esa ECOSPHERE

Analysis of high-frequency and long-term data in undergraduate ecology classes improves quantitative literacy

Jennifer L. Klug,[1,†] Cayelan C. Carey,[2] David C. Richardson,[3] and Rebekka Darner Gougis[4]

[1]Department of Biology, Fairfield University, 1073 N. Benson Road, Fairfield, Connecticut 06824 USA
[2]Department of Biological Sciences, Virginia Tech, 2125 Derring Hall, Blacksburg, Virginia 24061 USA
[3]Biology Department, SUNY New Paltz, 1 Hawk Drive, New Paltz, New York 12561 USA
[4]School of Biological Sciences, Illinois State University, Campus Box 4120, Normal, Illinois 61790 USA

Citation: Klug, J. L., C. C. Carey, D. C. Richardson, and R. D. Gougis. 2017. Analysis of high-frequency and long-term data in undergraduate ecology classes improves quantitative literacy. Ecosphere 8(3):e01733. 10.1002/ecs2.1733

Abstract. Ecologists are increasingly analyzing long-term and high-frequency sensor datasets as part of their research. As ecology becomes a more data-rich scientific discipline, the next generation of ecologists needs to develop the quantitative literacy required to effectively analyze, visualize, and interpret large datasets. We developed and assessed three modules to teach undergraduate freshwater ecology students both scientific concepts and quantitative skills needed to work with large datasets. These modules covered key ecological topics of phenology, physical mixing, and the balance between primary production and respiration, using lakes as model systems with high-frequency or long-term data. Our assessment demonstrated that participating in these modules significantly increased student comfort using spreadsheet software and their self-reported competence in performing a variety of quantitative tasks. Interestingly, students with the lowest pre-module comfort and skills achieved the biggest gains. Furthermore, students reported that participating in the modules helped them better understand the concepts presented and that they appreciated practicing quantitative skills. Our approach demonstrates that working with large datasets in ecology classrooms helps undergraduate students develop the skills and knowledge needed to help solve complex ecological problems and be more prepared for a data-intensive future.

Key words: freshwater ecology; Global Lake Ecological Observatory Network; ice phenology; lake metabolism; lake stratification; Project Environmental Data-Driven Inquiry and Exploration; quantitative skills; teaching modules.

Received 1 August 2016; revised 4 November 2016; accepted 3 January 2017. Corresponding Editor: Diane Ebert-May.

† **E-mail:** jklug@fairfield.edu

Introduction

Ecological research is becoming more data-intensive, as many ecologists now commonly acquire, manage, and analyze large volumes of quantitative and qualitative information (Michener and Jones 2012, Hampton et al. 2013, Schimel and Keller 2015). These data span both long periods of time (>1 decade) and high measurement frequencies. Although ecologists have historically used long-term data in their research (e.g., Magnuson et al. 2000), the increasing availability and duration of these long-term datasets have enabled new analyses on how systems are changing (Karasti and Baker 2008, Zimmerman 2008). In concert, ecologists are analyzing large datasets containing high-frequency data collected by automated sensors. Innovations in sensor technology and data analysis and increased data sharing are rapidly increasing the availability of high-frequency datasets for ecologists (Michener et al. 2011, Reichman et al. 2011, Weathers et al. 2013).

To harness the changing nature of ecology (Michener and Jones 2012), the next generation of ecologists needs to develop the quantitative literacy required to effectively use large datasets from long-term studies and high-frequency sensors. Quantitative literacy, in the context of our study, encompasses the skills needed to access, manipulate, and analyze large datasets, as well as the ability to use those data to ask and answer ecological questions. Challenges to developing quantitative literacy include learning how to use data analysis software (Stevenson et al. 2014), understanding the inherent variability in real data (Gougis et al. 2016), and determining the types of questions that can be addressed with large datasets (Langen et al. 2014). As a result of these challenges, a recent study of graduate students found that many lacked the skills necessary to work with large datasets (Hernandez et al. 2012), suggesting that having students begin to work with large datasets as undergraduates may be useful for developing quantitative literacy. Previous studies have suggested that activities that allow students to work with authentic data in undergraduate classrooms may be effective tools to improve quantitative literacy and teach ecological concepts (Ellwein et al. 2014, Langen et al. 2014). Emphasis on these types of activities has led to increasing availability of teaching materials that engage students in analysis of authentic ecological data (e.g., The EcoEd Digital Library [Klemow et al. 2009] and Teaching Issues and Experiments in Ecology [TIEE; D'Avanzo et al. 2006]).

One approach to introducing large datasets into undergraduate classrooms is to use teaching modules that allow students to analyze "real" data—data collected in the field by sensors and people for research and monitoring objectives, not created just for teaching exercises—to answer ecological questions (sensu Ellwein et al. 2014, Langen et al. 2014). Developing such modules is the focus of Project EDDIE (Environmental Data-Driven Inquiry and Exploration; http://projecteddie.org). EDDIE modules are designed to teach the quantitative skills that students need to ask and answer questions using large datasets and alleviate some of the barriers that currently limit the use of large datasets in undergraduate classrooms (Carey et al. 2015*a*, Bader et al. 2016). For example, individual modules include materials for instructors (annotated instructor's manuals and lecture slides) as well as student readings and handouts, reducing the preparation time for instructors. Strasser and Hampton (2012) found that lack of time was frequently cited by instructors as a barrier to teaching data-related skills. Similarly, modules include either data files or instructions for accessing data online (e.g., from the United States Geological Survey (USGS) stream nutrient data portal), alleviating the challenge of finding appropriate datasets for classroom activities (Langen et al. 2014, Stevenson et al. 2014).

Faculty may avoid teaching analytical skills needed to work with data because of a perceived trade-off between developing content knowledge and skills (e.g., Coil et al. 2010). However, EDDIE modules aim to teach both content knowledge and quantitative literacy. Initial feedback from instructors and students shows that completion of EDDIE modules increases student's appreciation and understanding of the importance of large datasets and improves their ability to work with data (Carey et al. 2015*a*, Bader et al. 2016, Carey and Gougis 2017). All teaching materials for the modules are freely available at http://projecteddie.org.

We developed three EDDIE modules specifically for use in freshwater ecology courses to better prepare undergraduate students to participate in the use of long-term and high-frequency data. Many freshwater ecologists have embraced the use of automated sensors (Weathers et al. 2013, Meinson et al. 2016) and are using high-frequency data to address important ecological research questions, including drivers of whole lake metabolism (Solomon et al. 2013), lake responses to extreme events (Jennings et al. 2012, Klug et al. 2012), feedbacks between nutrient loading and hypoxia (Gerling et al. 2016), and the effects of climate change on lake thermal stratification and productivity (O'Reilly et al. 2003). The three modules were designed to address these emerging themes in freshwater ecology and help students learn quantitative skills (Box 1, Fig. 1).

This study addresses whether the completion of the suite of modules improves students' quantitative literacy and understanding of ecological concepts. Specifically, we asked three research questions: Does participation in the modules increase self-reported student comfort and ability with spreadsheet software? Were there differential gains in self-reported student comfort and ability with spreadsheet software among students with different experience levels? Does participation in the modules improve student's perceptions of their understanding of the ecological concepts presented?

Box 1

Description of the three Project EDDIE modules used for this project.

	Lake ice phenology (Carey et al. 2015*b*)
Website	http://cemast.illinoisstate.edu/data-for-students/modules/ice-phenology.shtml
Description	Students explore long-term records of ice melting dates (ice-off) from lakes around the world and use linear regression to make predictions about ice-off dates in the future
Ecological concept learning objectives	• Understand ecological relevance of timing of ice-off and how global climate change affects long-term trends in ice-off dates • Calculate heat budgets and understand the interactions between ice-off date and heat storage
Quantitative skill development learning objectives	• Develop skills for spreadsheet navigation, data manipulation, graphing, and linear regression • Understand the importance of variability while using linear models to predict future scenarios
	Lake mixing (Carey et al. 2015*c*)
Website	http://cemast.illinoisstate.edu/data-for-students/modules/lake-mixing.shtml
Description	Students explore spatial and temporal patterns of lake mixing using high-frequency temperature data from lakes around the world and use a lake model to explore the ecological implications of climate change on thermal stratification
Ecological concept learning objectives	• Understand the drivers of lake mixing and thermal stratification by comparing and contrasting lake mixing regimes across different lakes • Predict how climate change will affect lake thermal stratification and the implications for distribution of organisms
Quantitative skill development learning objectives	• Develop skills for spreadsheet navigation, data manipulation, and graphing • Visually identify drivers of variation in time series data
	Lake metabolism (Richardson et al. 2015)
Website	http://cemast.illinoisstate.edu/data-for-students/modules/lake-metabolism.shtml
Description	Students explore high-frequency water quality datasets from lakes around the world to calculate estimates of metabolism (gross primary production (GPP) and respiration (R)) in lakes with different trophic status
Ecological concept learning objectives	• Understand the ecological consequences of eutrophication in aquatic ecosystems and the difference between structural vs. functional ecosystem metrics • Calculate and compare rates of GPP and R from lakes with different trophic status
Quantitative skill development learning objectives	• Develop basic skills used for data manipulation and numerical calculations including formula entry and spreadsheet navigation • Identify drivers of temporal and spatial variability

Methods

Study overview

Project EDDIE's three freshwater ecology modules: Lake Ice Phenology, Lake Mixing, and Lake Metabolism (Box 1), were taught in that order in upper-level ecology courses within one month at three universities, including a public research university (very high research activity), a public comprehensive master's college and university (larger program), and a private comprehensive master's college and university (larger program) in fall 2015 and spring 2016. Course sizes were small (13, 20, and 32 students) and had a high teacher-to-student ratio, with either just one instructor or an instructor and a graduate student teaching assistant. Each module had three scaffolding data analysis activities (A, B, and C), which were taught in each course, with some of the activities assigned as homework.

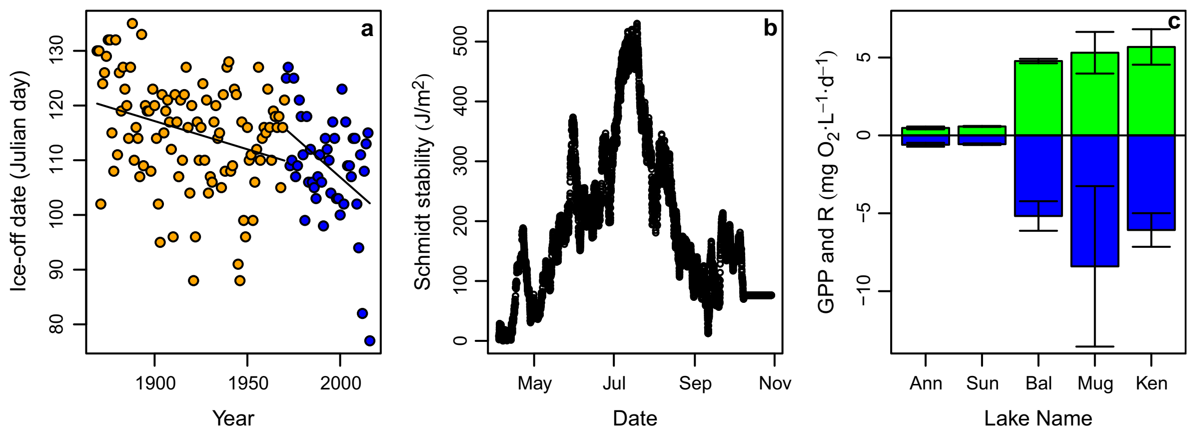

Fig. 1. Example student output from (a) the Lake Ice Phenology module showing earlier ice-off dates in Lake Sunapee, United States, with faster rates after 1970 (1.0 Julian days/decade) compared to prior to 1970 (0.3 Julian days/decade), (b) the Lake Mixing module showing variability in the strength of thermal stratification (Schmidt stability) in Lake Lillinonah, United States, and (c) the Lake Metabolism module showing average daily gross primary production and respiration in Lake Annie, United States (Ann); Lake Sunapee, United States (Sun); Lake Balaton, Hungary (Bal); Muggelsee, Germany (Mug); and Kentucky Lake, United States (Ken), with error bars as standard errors across nights for R (n = 4) and days for GPP (n = 3).

We performed a quantitative comparison before and after instruction of the three modules, augmented by analysis of qualitative data elicited through open-ended questions on a post-instruction questionnaire. Data collection procedures were approved by Institutional Review Boards at all instructors' institutions. Participants were recruited from the first three authors' upper-level ecology courses by showing a recruitment video developed by RDG in class. Within a week prior to the first module's instruction, instructors emailed their students a link to the pre-instruction questionnaire and/or posted it on their course management website (e.g., Blackboard). Following the instruction of the third module and completion of assignments, the students were given the link to the post-instruction questionnaire. Each questionnaire took approximately 20 min to complete. In all three courses, the assessments were optional, but students were offered a small number of extra credit points (<2% of final course grade) for completing the pre- and post-instruction questionnaires. To assign extra credit, RDG emailed the instructors a list of individuals who completed each questionnaire, but not the questionnaire results. The entire process (module instruction and pre-/post-assessments) took ~1.5 months in each course.

Modules description

The three modules of interest in this study were developed, tested, and revised by teams of freshwater ecologists over multiple years (Carey et al. 2015*a*) and are publicly available (Carey et al. 2015*b*, *c*, Richardson et al. 2015). Each module consists of time series data from long-term manual monitoring or high-frequency automated sensors accessed from published studies or widely available datasets (e.g., Benson and Magnuson 2000, Solomon et al. 2013). The sensor data are primarily from Global Lake Ecological Observatory Network (http://gleon.org) lakes from around the world. The modules have a similar structure, with readings, introductory material, and activities that span the levels of Bloom's taxonomy (Crowe et al. 2008). Each module also has learning objectives related to both ecological concepts and quantitative skill development (Box 1), with activities that require the students to describe and understand concepts and patterns in order to calculate, compare, contrast, and synthesize concepts across activities.

In each module, the students were responsible for exploring and visualizing datasets, summarizing data, calculating statistics, identifying trends, and interpreting calculations to develop their quantitative literacy, with a focus on improving their spreadsheet software manipulation skills and learning ecological concepts. The Lake Ice Phenology module examines long-term records in lake ice-off dates (i.e., the day of year when winter ice breaks up and melts), with a focus on using spreadsheet software to graph data, learn linear regression, discuss variability, and discuss the ecological implications of climate change. The students create graphs of several lakes, visually assess breakpoints, and compare rates of change using linear regression. For example, students examine the trends in the 147-yr record of Lake Sunapee (New Hampshire, USA) ice-off dates and may decide on 1970 for a breakpoint. Then, they would compare the pre- and post-1970 trends in ice-off dates (Fig. 1a). Finally, changes in ice-off dates are compared from a global selection of lakes on latitudinal, size, and water quality gradients. The Lake Mixing module examines thermal stratification across a series of lakes that differ in geomorphology and location. The students visually explore heat maps and generate time series plots of lake thermal stratification strength, as indicated by Schmidt stability (Read et al. 2011), and potential drivers such as wind speed and air temperature. For example, students could graph the Schmidt stability of Lake Lillinonah (Connecticut, USA) and look for drivers of the variability as a result of seasonal peaks, lake turnover, and storm events throughout the summer (Fig. 1b). The Lake Metabolism module allows students to calculate gross primary production and respiration using time series of dissolved oxygen data from high-frequency sensors. Student teams calculate metabolism rates from different lakes around the world that range in trophic status from oligotrophic to eutrophic and then compare rates among lakes, specifically examining how metabolism is mediated by nutrient concentrations (Fig. 1c).

Student assessment

The pre-/post-instruction questionnaire consisted of two sections. The first section gauged the students' comfort and ability in spreadsheet software, specifically Microsoft Excel (hereafter, Excel), the software used for these modules. The students were asked to rank their comfort level with Excel using the following scale: 1 = I don't know how to do anything in Excel; 2 = I only know how to do a few things in Excel; 3 = I know how to do several things in Excel well; 4 = I feel very competent in Excel but would not feel comfortable teaching others how to use Excel; 5 = I feel very competent in Excel and would feel comfortable teaching others how to use Excel. The participants were also asked to rank their ability using 10 different functions in Excel: calculate an average, median, standard deviation, variation; perform a correlation; find a maximum value in a data array; draw a trend line; analyze an equation for a trend line; create a bar graph; and create a line graph. They ranked their ability on each of these items using the following scale: 1 = I feel incompetent doing this task; I would not know where to start; 2 = I could figure out how to do this task; 3 = I feel somewhat competent doing this task; 4 = I feel very competent doing this task independently; or 5 = I feel so competent doing this task that I could teach others.

The second section of the pre-/post-instruction questionnaire included three identical open-ended questions (one for each module). The students were asked, "What was your favorite aspect or component of the MODULE activity," where MODULE was the title of each of the three different modules: Lake Ice Phenology, Lake Mixing, and Lake Metabolism. Students completed the open-ended responses in a blank text box. Across all three modules, we identified the five most common themes in the open-ended responses as: "better understanding of scientific concepts," "practicing quantitative skills (statistics, practice with spreadsheet software, and graphical analysis)," "working with real data," "comparing results across different lakes," and "working collaboratively with peers." Three coauthors (CCC, JLK, and DCR) and an independent education researcher categorized each response into one or more of the five themes. We compared the four researchers' results for every response; if fewer than 75% of the individuals agreed on the coding for a student response, we then discussed how the response should be categorized and came to a consensus on the coding for that response.

Data analyses

We used Wilcoxon signed-rank tests to compare the students' pre- vs. post-instruction scores on their self-reported spreadsheet comfort and ability on the 10 Excel tasks. In addition, we

Table 1. Statistical comparison of pre- and post-instruction responses for all students, aggregated across the three courses.

Item	Wilcoxon signed-rank test statistic	P-value	n	Pre-module mean ± 1 SE	Post-module mean ± 1 SE	Effect size
Excel comfort	89.50	**0.0003**	40	2.214 ± 0.165	2.780 ± 0.173	10.01
Calculate an average	51.50	**0.03**	43	4.186 ± 0.164	4.465 ± 0.146	5.55
Calculate a median	30.00	0.26	43	3.930 ± 0.180	4.116 ± 0.150	3.23
Calculate a standard deviation	55.00	**0.018**	43	3.651 ± 0.208	3.953 ± 0.194	5.93
Calculate variation	61.00	**0.03**	43	3.186 ± 0.200	3.512 ± 0.195	6.58
Perform a correlation	130.50	**<0.0001**	43	3.116 ± 0.195	3.860 ± 0.175	14.07
Find a maximum value in a data array	82.50	**0.0007**	42	3.744 ± 0.189	4.262 ± 0.118	9.00
Draw a trend line	88.50	**<0.0001**	43	4.209 ± 0.151	4.791 ± 0.078	9.54
Analyze an equation for a trend line	157.50	**<0.0001**	43	3.884 ± 0.165	4.651 ± 0.093	16.98
Create a bar graph	35.00	**0.004**	43	4.465 ± 0.130	4.837 ± 0.066	3.77
Create a line graph	35.00	**0.004**	42	4.465 ± 0.130	4.833 ± 0.067	3.82
Mean ability across Excel tasks	265.00	**<0.0001**	43	3.884 ± 0.140	4.322 ± 0.102	28.58

Notes: For the first item, students were asked to rank their comfort level using Microsoft Excel. For all of the other items, students were asked to rank how competent they felt performing each task in Excel. All ranks were conducted on a scale from 1 (lowest) to 5 (highest). Significant changes ($P < 0.05$) between pre- and post-instruction responses are highlighted in boldface. Effect sizes were calculated as $Z/\sqrt{n}$, following Rosenthal (1994).

averaged scores from the 10 items asking about ability on individual Excel tasks to create a mean Excel ability score on a 1–5 scale (Cronbach's alpha = 0.94). The data were pooled across the three courses because of small sample sizes of respondents in each course (n = 8, 14, and 21). Finally, we calculated gains in the Excel comfort and mean Excel ability scores by taking the difference between the pre- and post-instruction scores. Here, a positive number indicates a gain in comfort or self-assessed ability, while a negative number indicates a decrease. We compared those increases or decreases to the initial pre-instruction scores for the respective metrics and used linear regression to compare the students' pre-instruction scores to gains. All statistics were completed in JMP Pro (v.11.0.0, SAS Institute Inc., Cary, North Carolina, USA) statistical software.

Results

After completion of the modules, students reported increased overall comfort and ability using Excel (Table 1). Post-module scores on comfort level, mean ability, and ability on nine out of 10 individual Excel tasks were significantly higher than pre-module scores (Table 1). Furthermore, students that scored lower on mean Excel ability prior to module instruction exhibited the greatest gains in Excel ability (Fig. 2; gain in mean Excel ability = 2.24 (±0.31) − 0.46 (±0.08) × pre-mean Excel ability; $n = 43$, $R^2 = 0.46$, $P < 0.0001$). We observed a similar relationship for student comfort with Excel (data not shown, gain in Excel comfort = 1.71 (±0.34) − 0.50 (±0.14) × pre-mean Excel comfort; $n = 40$, $R^2 = 0.26$, $P = 0.0009$).

In conjunction with overall increase in comfort and ability, the distribution of comfort and ability scores shifted upward with fewer students at the lower tail. Almost half of the students improved in comfort (43%) and most improved in mean

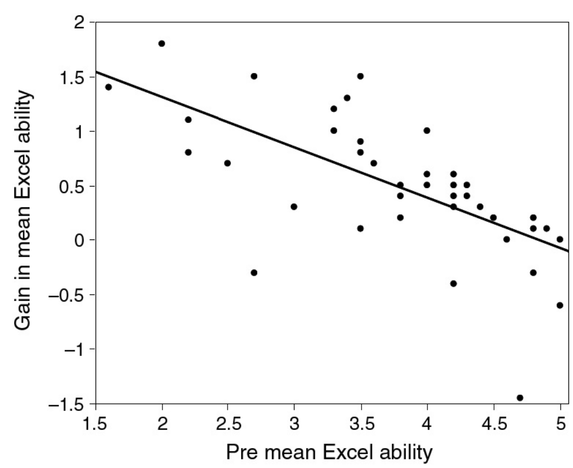

Fig. 2. Relationship between students' pre-instruction ability on tasks in Excel (x-axis) and their improvement in Excel ability (y-axis) [gain in mean ability = 2.24 (±0.31) − 0.46 (±0.08) × pre-mean ability $n = 43$, $R^2 = 0.46$, $P < 0.0001$].

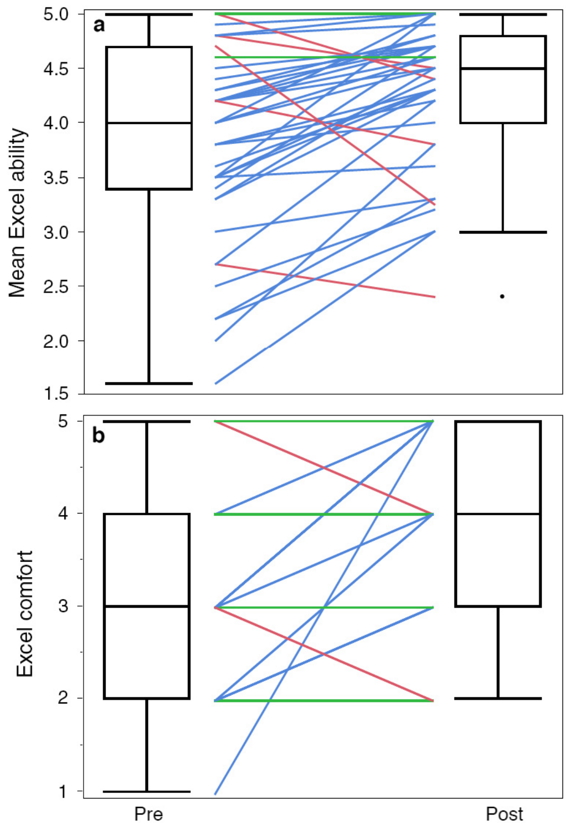

Fig. 3. Distribution of student scores on (a) Excel ability (n = 43) and (b) comfort (n = 40) using self-reported Excel ability before (pre) and after (post) completing the three EDDIE modules in freshwater ecology courses. Boxes show the median score and 25th and 75th percentiles, and whiskers show the first quartile minus 1.5× the interquartile range and the third quartile plus 1.5× the interquartile range. One outlier is denoted as a point in panel a. Colored lines connect individual students' pre- and post-module scores. Blue lines indicate student improvement, green lines represent no change, and red lines indicate student declines in ability and comfort using Excel between assessments. Lines can be overlapping in each panel.

ability (74%), some dramatically so (Fig. 3). Although students who started in the lowest quartile prior to module instruction showed greater gains than the students who started in the highest quartile (Fig. 2), their post-module scores were still lower when compared to students who started in the highest quartile (post-Excel comfort; $t_{13} = -4.31$, $P = 0.0008$; post-Excel ability; $t_{18} = -4.09$, $P = 0.0006$).

When asked to describe their favorite aspects or components of individual modules, the most common response was related to "better understanding of scientific concepts" (37–44% of students on each individual module), followed by "practicing quantitative skills" (23–28%). Students also identified "working with real data," "comparing results from different lakes," and "working collaboratively with their peers" as favorite aspects. There were no large differences in favorite aspects among modules, with the exception of "working with real data." Twenty-one percent of students identified "working with real data" as a favorite aspect of the ice-off module, whereas only 7% reported that for the Lake Mixing module and 5% for the Lake Metabolism module.

Aggregating across modules, seventy percent of all students wrote that "better understanding of the scientific concepts" was their favorite aspect of at least one module, suggesting that most students felt the modules improved their understanding of the ecological concepts presented. For example, one student wrote that the Lake Metabolism module improved their understanding of how metabolism was affected by trophic state, noting that: "...looking at the data we could tell which lakes were oligotrophic vs. eutrophic and understand how consumption could be higher than production of GPP." Another wrote of the ice-off module, "my favorite part of this activity was learning of the effects of climate change and how it affects lakes worldwide."

Similarly, the majority (56%) of students reported that their favorite aspect of at least one module was practicing quantitative skills. Examples of student responses in this category included, "I liked that this activity helped develop my skills in Excel and interpretation of graphs similar to the one we created," "I liked being able to come up with predictions of ice-off dates in the future," and "being able to learn to graph multiple datasets and get two separate regression lines was really important and cool."

About a quarter of students identified working with real data (26%) and comparing results from different lakes (29%) as a favorite aspect of at least one module. One student wrote of the ice-off module, "I liked having real data to work with more hands on. It was easier to see what was going on in the lake system and I liked comparing the different lakes and trying to determine the factors that may have caused the variation," whereas another student wrote of the Lake Mixing module, "I appreciated working with actual data to better understand how dimictic and polymictic lakes can vary in their stratification patterns."

Discussion

In this study, teaching three ecological instructional modules improved students' comfort working with large datasets and the students' ability to carry out quantitative tasks. Furthermore, the students that began with the lowest self-reported ability experienced the greatest gains after participating in the three modules. Extensive and repeated experience working with real data in a spreadsheet program across three different contexts has the potential to effectively help students who begin instruction with the fewest quantitative skills. Ultimately, the modules work toward leveling student skill and comfort levels performing quantitative tasks. Finally, the qualitative responses indicate that students perceive increased conceptual understanding of ecological concepts and increased quantitative skills as the primary learning gains during these three ecological teaching modules, indicating our learning objectives are being addressed through these modules.

Many students reported large gains in comfort with spreadsheet software and in performing data exploration and statistical tasks (Fig. 3). These are skills that are important to students in ecological classrooms and with interests in ecological or environmental science careers, as well as many other career paths. These gains were made by repeatedly working with spreadsheets and manipulating data by specifically examining trends, exploring variation, doing calculations, and interpreting the meaning of statistics. However, for each module, the ecological concept learning objectives were different (Box 1); the sampling frequency, structure, and content of the datasets were different (Fig. 1); and the data manipulation tasks varied for each module. Participating in all three modules within one semester gave the students opportunities to practice using the spreadsheet software to explore real data in different contexts.

Our results show gains in self-reported ability and comfort with using spreadsheet software after completion of three modules but it is likely that some gains would be realized after completion of fewer modules. Preliminary results from a study using other EDDIE modules suggest that student self-reported Excel comfort and Excel ability scores increase after completion of one EDDIE module (Bader et al. 2016). Anecdotal observations from the authors of this study also suggest improvement after fewer than three modules. For example, we found that students asked fewer Excel-related questions in the classroom when completing the last module compared to the first.

Confidence and comfort are critical components of quantitative literacy (Wilkins 2000, Parsons et al. 2009). Increasing student confidence (and lowering anxiety) in quantitative tools and concepts can facilitate student learning when dealing with more rigorous concepts and increase the student's ability to apply quantitative skills to software and situations (Bos and Schneider 2009). Collectively, increasing confidence as well as skills and knowledge results in improvements to academic performance and post-undergraduate success in quantitative literacy (Tariq and Durrani 2012). Students that progress to use large datasets to ask research questions will likely need to move beyond basic spreadsheet software and use scripted programming languages such as R, Python, or Matlab. Carey and Gougis (2017) showed that undergraduates can gain confidence and ability with a more advanced software platform (R) by completing an EDDIE module on lake modeling, suggesting that the gains we saw in our study are generalizable across software platforms.

Qualitative feedback from students highlights their appreciation of the module activities. Students liked that the modules helped them better understand the scientific concepts presented in the modules and practice their quantitative skills. Further, students identified working with real data and working collaboratively as favorite aspects, suggesting that students are gaining skills and approaches that are valuable for modern ecological science. For example, ecological research projects are increasingly collaborative across institutions and disciplines (Borrett et al. 2014) and working in teams requires practice and training (Cheruvelil et al. 2014, Read et al. 2016). Additionally, many ecological questions require comparisons across ecosystems and messy, highly variable data (Hampton et al. 2013). The students recognize that these are valuable components of the modules: In all three of the classes studied here, the instructors facilitated a dialogue with the students about how these topics are handled in real-world ecological research and science in general.

In order for professionals to work effectively with complex data, they need to develop quantitative data skills by extensively practicing those skills during their undergraduate science curricula (AAAS 2011). The students in our study found value in working with real data. However, the majority of the time, students in undergraduate and secondary education classrooms are trained with problems where there is little variability, relationships are clearly defined, and datasets are relatively simple (Hoskinson et al. 2013). These routine problems or exercises often have a single, correct answer with a clear method for getting to that answer. In our modules, the students worked with real, variable data often without a clear path to approach analyses. This was a problem for many students; for example, one student mentioned, "I didn't like missing data." These negative sentiments provide opportunity to talk about the challenges of data collection (e.g., the lightning strike that resulted in sensor malfunction and missing data). Many students are unaware of the extent of data available. One student reported directly to their instructor that "the modules made me want to learn more about how these large datasets make a difference in the real world. Who is using this data and how they are changing something in environmental science because of the results?" Thus, there appears to be great value in allowing students to struggle with complex problems and real data.

The activities appeared to challenge all the students but perhaps in different ways. In the classrooms for this project and many others, there is often a bimodal distribution of quantitative skills with the lower mode consisting of poorly prepared students and an upper mode of students with more developed quantitative skills (Maltese et al. 2015). At the lower end of the distribution, the students with the most to gain exhibited the greatest gains in our study. Similarly, Beck and Blumer (2012) showed that the largest gains in scientific reasoning skills following inquiry-based activities in ecology laboratory courses were observed in students who started in the lowest quartile. Students often learn quantitative skills independently of their biology and ecology classes and struggle with transferring skills to their ecology classes (Hester et al. 2014). However, integration of quantitative skill development into the curriculum (as in the modules in this study) can facilitate both learning and application of those quantitative skills to ecologically relevant problems (Hester et al. 2014). The students in the upper mode were able to reinforce their quantitative skills and perceive gains in their conceptual understanding.

A few students in our study decreased in their comfort and self-reported ability using spreadsheet software from pre- to post-instruction (Fig. 3). Most of these students started out at the upper third of scores on the pre-instruction questionnaire (Fig. 2). These students might have been initially overconfident before attempting the modules and realized that they knew less than they thought they did during the post-instruction questionnaire. Previous work with undergraduate psychology students shows that lower-skilled individuals may overestimate their abilities because they lack the skills needed for accurate self-assessment (Kruger and Dunning 1999). Interestingly, Kruger and Dunning (1999) found that self-assessment ability improved as the ability to perform tasks improved; thus, it is possible that some of the self-reported declines in our study were related to improvement in self-assessment, rather than a decrease in actual skill level.

The Project EDDIE modules are fundamentally flexible in their design. They require no special equipment; usually one computer per pair of students is sufficient. They can be taught in a computer laboratory or with student or departmental laptops in varied classroom arrangements. Subsets of the activities within the module can be taught depending on the time available, class setting (lecture vs. laboratory), or level of course. The modules in this study were all taught in freshwater ecology classes, but one co-author (JK) has successfully adopted components in a general education environmental science course. The three freshwater ecology modules here can be taught alone or in sequence. They can also be paired with more traditional laboratory exercises. For example, some freshwater ecology courses cover concepts of thermal stratification in lakes with an accompanying exercise where students heat water in a model lake (i.e., an aquarium) with a heat lamp and watch as stratification occurs over time on a small scale (Wetzel and Likens 2000). For classes on the semester system, sampling stratification in lakes in fall and spring semesters is often impractical so the lake model may be students' only chance to collect their own temperature data. Pairing the small-scale model data with the high-frequency time series from the Lake Mixing module can enhance student understanding of the variability in lake mixing patterns and allow students to put laboratory experiments and local data into a global ecological context. Furthermore, instructors can incorporate additional data into each module (e.g., from the National Snow and Ice Database), providing opportunities for place-based learning (Gosselin et al. 2015), particularly valuable if course or institution resources are limited or field trips are infeasible.

Inquiry-based learning has been shown to help students achieve the learning outcomes expected of the next generation of scientists, including the ability to test hypotheses (Minner et al. 2010), acquire science process skills (Gormally et al. 2009), and use evidence-based reasoning (Beck and Blumer 2012). However, there are often challenges for both the students and instructors when changing a classroom to a student-centered, active learning environment involving data exploration (Felder and Brent 1996, Crawford 2007, Spronken-Smith et al. 2011, Gormally et al. 2016). Students are accustomed to a didactic classroom where the instructor attempts to transfer knowledge to the students with clear processes and answers (i.e., the instructor tells the students everything they need to know). Students may resist when they are forced to take a more active role in their learning and given assignments that have multiple correct answers and processes for arriving at an answer (Gormally et al. 2016). The EDDIE modules are structured to provide an introduction both to the content and to the active learning process to overcome these challenges. It is particularly useful to establish expectations early in the course regarding the frustration that students will likely encounter as they grapple with real data. Additionally, there is a fine balance to be struck between giving students step-by-step instructions and providing less direction. The latter grants opportunity for creative problem-solving but also has the potential for a student to reach an impasse or feel defeated by the activity. These decisions are left to the discretion of their instructor, who is best equipped to determine their particular students' prior knowledge and experience with spreadsheets, statistical concepts, and active learning formats. One suggestion to alleviate these potential roadblocks is to ask students who have completed the activity to act as learning assistants to others who are still finishing, an approach that has shown benefits for both the learning assistants and the students receiving help (Talbot et al. 2015).

Challenges may also be encountered by the instructor teaching these modules. Instructors need to recognize the extra time required for data exploration in class when designing their curricula. The modules require instructor preparation, effort, willingness to circulate and interact with many students as they complete activities, and flexibility in course timing. Instructors can ease the shift to a student-centered classroom by preparing in advance, completing the activities themselves prior to the class, and managing student expectations and motivation in the syllabus, prior to, and during each class (Roehl et al. 2013). Finally, these modules are built for use with Microsoft Excel, but there are many versions of spreadsheet software that students use including earlier versions of MS Excel, Apple Numbers, or free software like Google Sheets or Apache OpenOffice Calc. One solution may be to hold the class in a computer laboratory where the instructor is familiar with the existing software, if such computer laboratories are available. The modules can be completed using other software, but instructors should be prepared to help students troubleshoot.

Our results are based on student assessment of their comfort and ability using spreadsheet software. Self-reported student data are common in educational research (e.g., Beck and Blumer 2012, McCright 2012) and are valuable in gauging effectiveness of the modules in meeting their learning objectives. However, self-reported data do have limitations because self-reported ability may not represent actual ability, as individuals can overestimate their abilities (Gross and Latham 2012), as noted above. Future research using other forms of assessment, such as use of quantitative instruments or spreadsheet-driven tasks that directly measure ability, would be valuable in assessing the degree to which completion of EDDIE modules increases quantitative literacy. For example, researchers could ask students to create a line graph and capture their cursor movements and keyboard actions through screen-recording software (e.g., CamStudio) to evaluate not just the finished product but also the process by which they created the product. In addition, future research could address whether post-module gains are sustained by assessing student comfort and ability at the end of the semester or in future coursework; however, it may be difficult to control for the effects of instruction given in other classes between assessments. Although it is plausible that the gains we observed are due to module instruction, we cannot make strong claims about causation due to the lack of a control group. However, consistent gains in students who use EDDIE modules across a variety of classrooms and course experience levels (Carey et al. 2015*a*, Bader et al. 2016, Carey and Gougis 2017) suggest that our results are reliable and repeatable.

Conclusions

Developing scientific literacy is a critical component of an undergraduate education, regardless of career path (Feinstein et al. 2013). We suggest that students can develop substantial gains in quantitative literacy by manipulating and analyzing large datasets in undergraduate classes. We advocate the use of flexible teaching modules, which allow instructors to tailor content and context for their students, as important tools to help undergraduate students develop the skills and knowledge needed to navigate our increasingly data-rich society.

Acknowledgments

We thank the students who participated in this study, Randy Fuller for contributing to the initial development of the Lake Mixing module, Janet Stromberg for help with qualitative data analysis, and Jon Doubek and Kate Hamre for teaching assistance. We also thank our Project EDDIE colleagues, Nick Bader, Devin Castendyk, Randy Fuller, Cathy Gibson, Luke Nave, Catherine O'Reilly, Dax Soule, Tom Meixner, and Kathleen Weathers, for helpful discussions. Project EDDIE is funded by NSF TUES 1245707, with administrative support from CeMaST (Center for Mathematics, Science and Technology at Illinois State University).

Literature Cited

AAAS [American Association for the Advancement of Science]. 2011. Vision and change in undergraduate biology education: a call to action. AAAS, Washington, D.C., USA.

Bader, N. E., D. Soule, D. Castendyk, T. Meixner, C. O'Reilly, and R. D. Gougis. 2016. Students, meet data. Eos 97. https://doi.org/10.1029/2016EO047175

Beck, C. W., and J. S. Blumer. 2012. Inquiry-based ecology laboratory courses improve student confidence and scientific reasoning skills. Ecosphere 3:112.

Benson, B., and J. Magnuson. 2000. Global lake and river ice phenology database, Version 1. NSIDC: National Snow and Ice Data Center, Boulder, Colorado, USA. https://doi.org/10.7265/N5W66HP8

Borrett, S. R., J. Moody, and A. Edelmann. 2014. The rise of network ecology: maps of the topic diversity and scientific collaboration. Ecological Modelling 293:111–127.

Bos, A. L., and M. C. Schneider. 2009. Stepping around the brick wall: overcoming student obstacles in methods courses. PS: Political Science and Politics 42:375–383.

Carey, C. C., and R. D. Gougis. 2017. Simulation modeling of lakes in undergraduate and graduate classrooms increases comprehension of climate change concepts and experience with computational tools. Journal of Science Education and Technology 26:1–11.

Carey, C. C., R. D. Gougis, J. L. Klug, C. M. O'Reilly, and D. C. Richardson. 2015*a*. A model for using environmental data-driven inquiry and exploration to teach limnology to undergraduates. Limnology and Oceanography Bulletin 24:32–35.

Carey, C. C., J. L. Klug, and R. L. Fuller. 2015*c*. Project EDDIE: dynamics of Lake Mixing. Project EDDIE Module 3, Version 1. http://cemast.illinoisstate.edu/data-for-students/modules/lake-mixing.shtml

Carey, C. C., J. L. Klug, and D. C. Richardson. 2015*b*. Project EDDIE: Lake Ice Phenology. Project EDDIE Module 1, Version 1. http://cemast.illinoisstate.edu/data-for-students/modules/ice-phenology.shtml

Cheruvelil, K. S., P. A. Soranno, K. C. Weathers, P. C. Hanson, S. J. Goring, C. T. Filstrup, and E. K. Read. 2014. Creating and maintaining high-performing collaborative research teams: the importance of diversity and interpersonal skills. Frontiers in Ecology and the Environment 12:31–38.

Coil, D., M. P. Wenderoth, M. Cunningham, and C. Dirks. 2010. Teaching the process of science: faculty perceptions and an effective methodology. CBE—Life Sciences Education 9:524–535.

Crawford, B. A. 2007. Learning to teach science as inquiry in the rough and tumble of practice. Journal of Research in Science Teaching 4494:613–642.

Crowe, A., C. Dirks, and M. P. Wenderoth. 2008. Biology in bloom: implementing Bloom's Taxonomy to enhance student learning in biology. CBE—Life Sciences Education 7:368–381.

D'Avanzo, C., B. W. Grant, D. Morris, S. Musante, J. Taylor, J. Riney, and D. Udovic. 2006. Design and evaluation of TIEE, a peer-reviewed electronic teaching resource. Frontiers in Ecology and the Environment 4:189–195.

Ellwein, A. L., L. M. Hartley, S. Donovan, and I. Billick. 2014. Using rich context and data exploration to improve engagement with climate data and data literacy: bringing a field station into the college classroom. Journal of Geoscience Education 62: 578–586.

Feinstein, N. W., S. Allen, and E. Jenkins. 2013. Outside the pipeline: reimagining science education for nonscientists. Science 340:314–317.

Felder, R. M., and R. Brent. 1996. Navigating the bumpy road to student-centered instruction. College Teaching 44:43–47.

Gerling, A. B., Z. W. Munger, J. P. Doubek, K. D. Hamre, P. A. Gantzer, J. C. Little, and C. C. Carey. 2016. Whole-catchment manipulations of internal and external loading reveal the sensitivity of a century-old reservoir to hypoxia. Ecosystems 19: 555–571.

Gormally, C., P. Brickman, B. Hallar, and N. Armstrong. 2009. Effects of inquiry-based learning on students' science literacy skills and confidence. International Journal for the Scholarship of Teaching and Learning 3:16. https://doi.org/10.20429/ijsotl.2009.030216

Gormally, C., C. S. Sullivan, and N. Szeinbaum. 2016. Uncovering barriers to teaching assistants (TAs) implementing inquiry teaching: inconsistent facilitation techniques, student resistance, and reluctance to share control over learning with students. Journal of Microbiology and Biology Education 17:215–224.

Gosselin, D., S. Burian, T. Lutz, and J. Maxson. 2015. Integrating geoscience into undergraduate education about environment, society, and sustainability using place-based learning: three examples. Journal of Environmental Studies and Sciences. https://doi.org/10.1007/s13412-015-0238-8

Gougis, R. D., J. F. Stomberg, A. T. O'Hare, C. M. O'Reilly, N. E. Bader, T. Meixner, and C. C. Carey. 2016. Post-secondary science students' explanations of randomness and variation and implications for science learning. International Journal of Science and Mathematics Education. https://doi.org/10.1007/s10763-016-9737-7

Gross, M., and D. Latham. 2012. What's skill got to do with it?: information literacy skills and self-views of ability among first-year college students. Journal of the American Society for Information Science and Technology 63:574–583.

Hampton, S. E., C. A. Strasser, J. J. Tewksbury, W. K. Gram, A. E. Budden, A. L. Batcheller, C. S. Duke, and J. H. Porter. 2013. Big data and the future of ecology. Frontiers in Ecology and the Environment 11:156–162.

Hernandez, R. R., M. S. Mayernik, and M. Murphy-Mariscal. 2012. Advanced technologies and data management practices in environmental science: lessons from academia. BioScience 62:1067–1076.

Hester, S., S. Buxner, L. Elfring, and L. Nagy. 2014. Integrating quantitative thinking into an introductory biology course improves students' mathematical reasoning in biological contexts. CBE—Life Sciences Education 13:54–64.

Hoskinson, A. M., M. D. Caballero, and J. K. Knight. 2013. How can we improve problem solving in undergraduate biology? Applying lessons from 30 years of physics. CBE—Life Sciences Education 12:153–161.

Jennings, E., S. Jones, L. Arvola, P. A. Staehr, E. Gaiser, I. D. Jones, K. C. Weathers, G. A. Weyhenmeyer, C.-Y. Chiu, and E. De Eyto. 2012. Effects of weather-related episodic events in lakes: an analysis based on high-frequency data. Freshwater Biology 57:589–601.

Karasti, H., and K. S. Baker. 2008. Digital data practices and the long term ecological research program growing global. International Journal of Digital Curation 3:42–58.

Klemow, K., J. Riem, T. Mourad, and D. Kirschtel. 2009. Assessing faculty use of digital resources to teach undergraduate ecology: Can ESA's EcoEd digital library save the day? Bulletin of the Ecological Society of America 90:286–295.

Klug, J. L., et al. 2012. Ecosystem effects of a tropical cyclone on a network of lakes in northeastern North America. Environmental Science and Technology 46:11693–11701.

Kruger, J., and D. Dunning. 1999. Unskilled and unaware of it: how difficulties in recognizing one's own incompetence lead to inflated self-assessments. Journal of Personality and Social Psychology 77: 1121–1134.

Langen, T. A., et al. 2014. Using large public datasets in the undergraduate ecology classroom. Frontiers in Ecology and the Environment 12:362–363.

Magnuson, J. J., et al. 2000. Historical trends in lake and river ice cover in the northern hemisphere. Science 289:1743–1746.

Maltese, A. V., D. Svetina, and J. A. Harsh. 2015. Data visualization literacy: investigating data interpretation along the novice-expert continuum. Journal of College Science Teaching 45:84–90.

McCright, A. 2012. Enhancing students' scientific and quantitative literacies through an inquiry-based learning project on climate change. Journal of the Scholarship of Teaching and Learning 12:86–102.

Meinson, P., A. Idrizaj, P. Nõges, T. Nõges, and A. Laas. 2016. Continuous and high-frequency measurements in limnology: history, applications, and future challenges. Environmental Reviews 24: 52–62.

Michener, W. K., and M. B. Jones. 2012. Ecoinformatics: supporting ecology as a data-intensive science. Trends in Ecology and Evolution 27:85–93.

Michener, W. K., D. Vieglais, T. Vision, J. Junze, P. Cruse, and G. Janée. 2011. DataONE: Data Observation Network for Earth—Preserving data and enabling innovation in the biological and environmental sciences. D-Lib Magazine 17:1–9.

Minner, D. D., A. J. Levy, and J. Century. 2010. Inquiry-based science instruction – What is it and does it matter? Results from a research synthesis years 1984–2001. Journal of Research in Science Teaching 47:257–279.

O'Reilly, C. M., S. R. Alin, P.-D. Pilsnier, A. S. Cohen, and B. A. McKee. 2003. Climate change decreases aquatic ecosystem productivity of Lake Tanganyika, Africa. Nature 424:766–768.

Parsons, S., T. Croft, and M. Harrison. 2009. Does students' confidence in their ability in mathematics matter? Teaching Mathematics and Its Applications 28:53–68.

Read, J. S., D. P. Hamilton, I. D. Jones, K. Muraoka, L. A. Winslow, R. Kroiss, C. H. Wu, and E. Gaiser. 2011. Derivation of lake mixing and stratification indices from high-resolution lake buoy data. Environmental Modelling and Software 26:1325–1336.

Read, E. K., M. O'Rourke, G. S. Hong, P. C. Hanson, L. A. Winslow, S. Crowley, C. A. Brewer, and K. C. Weathers. 2016. Building the team for team science. Ecosphere 7:e01291.

Reichman, O. J., M. B. Jones, and M. P. Schildhauer. 2011. Challenges and opportunities of open data in ecology. Science 331:703–705.

Richardson, D. C., J. L. Klug, and C. C. Carey. 2015. Project EDDIE: Lake Metabolism. Project EDDIE Module 2, Version 1. http://cemast.illinoisstate.edu/data-for-students/modules/lake-metabolism.shtml

Roehl, A., S. L. Reddy, and G. J. Shannon. 2013. The flipped classroom: an opportunity to engage millennial students through active learning strategies. Journal of Family and Consumer Sciences 105: 44–49.

Rosenthal, R. 1994. Parametric measures of effect size. Pages 231–244 *in* H. Cooper and L. V. Hedges, editors. The handbook of research synthesis. Russell Sage Foundation, New York, New York, USA.

Schimel, D., and M. Keller. 2015. Big questions, big science: meeting challenges of global ecology. Oecologia 177:925–934.

Solomon, C. T., et al. 2013. Ecosystem respiration: drivers of daily variability and background respiration in lakes around the globe. Limnology and Oceanography 58:849–866.

Spronken-Smith, R., R. Walker, J. Batchelor, B. O'Steen, and T. Angelo. 2011. Enablers and constraints to the use of inquiry-based learning in undergraduate education. Teaching in Higher Education 16: 15–28.

Stevenson, R. D., K. E. Klemow, and L. J. Gross. 2014. Harnessing bits and bytes to transform ecology education. Frontiers in Ecology and the Environment 12:306–307.

Strasser, C. A., and S. E. Hampton. 2012. The fractured lab notebook: undergraduates and ecological data management training in the United States. Ecosphere 3:116.

Talbot, R. M., L. M. Hartley, K. Marzetta, and B. S. Wee. 2015. Transforming undergraduate science education with learning assistants: student satisfaction in large-enrollment courses. Journal of College Science Teaching 44:24–30.

Tariq, V. N., and N. Durrani. 2012. Factors influencing undergraduates' self-evaluation of numerical competence. International Journal of Mathematical Education in Science and Technology 43:337–356.

Weathers, K. C., et al. 2013. The global lake ecological observatory network (GLEON): the evolution of grassroots network science. Limnology and Oceanography Bulletin 22:71–73.

Wetzel, R. G., and G. Likens. 2000. Limnological analyses. Third edition. Springer, New York, New York, USA.

Wilkins, J. L. 2000. Special issue article: preparing for the 21st century: the status of quantitative literacy in the United States. School Science and Mathematics 100:405–418.

Zimmerman, A. S. 2008. New knowledge from old data the role of standards in the sharing and reuse of ecological data. Science, Technology and Human Values 33:631–652.