. Unfortunately, this does not seem to be true. Although many students are far more adaptable to computers and their applications than older

generations, they seem daunted by more sophisticated uses, such as data manipulation and analysis for the purpose of scientific research (Goodall, 2008). It is widely believed that this is at least partially due to the insufficient preparation of the students in terms of basic mathematical skills, and especially logic and statistics, which are indispensable in life sciences research (Rao, 2008).

This article describes an innovative undergraduate-level course, a four-credit-hour elective open to all life science majors, titled Computer Literacy for Life Sciences, which aims at partially remedying this situation. Although it is not easy to fill in all the gaps resulting from many years of underpreparation in the area of quantitative skills development, this course is intended to provide the students with some basic proficiency needed for scientific research in the digital era.

The overall purpose of Computer Literacy for Life Sciences is for students to develop a hands-on (and brains-on!) working experience in using standard computer software tools as well as computer techniques and methodologies used in life sciences research. Researchers should be familiar with the computer's operating system and its main and most useful applications. They must know how to effectively manipulate documents and data files, including those downloaded from libraries and data repositories available on the World Wide Web (WWW). Scientists must be able to efficiently manage and query the data resulting from wet lab

DOI: 10.1187/cbe.10-03-0050

Address correspondence to: Tomasz G. Smolinski (tsmolin@emory.edu).

© 2010 T. G. Smolinski. CBE—Life Sciences Education © 2010 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>).

experiments as well as those derived from computer simulations. Such data are then subject to statistical analysis, or other means of data mining, to answer the underlying scientific questions. Finally, researchers should be able to present their discoveries in the form of reports or scientific articles and audiovisual presentations.

To fulfill all of the above-mentioned objectives, the students in the Computer Literacy for Life Sciences course are mentored throughout the process of scientific pursuit, which is implemented in the form of several group virtual research projects to be completed with the use of state-of-the-art computer techniques and applications. Each virtual project includes the following components: background literature research, data manipulation, formulation of scientific questions and verifiable hypotheses, statistical data analysis, and presentation of scientific discoveries.

Several aspects of this course, such as the idea of semester-long virtual research projects, modules, and skills-checks as well as the grading scheme, were based on a somewhat similar course developed at San Jose State University (MacDonald *et al.*, 2010). Many of the experiences derived from a course on information literacy in biology education described by Porter (2005) also were very useful in the preparation of Computer Literacy for Life Sciences.

The first installment of Computer Literacy for Life Sciences was taught at Emory University in fall semester 2009. The enrollment in the course was 12 highly motivated students from two majors: biology, and neuroscience and behavioral biology (NBB).

COURSE STRUCTURE AND ORGANIZATION

The course was taught once a week in a 3-h-long session consisting of a short (up to 30 min) lecture on the main concepts of the particular session followed by a 2.5-h hands-on computer laboratory under the supervision of, and with help from, the instructor.

The course was taught in a modular manner, with each module representing a major stage of a research project, as outlined below. The students were divided into several groups, each working on a specific virtual research project, selected from a list of projects provided by the instructor. In each module, the students were to acquire a set of skills that would allow them to complete the specific phase of their project and traverse to the next step, until a successful conclusion of their pursuit. Each of the stages was concluded by a short group presentation to the class, as well as an in-class practical competency exam (described below).

The presentations were designed to give each group an opportunity to update the class on their progress, as well as to discuss any problems or difficulties they might have encountered. To avoid a situation in which only some group members are actually performing all the work and thus learning the material, each member of the group was expected to partake in the presentations.

Assessment

The students were expected to complete readings, cooperate in the group project, prepare presentations for each of the course modules, participate in class discussions, and successfully pass the competency exams. Table 1 presents the

Table 1. Grade computation scheme

	Points
Group presentations (three at 30 points each)	90
Competency exams (three at 30 points each)	90
Final presentation	135
Final article	135
Level of effort made, interest, participation, and evidence of growth	50
Total	500

grade computation scheme used, and a discussion of the specific subareas of assessment follows.

Group Presentations. As mentioned, the module group presentations were implemented to allow each team to provide an update on their progress as well as to have the opportunity to discuss any problems or difficulties they might have encountered, so that the whole class could brainstorm, learn, and benefit from the experiences of others. The presentations were graded based on a rubric that included the following categories: organization, content knowledge, supporting material, mechanics, and delivery. In the module 1 group presentations, only the instructor provided the scores, but all other presentations also were scored by the other classmates. The purpose for this was twofold: 1) the grading students provided more feedback and thus increased the perceived fairness of the grades (all the grades, both from the instructor and the students, were included in the final score); but more importantly, 2) they were forced to think about the aspects of the presentations that were included in the rubric and inadvertently reflect on their own performance in those categories.

Competency Exams. Similarly to the group presentations concluding each of the first three modules, there were three competency exams administered during the semester, to be completed by each student individually. The competency exams were based on skills-check handouts provided for each module by the instructor, before the exam. Each handout contained a list of tasks that a student was expected to be able to complete after attaining the learning objectives of a given module. Only the tasks included in a skills-check were to be included on an actual exam, although some variations (e.g., different input files, different variables) were to be expected. Thus, if the students mastered the material on a given skills-check, they should have no trouble with the exam to follow. The skills-checks were practiced by the students in class during special review sessions as well as in their own time.

Final Presentation. Toward the end of the term, all groups had a chance to orally present their discoveries in the form of a final team presentation. This presentation, in contrast to the three prior presentations, was to provide an overview of the complete project, similarly to what might be presented at a scientific conference. The students were expected to discuss their entire experience, without simply repeating all of the material from the previous presentations, and concentrate on their hypotheses and how they attempted to verify them. Chronologically, the presentations were scheduled before the final

Table 2. Modules and their corresponding main learning goals

Module	Main learning objectives
	After attaining the learning objectives of the module, students should be able to
1. Operating systems and Web-based project repositories; background literature research	Use computer operating systems to effectively manage data files and other documents Build and publish an online research project repository (i.e., website) Perform background literature searches using various online resources
2. Databases and data querying	Use Microsoft Access to import, browse, sort, search, and export data files Use the Structured Query Language (SQL) to explore and summarize data Use the SQL to extract subsets of data that are of interest to the underlying research project
3. Data analysis	Use Microsoft Excel to perform basic statistical data analyses using descriptive statistics and plotting methods Use SPSS to perform more complex univariate, as well as multivariate, analyses Form verifiable scientific hypotheses, translate them into statistical hypotheses, and test their significance
4. Presentation of scientific discoveries	Interpret and explain the meaning of the results of statistical analyses and tests Present an overview of a research project, including the underlying questions, approaches, methodologies, and results of analyses in the form of an audiovisual presentation Describe a research project in the form of a co-authored journal-like article

article, so that the students could obtain valuable feedback to be included in the final product.

Final Article. At the conclusion of the semester, each group had to summarize their projects and discoveries in the form of a short journal-like coauthored article. The students were given proper instruction and resources for how to prepare such a publication beforehand. The articles were to contain a discussion of the underlying big-picture problem, the state of current research in the field, a presentation of the group's ideas and hypotheses, a description of all the methods used throughout the project, and a summary of the main discoveries resulting from the entire virtual research project.

Level of Effort Made, Interest, Participation, and Evidence of Growth. The last category included in the grading scheme was a somewhat subjective method to reward the students for their added effort, realized in the form of numerous extra-credit assignments, and overall interest in the subject matter.

Based on the above-described scheme, a standard percentage grading scale rounded to the nearest whole number (by rounding up from 0.5) was used to calculate the final grades with "A+" = 97–100%, "A" = 92–96%, "A–" = 89–91%, B+ = 87–88%, B = 82–86%, and so on.

Modules

The course was divided into four separate modules, each having a different set of overarching learning objectives. The modules and their corresponding main learning goals are presented in Table 2, and a detailed description of their contents follows.

Module 1: Operating Systems and Web-based Project Repositories; Background Literature Research. In the beginning of the course, the students had a chance to familiarize themselves with the two operating systems widely used in the scientific community, Windows (Microsoft, Redmond,

WA) and Unix/Linux (developer Linus Torvalds and thousands of collaborators), with the emphasis on effective file management. Both operating systems provide a collection of tools that can be applied to scientific pursuits and they also often complement each other. Therefore, it is important to have some degree of mastery over both and to be able to make them interact with each other.

Also at this stage, the students were asked to select a virtual research project from the list provided by the instructor to be pursued by them and their teammates for the remainder of the course. There were 14 project proposals from the following three areas: basic biology, biomedicine, and neuroscience (due to the obligation imposed by the NBB major, which required a neuroscience component in the course). To be on the list, a topic had to be supported by a relatively large number of works in the literature and have underlying data available. The data sets, and the corresponding principal publications in addition to the instructor's own neuroscience-related projects, were obtained from WWW resources, such as the University of California Irvine Machine Learning Repository (<http://archive.ics.uci.edu/ml>). Table 3 lists the 14 proposed virtual research project themes.

Not surprisingly, because most of the 12 students enrolled in the course were premed students, the most popular topics were among the biomedicine-related topics. Actually, based on the students' selections, four groups were formed, each dealing with one of the biomedicine topics listed in Table 3.

After the selections had been made, the students were asked to perform background literature research on their topics. The underlying background information about a scientific inquiry is usually acquired by using computer tools to research the literature or the WWW. The Internet is also abundant with existing and publicly available data sets derived from experiments performed by others in the scientific community that not only can provide invaluable insight into the inquiry of interest but also may be used for further

Table 3. Proposed virtual research project topics

Area	Topic
Basic biology	Prediction of the cellular localization sites of proteins in yeast
	Prediction of the protein localization sites in <i>Escherichia coli</i> bacteria
	Analysis of the <i>E. coli</i> promoter gene sequences (DNA) with partial domain theory
	Prediction of the protein secondary structure based on a sequence of amino acids
	Classification of mushrooms based on their physical characteristics
Biomedicine	Classification of soybean diseases based on plants' physical characteristics
	Analysis of the cell nuclei from digitized images of a FNA of a breast mass to determine malignancy of cancer
	Breast cancer surgery survival rates analysis
	Diagnosing heart abnormalities based on cardiac SPECT images
Neuroscience	Diagnosing heart arrhythmia based on clinical patient data
	Analysis of cortical evoked potentials (micro-electroencephalogram [EEG]) from rats 1: vibrissa stimulation
	Analysis of cortical evoked potentials (micro-EEG) from rats 2: auditory stimulation
	Analysis of multidimensional parameter spaces of invertebrate pace-making neurons
	Analysis of relationships between ionic currents in spiking neurons

analysis complementing previous studies. Topics such as browsing and searching online scientific databases as well as retrieving and storing the appropriate information were covered in this part of the course. The students learned how to use this kind of information to form their own scientific questions and delineate plans for answering them.

At this stage, in addition to the Microsoft Windows operating system, which was used in virtually all other tasks in the course, the students were also introduced to the Unix/Linux-type operating systems. Using the popular Tectia SSH Secure Shell application, they learned how to connect to a Linux-based server provided by Emory University, transfer files between their local Windows-based workstation and the server, and effectively manipulate those files on the remote server. These skills were not only important for completion of the final part of this module (described in the next paragraph) but also are, in general, fundamental if one plans to perform computational simulations, often done on Unix/Linux supercomputers, as the basis for life sciences research.

Using their freshly acquired knowledge about the two operating systems and their interconnectivity, the students learned how to create a simple website to serve as a repository of their project's records and documentation. They learned the basics of HyperText Markup Language (HTML)—a principal programming language used in web-page development—as well as some aspects of website design provided by the popular Web development application, Macromedia (currently Adobe Systems, Mountain View, CA) Dreamweaver. The main purpose of this task was threefold: the students were not only expected to 1) create a resource to store all the materials pertaining to their projects (accessible to them from wherever they were), and 2) learn the basics of website development, but also to 3) have a smooth introduction to teamwork and begin cooperating with their group-mates.

It is also important to point out that the students, as opposed to using the Dreamweaver as the only method of webpage development, were expected to use HTML for a reason. Most of the students in the class had never used a programming language before and learning HTML was a method for easing them into the concept of coding by allow-

ing them to program something that would be simple, yet palpable and exciting. This experience also was expected to make the introduction of the other programming language used in the course, SQL (described below), easier.

Module 2: Databases and Data Querying. Research studies can be performed in the laboratory setting as wet experiments, or they can take the form of computer simulations, but regardless of the method, the results are rather large amounts of computerized data. Such data can be effectively stored in database management systems that allow for easy access and retrieval. In this section of the course, the students learned the basics of relational database management systems (RDBMS) along with how to efficiently query such systems to obtain the desired portions of the data. Specifically, Microsoft Access, with its Structured Query Language (SQL; the most commonly used data querying language) capabilities was utilized for this purpose.

The students also learned how to upload data retrieved from the WWW in the previous module into the RDBMS and then effectively query the system to extract the relevant parts of the data. Finally, the students became skilled at exporting such subsets of data to other types of software for the purpose of data analysis.

Most of the material presented in this module was acquired by the students in the form of tutorials and in-class and/or take-home assignments (referred to as the tutorial/assignment model). A tutorial includes a collection of specific instructions to complete a specific task. For example, some of the tutorials included in this module were as follows: creating a new, blank database in Microsoft Access 2007; uploading data into Access from external sources; or sorting, searching, and filtering data in Access. In the SQL portion of the module, some of the tutorials were as follows: simple queries using SQL; data extraction using SQL; and data summarization using SQL. As mentioned, the tutorials list the specific steps needed to complete a given task (e.g., upload data to Access). Initially, the students were given the tutorials and asked to complete the steps individually, under the supervision and with help of the instructor, as needed. However, the students insisted, and it turned out to be more efficient, on completing the tutorials as a group, to make sure that everybody is "on the same page" and all

questions are answered. A few examples of this module's tutorials can be found at www.biology.emory.edu/research/Prinz/Tomasz/cl_for_ls.

In addition to tutorials, the students had to complete assignments that were either take-home or to be performed in-class. The purpose of the assignments was for the students to extrapolate from the knowledge and skills they acquired by completing the tutorials to new, yet similar, circumstances and problems. For example, based on several previously completed tutorials, the students were asked to upload a new data set into Access, create a series of cooperating SQL queries to extract specific information about the data, and export that information from Access to a text file. The assignments were to be performed either individually, in pairs, or in groups, depending on the particular material. An example of such an assignment can be found at www.biology.emory.edu/research/Prinz/Tomasz/cl_for_ls.

After attaining the learning objectives of this module, the students should be able to import the data of interest into Access from an external source, manipulate them to extract particular subsets of the data, and export those pieces of data into external data files, to be used for further analyses.

Module 3: Data Analysis. Calculations and manipulations needed for data analysis can be performed using various statistical programs and spreadsheets. In this part of the course, the students experimented with two different approaches, using two different programs, to statistical data analysis. The first approach was based on Excel (Microsoft) to extract and graph basic statistics from data samples. The students learned how to get to know the data and how to summarize their essential characteristics, all by using the standard and relatively simple functions provided by Excel.

The second approach used a more sophisticated statistical data analysis software package, SPSS (SPSS, Chicago, IL), to investigate interactions between variables in the data. The students learned how to examine interdependencies between variables, by using various measures and methods such as correlations and regression analysis.

In this module, the same tutorial/assignment approach was used. In a series of tutorials, the students learned several fairly basic statistical methods used for data analysis. In Excel, the students were introduced to performing simple descriptive statistical analysis (e.g., in terms of means, medians, modes, standard deviations), as well as creating basic plots to graphically describe the data.

In SPSS, the students applied more sophisticated statistical techniques including: box-and-whisker analysis, correlations, *t* tests, ANOVA, and regression analysis. The purpose here was by no means to turn the students into statistical analysis experts, but to make them capable of choosing and performing appropriate statistical analyses and tests (out of this quite limited arsenal of tests) using one of the two introduced software packages, and to be able to understand the meaning of the results. Therefore, a special emphasis was placed on stating testable hypotheses and the interpretation of a hypothesis testing result.

In several assignments that accompanied the tutorials, students were asked to brainstorm and propose several hypotheses pertaining to their virtual research projects, and test those hypotheses using Excel or SPSS.

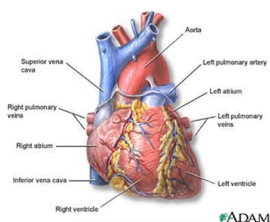
Module 4: Presentation of Scientific Discoveries. Finally, in the last part of the course, the students had a chance to summarize their inquiries and discoveries in the form of a short (4–5 pages) coauthored journal-like article and a final class presentation. The articles, as well as the presentations, were to be organized into sections corresponding to the modules of the course. Each section contained a description of the methods used and a discussion of how they were applied to a particular problem. The articles and the presentations concluded with a summary of the main discoveries resulting from the entire virtual research project.

DISCUSSION

As mentioned, in the first module, students had the opportunity to familiarize themselves with Windows and Unix/Linux operating systems, perform a background literature research on their group virtual topics, and create a simple website to serve as a repository of their project's records and documentation. The module proved to be a success, as not only did the students develop interesting and comprehensive websites for their projects (two examples shown in Figure 1) but also the stage was set for further group cooperation on the virtual research projects.

In the second module, the students were presented with the actual data sets underlying their projects and, utilizing the skills they had learned based on tutorials and assignments, were asked to explore the data and extract the "important" subsets from the data sets. Many of the data sets behind the virtual projects contained superfluous attributes, such as names and IDs, hardly useful for classification. The students were expected to identify such fields and remove them from consideration, which all groups successfully performed by the means of the SQL language. In addition, because some of the fields in the data sets contained missing entries, the students, after researching various methods for dealing with missing data (which was one of their take-home assignments), implemented some of those approaches using SQL.

The third module, concerned with the actual analysis of the data pertaining to the group projects, allowed the students to utilize their newly acquired knowledge of several statistical techniques (e.g., graphing methods, correlations, *t* tests, analysis of variance [ANOVA], and regression analysis) to test and interpret their hypotheses. Based on several tutorials and assignments, the students learned how to perform statistical analyses and tests in either Excel or SPSS, and which of those methods would work best for their particular data set. Because each of the projects was quite different in terms of the characteristics of the underlying data, the outcomes of analyses ranged from a description of a set of quite clear and concise classification rules (Analysis of the cell nuclei from digitized images of a fine needle aspirate [FNA] of a breast mass to determine malignancy of cancer and Diagnosing heart abnormalities based on cardiac single proton emission computed tomography [SPECT] images), through comparisons of different regression methods (Breast cancer surgery survival rates analysis), to the creation of new attributes based on the existing ones (e.g., body mass index, based on other physical attributes of patients) for the purpose of creation of the best possible classifier



Abstract:

A non-invasive way of identifying patients who are at high risk of heart arrhythmia is necessary to start treatment as early as possible. The measurement of T-wave alternans has been found to predict whether patients are at higher risk for heart arrhythmia. The T-wave measures change in T-wave morphology, reflecting the occurrence of action-potential alternans, which promote arrhythmias. "A Supervised Machine Learning Algorithm for Arrhythmia Analysis" uses the T-wave as a component in its algorithm to predict and differentiate among the different types of arrhythmias. The T-wave contributes to the larger picture of the algorithm.

In this paper, they use a VFIS algorithm analyzing a 12-lead ECG. The results are presented in training cases, and each of the 249 attributes is analyzed for its relevancy of a particular attribute that it can cast before. Models like such as this because they enable physicians to make decisions contrary they need to be overseen by specialists. This is a preliminary study.

Extensive research is being conducted to refine computational models.

[Machine learned algorithm for arrhythmia analysis](#)

Breast Cancer Surgery Survival Rates Analysis



Project Description:

In the United States, approximately one in eight women has a risk of developing breast cancer. Over the past years, the survival rate for breast cancer victims has been increasing due to a variety of factors (i.e., better technology, treatment, etc.). Analysis of survival rates is imperative and is being used to devise methods to predict such rates. These methods are of value in medicine and can be used to assist in determining prognosis, developing successful treatment, etc. Several prediction methods have been analyzed, including artificial neural networks, decision tree algorithms, and logistic regression models. Results have indicated that the decision tree is the best predictor, with accuracy around 90%, and artificial neural networks are a close second. However, for these models to be of any use, medical professionals must collaborate and be open to using them. Further research can be conducted on how these rates are correlated with other cancers. We are interested in seeing how these rates can be improved by incorporating more clinically relevant data; however this raises ethical concerns about giving such information to patients.

Resources:

[Predicting Breast Cancer Survivability Using Data Mining Techniques](#)

[Predicting Breast Cancer Survivability: A Comparison of Three Data Mining Methods](#)

[Prognostic and Predictive Factors in Primary Breast Cancer and Their Role in Patient Management: The Nottingham Breast Team](#)

Figure 1. Virtual research project repository websites developed by the students in two selected groups.

(Diagnosing heart arrhythmia based on clinical patient data).

Finally, each group delivered a presentation and returned a comprehensive and interesting article in which they provided an overview of their project, as well as the results of their studies.

As an illustration of the entire virtual research experience, consider the group working on the heart arrhythmia project. After building their online project repository, the students in that group obtained a data set containing 452 patient records in terms of their age, sex, height, and weight, as well as different characteristics of the patients' electrocardiogram (ECG). The data set also included the information about the patients' diagnoses: either normal or one of 15 types of arrhythmia (Güvenir *et al.*, 1997).

The students uploaded the data set into Access and used SQL to explore the data and to perform preprocessing in terms of missing entries or other inconsistencies, to prepare the data set for analysis. Importantly, in addition to the relatively simple data manipulation procedures the students carried out in this step, they already started forming hypotheses as to the importance of specific attributes in terms of the overall diagnosis. For example, based on various combinations of the SQL queries that the members of this group

implemented, they hypothesized that weight and height by themselves may be quite misleading predictors of arrhythmia, but if combined into the body mass index, the diagnosis accuracy could be improved.

In the next step, the students in the arrhythmia group used Excel and SPSS to further analyze their data. They looked more closely at the distributions of values in all the attributes and, based on those analyses, hypothesized that three additional fields (QRS-duration, T-interval, and P-interval, all representing specific characteristics of the ECG) also may be useful for diagnostic purposes. To test their hypotheses, the students performed one-way ANOVA tests and built a multinomial logistic regression model and indeed confirmed the significance of the selected attributes.

It is important to point out that the analyses performed by the students in this group, as well as the other three groups, were based on their own open-ended hypotheses and ideas and were not merely a recreation of other published studies. Although the students had a rather limited selection of data analysis methods at their disposal (due to the time constraints of the course) and did not actively seek other bioinformatics or data mining approaches beyond what was covered in the class, they learned how to choose and apply correct statistical tests to evaluate their own hypotheses.

CONCLUSIONS

The Computer Literacy for Life Sciences course has proved to be a very successful endeavor, as indicated by the students' positive feedback on unofficial surveys and official evaluations, as well as their high performance on assignments and tests. In their anonymous comments, students expressed their appreciation for the tutorial/assignment approach and claimed it made "a tough class easy" and helped them "navigate the unfamiliar material." They also appreciated the "freedom to apply knowledge and experiment with data." Most students enjoyed the group work and hands-on experience.

The most difficult aspect of the course was the timing and length of the sessions. One 3-h-long class-meeting late in the afternoon was very taxing for both the students and the instructor. In the second installment of the course, which was taught again in spring semester 2010, sessions were divided into two 1.25-h class-meetings twice a week. This has proved to be a more appropriate time and material distribution.

Finally, to answer the underlying question of this article, "Are the digital-era undergraduates computer literate enough to face today's research?," based on my observations in the classroom, I would have to draw the conclusion that no, this is not quite the case, because in spite of their familiarity with computers, the students of today seem to be quite helpless in the face of more advanced computer applications, especially when they require some background knowledge of mathematics, statistics, or computer science (as opposed to just computer use). However, because stu-

dents are so familiar with and accustomed to computers, with a little help and guidance, as provided by the Computer Literacy for Life Sciences course, for example, they can achieve proper research-centered computer literacy quite easily.

ACKNOWLEDGMENTS

The support for the development of this course was provided by the Howard Hughes Medical Institute grant 52005873 (to Pat Marsteller, Director of the Emory College Center for Science Education). I also acknowledge the advisory support of the Faculty Institutes for Reforming Science Teaching, Fourth Edition (FIRST IV) program.

REFERENCES

- Goodall, B. (2008) Computer literacy doesn't mean information literacy, report says. *Chron. Higher Educ.*, January 16, 2008.
- Güvenir, H. A., *et al.* (1997) A supervised machine learning algorithm for arrhythmia analysis. In: *Proceedings of the Computers in Cardiology Conference*, 7–10 September 1997, Lund, Sweden, 433–436.
- MacDonald, S. *et al.* (2010). *Biology 5, Computer Literacy in Biology* (lower division biology course at San Jose State University). <http://vireo.biology.sjsu.edu/bio5> (accessed 15 March 2010).
- Porter, J. R. (2005). Information literacy in biology education: an example from an advanced cell biology course. *Cell Biol. Educ.* 4, 335–343.
- Rao, U. (2008) Major affairs: mathematics and CS in biology. *The Cornell Daily Sun*, December 3, 2008.