JOURNAL OF CHEMICAL EDUCATION

Article

pubs.acs.org/jchemeduc

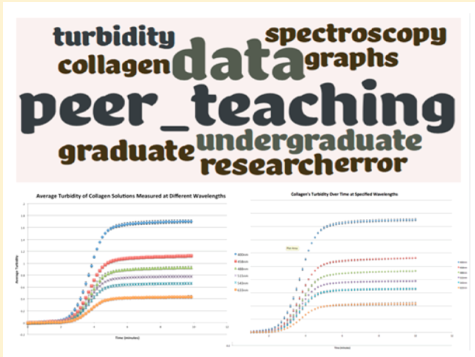
Bridging the Gap between Instructional and Research Laboratories: Teaching Data Analysis Software Skills through the Manipulation of Original Research Data

Sarah J.R. Hansen,* Jieling Zhu, Jessica M. Karch, Cristina M. Sorrento, Joseph C. Ulichny, and Laura J. Kaufman

Department of Chemistry, Columbia University, 3000 Broadway, New York, New York 10027, United States

Ⓢ *Supporting Information*

ABSTRACT: The gap between graduate research and introductory undergraduate teaching laboratories is often wide, but the development of teaching activities rooted within the research environment offers an opportunity for undergraduate students to have first-hand experience with research currently being conducted and for graduate students to develop communication and pedagogical skills. We describe the development, evaluation, and redesign of a Microsoft Excel-based activity where data analysis from contemporary research was adapted into a data analysis training exercise, requiring undergraduate students to engage with the data by performing a series of analytical tasks guided by commentary from the researcher. In addition to learning common data analysis techniques, this exercise provided an opportunity for undergraduate students to practice problem solving and collaboratively work through trouble-shooting. Throughout the activity development, multiple iterations allowed the authors to explore the impact of different pedagogical approaches including, but not limited to, student-prepared questions for the graduate student author that were used as a starting point for an in-class discussion of the research findings, techniques, considerations, and future goals. This conversation between general chemistry students and the graduate student researcher was coupled with an in-person laboratory tour or video introduction by the researcher. These activities humanized the research enterprise for the undergraduate students and serve as a model for future collaborations between research and instructional laboratories.



KEYWORDS: *First-Year Undergraduate/General, Graduate Education/Research, Curriculum, Interdisciplinary/Multidisciplinary, Laboratory Instruction, Laboratory Computing/Interfacing, Computer-Based Learning, Biophysical Chemistry*

■ INTRODUCTION

Competence in computing and data analysis has become increasingly important in preparing chemistry students for careers in industry and academic research. Problem solving, which includes the ability to "analyze data using appropriate statistical methods" and "to understand the fundamental uncertainties in experimental measurements", is included in the American Chemical Society's (ACS) "Guidelines for Bachelor's Degree Programs" list of skills that should be developed in an undergraduate chemistry program.[1]

In this paper, we describe a study in which ongoing scientific research was adapted for a general chemistry laboratory course. The instructors and teaching assistants (TAs) for the course collaboratively developed an assignment for students with the authors of a paper published in *Biophysical Journal*.[2] This research article provided an opportunity to introduce the analysis of spectroscopy data in the context of a recent study published by a graduate student in the Chemistry department. Using raw data provided by the research group and directions created by the paper's lead author, students analyzed the data, then wrote a summary and commentary paper. The activity description and directions, as provided to the students, are provided in the Supporting Information.

Several papers have been published describing the use of Microsoft Excel and other data analysis programs in teaching undergraduate chemistry and chemical engineering courses.[3−7] These activities have included examples where undergraduate chemistry students use linear and nonlinear regression models to calculate solution enthalpies according to the van't Hoff equation to investigate solubility trends for various anhydrous salts, to analyze the temperature dependence of heat capacities, or to investigate the kinetics of several chemical and radioactive decay reactions.[3,5,7] Computational programs used in the data analysis activities have included Gaussian '09, Microsoft Excel, Excel Visual Basics Application, and MATLAB.[4,6,7] Here, we designed an activity to teach data analysis using Excel or similar computational programs by selecting a journal article from a research lab in the home Chemistry department.

Received: September 4, 2015
Revised: December 2, 2015

The activity was designed to teach undergraduates about statistics relevant to analytical chemistry as well as to introduce them to primary literature in the field, another key skill identified in the ACS Guidelines. There have been many papers published that focus on teaching students to locate, analyze, evaluate, and cite scientific articles, usually utilizing database searches, outlining exercises, and student papers.[8−13] Increasing student exposure to primary literature can benefit all students, including those who are not pursing an undergraduate chemistry degree.

Here, we describe a practice in science education in which undergraduate students develop skills in computing and data analysis, as well as the ability to critically read and evaluate scientific literature, through interaction with a graduate student author. This collaboration was motivated by the desire to symbiotically improve both undergraduate and graduate education. We sought to teach undergraduate students how to use Excel (or similar analysis software) for data analysis within the context of real laboratory experiments. Many students begin this course with little or no experience with any data analysis programs. Because several other assignments in the course require software to generate graphs and analyze data, we first wished to introduce such software in a manner that went beyond the use of tutorials and provided an authentic research context. Second, we wanted to introduce students to reading and evaluating primary literature. For many of them, this would be their first time reading a scientific research article. Third, we intended that the collaboration would give students the opportunity to learn more about departmental research and researchers.

This type of collaboration also facilitates graduate-level education, as it provides graduate students with the opportunity to gain pedagogical experience by becoming involved with course development. Constructing chemistry curricula has been identified as a skill that can benefit chemistry graduate students and helps move advanced graduate students beyond their role as TAs.[14,15] Engagement with teaching has been found to positively impact critical research skills, such as generating hypotheses and designing valid experiments,[16,17] which highlights the importance of both research and teaching during the graduate school experience. Specifically, the graduate student involved helped identify learning objectives and then redesigned the activity based on achieved learning outcomes. Additionally, in adapting research for undergraduate education, the graduate student gained experience describing her research and making research problems accessible to introductory students and nonscientists.

■ CURRICULUM DEVELOPMENT

The data analysis activity was used in a course designed for introductory level undergraduates and served as a supplement to rather than a replacement for any existing lab activity. This course included students with widely divergent backgrounds and skill levels. For example, while the typical undergraduate was a student attending college directly out of high school, the course also included postbaccalaureate students considering a change in career after being out of school for many years. Students in the course varied from having no experience to extensive expertise with data analysis software. Three other experiments in the current curriculum require data analysis software, so it was critical that all students attain a sufficient level of confidence with such software. Several skills developed in this activity, such as using functions, referencing cells and workbooks, and making graphs, were also utilized in subsequent experiments. The data analysis activity described here was strategically situated before the spectroscopy experiment, the first experiment in the course requiring students to generate a graph. In addition to orienting students to Excel and similar data analysis software, this timing introduced and contextualized terminology that was to be discussed in the next class session, including wavelength, path length, and absorbance.

The article selected for the development of this exercise was deemed appropriate for several reasons: (1) the chemical concepts related directly to material covered elsewhere in the General Chemistry and General Chemistry Laboratory curricula, (2) the vocabulary introduced ideas to be discussed in the next lecture, thus providing context for the students when preparing for their next laboratory experiment, and (3) a subset of the experiments could be straightforwardly analyzed starting with raw data collected by the researchers.

The published article, whose lead author was a graduate student researcher, contained several sets of experiments, but only one was chosen for the data analysis activity. The researcher, after discussions with the course instructor, decided which type of analysis to incorporate into the activity. The particular analysis chosen was straightforward to implement, provided students the opportunity to analyze contextualized data, and connected to topics covered in the first weeks of the laboratory course (specifically, absorption, wavelength, and error analysis). Once a set of raw data was chosen, the graduate student researcher wrote a short introduction, tying in the research article topic to popular biomedical applications, and created an outline of steps that students would complete. The course instructor and researcher deliberated about how detailed these instructions would be and worked together to identify the teaching goals and strategy.

The portion of this article selected for the data analysis activity related to turbidity increase during collagen fibrillogenesis. Turbidity measurements during the collagen self-assembly process were collected at six wavelengths to reveal fibril diameter over the course of fibrillogenesis.[2] This biophysical experiment provided a link between the medical interests of many students and the visible spectroscopy experiment scheduled for the lab session following this data analysis activity. In the final iterations of this activity (Spring and Summer 2015), students were introduced to the article with a 17 min video recorded by the graduate researcher: this presentation (with audio) provided background on the role of collagen in the body, the research lab's interests, and an overview of the experimental design with data collection. During the activity, students were provided the handout with introduction and task-based directions described above (Supporting Information provides the current, revised versions). The directions avoided specific instructions on which buttons to click. The first set of students who participated in this activity (Summer 2014) were able to submit questions directly to the graduate student researcher. These questions and the researcher's responses were provided to future students as an additional resource when writing the final paper that required both a summary and a commentary section. The directions for the final paper (Supporting Information provides the current, revised version) placed emphasis on student reflection on their engagement with the data, and asked students to explicitly identify aspects of the

DOI: 10.1021/acs.jchemed.5b00710
J. Chem. Educ. XXXX, XXX, XXX−XXX

article they found challenging as well as to find links to material studied in the laboratory course.

RESEARCHER COLLABORATION

The initial adaptation of the published journal article for this activity was completed by the primary author of the article. The goals from this perspective for adapting the research publication for a laboratory classroom setting were two-fold: first, to choose one out of several experiments presented in the article to adapt with the objective of introducing undergraduate students to basic data analysis skills; second, to develop the graduate student's own skills in curriculum development. In achieving these goals, the primary author of the article collaborated extensively with laboratory instructors and TAs to identify factors for a successful adaptation. The graduate student researcher spent approximately 20 h in discussing and preparing the activity.

For the graduate student researcher, developing the article into a data analysis activity was an opportunity to be involved in curriculum development and approach teaching from a unique perspective. It was also an opportunity to explain the research to nonscientists. In subsequent semesters, with more students completing the exercise, it was not possible to include a lab tour and an introductory video was utilized instead.

CURRICULUM IMPLEMENTATION

With a draft of the instructions written by the graduate student researcher, the instructors and TAs implemented the assignment as part of the curriculum of a six-week general chemistry laboratory course during the summer session of 2014. The summer session was ideal for piloting this activity given the small class size; increased instructor to student ratio; and short course length with two six-week courses offered back to back, providing an opportunity to pilot the activity twice with adaptations between. There were two TAs (here, fourth-year undergraduate chemistry majors) and one faculty instructor for each six-week course. We note that well-prepared undergraduate students often serve as TAs in the department's summer session courses. The implementation of this assignment was an opportunity for these TAs to gain experience in the iterative process of course design, implementation, and redesign.

We carried out the first implementation of this data analysis activity by assigning mandatory time in the department's computer lab for the students to work on the activity for a 2 h period. Students were given and told to read the original article and instructions and to complete the activity described in the instructions. Students were not required to complete the activity in that 2 h period, but instead instructed to complete as much as possible in class, with the understanding that there would be an additional dedicated drop-in office hour with the instructor and a TA available to answer questions.

Because many students had little or no experience either using Excel or reading scientific journal articles, there was initial resistance to carrying out the assignment. The instructor and TAs had to convince students that struggling through reading the article and understanding the steps in the assignment was not only normal, but actually integral to the process of reading and understanding scientific literature and to processing data. However, their feedback was also used to clarify and modify the instructions for the next set of students. Although the second set of students struggled with the assignment, they completed the activity in less time than the first section, despite having no more experience on average with Excel.

The next implementation of the assignment occurred during the fall semester. Several logistical changes were made to compensate for the larger class size. Time was reserved in the department's computer lab during every lab section for 2 weeks. After students completed the week's experiment, they were encouraged (but not required) to go to the computer lab and work on the data analysis assignment. The number of steps in the assignment was also reduced, removing the creation of graphs that were critical for the researchers but did not introduce new skills to the students. In addition to the undergraduate and graduate TAs assigned to each section, either the instructor or a TA familiar with the assignment from the initial summer sessions was available to answer student questions. Students were provided with modified instructions and the original article as well as with newly produced instructional videos. In these videos, critical steps of the assignment were demonstrated (links to these videos are embedded in Supporting Information documents). To avoid students copying the numbers they saw in each video, care was taken to provide details on working with Excel and using specific commands without displaying the numbers in the sheet. The videos were crafted without audio to facilitate multiple students viewing them concurrently in a single computer lab.

At the start of the activity, students directed their questions to instructors and TAs, and one-on-one help was provided for the initial steps requiring basic software techniques, e.g., inserting formulas or linking data between sheets in the same file. However, as students became more comfortable with the language of the program, they began to teach each other how to complete steps and troubleshoot difficulties they encountered. Initially, this peer teaching emerged spontaneously; once it was noticed, the instructors encouraged it further by, for example, asking a student on step 4 to help a student struggling with step 3.

After performing the data analysis activity described in this article, students were given a writing assignment based on the *Biophysical Journal* article from which this data analysis activity was developed. In the initial implementation, students wrote a one-page summary followed by a one-page commentary in which students reflected on the process of engaging with the complex and technical scientific literature. Analysis of the writing assignments suggested students struggled to paraphrase the research when writing their summaries, and in later iterations, the writing assignment was redesigned to decrease the summary section to half a page, providing a page and a half of space for the commentary section. This change decreased student paraphrasing of the original article and offered more opportunity for reflection on the reading and data analysis process.

The impact of this activity on student understanding was observed in part through this writing assignment. Student integration of the article and data analysis skills was evident. In the writing assignment, many students discussed the relationship between absorbance and turbidity. This was integral to the article as well as to the last step of the data analysis activity, which required recreating a graph from the article with multiple data sets, showing the relationship between absorbance and time for each of the wavelengths investigated.

DOI: 10.1021/acs.jchemed.5b00710
J. Chem. Educ. XXXX, XXX, XXX−XXX

Table 1. Implementation of Data Analysis Activity per Semester

Semester	Number of Students	Number of Steps	In-Class Support, h	Additional Support, h	Av. Time Students Spent[b] on Activity In-Class/Out-of-Class, h
Summer 2014	29	11	4	3	3.4/4.2
Fall 2014	125	5	2	2	2.0/2.0
Spring 2015	272	6	2	—[a]	1.8/1.1
Summer 2015	125	6	2	—[a]	1.7/2.1

[a]These semesters only office hours were available for additional support. [b]Students reported this information on the survey.

■ CURRICULUM ASSESSMENT AND REDESIGN

After completing the data analysis activity, students were asked to complete a survey (for which IRB approval was obtained) that took approximately 6 min. The survey was administered 2 days after the activity each of the four semesters in which the activity was performed (Summer 2014, Fall 2014, Spring 2015, and Summer 2015) and left open for 3 weeks. In total, 190 of the 545 enrolled students completed the survey; the authors believe the low response rate is due to the voluntary and anonymous nature of the survey and is consistent with responses to voluntary and anonymous evaluations of the course. In addition to survey data, the course instructors read the two-page papers written by each student. These papers provided insight into student engagement with both the article and the data analysis activity.

Initial survey data indicated that Summer 2014 students were frustrated by lack of specificity in the directions and spent an average of 7.6 h completing the 11 steps. Subsequently, the directions were edited to provide additional tips, and an introductory video was recorded by the researcher. Following this, the Fall 2014 students spent an average of 4.0 h completing the activity. The goal was to streamline the activity so it could be completed within 2 h by the majority of the students (with additional help provided during office hours). By clarifying the directions and removing redundant skills, the time required for the activity was reduced while critical components were maintained as assessed by the self-reports of skills and the quality of the written papers. Specifically, the activity was decreased from 11 to five steps by removing data processing that repeated data analysis skills already covered then increased to six steps to improve clarity. By the Spring of 2015, the majority of the students who took the survey reported completing the activity in less than 4.0 h (Table 1). This suggests either further refinement and instruction are needed or the instructors should adjust their time expectation for the assignment.

■ SURVEY RESULTS AND DISCUSSION

A majority of students responding to the survey (75%) suggested this activity be maintained in the course curriculum citing a long list of skills gained by engaging with data analysis software in the manner described here. Specifically, students mentioned linking between sheets, entering equations, and adding custom error bars to a graph as new skills they had developed. Students in the first iteration of the activity, who had an opportunity to meet with and discuss the research directly the graduate student author, commented very positively on this experience, suggesting that this humanized the research experience. Several students commented on the fact that this interaction revealed that the graduate student scientist was a "regular person". One comment capturing this sentiment was, "Just seeing that she was like any one of us, and was able to conduct an experiment that seemed complicated and daunting at first, even to her made me feel confident that with the right preparation and guidance, I would be able to do something similar one day." Such responses highlighted the importance of students meeting and talking to someone who is engaged in research.

Nearly all students who completed the survey reported that their confidence and skills using Excel increased or remained the same after completing this data analysis activity (Figure 1, Table 2). A significant difference was observed in the self-reported before-and-after Excel skills between the students who suggested we maintain this activity in the course and the students who suggested we remove it (Table 3). Both populations showed reported gain in their Excel skills but those who suggested maintaining the activity showed the greater gain. These students also had slightly lower self-reported Excel skills (2.2 for the students suggesting the activity be maintained versus 2.4 for those suggesting we remove it). Because a self-reported measure of gain was used, it is possible student frustration impacted how they reported their skills.

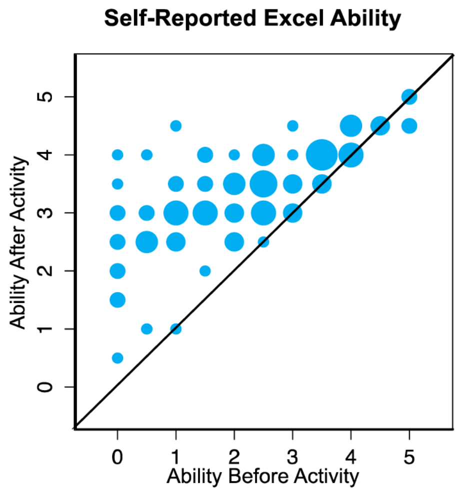


Figure 1. Self-reported student ability using Excel before and after carrying out the data analysis activity. Spot size corresponds to number of students, with the smallest spot representing 1 student and the largest representing 6 students.

Table 2. Self-Reported Excel Ability before and after Participating in Excel Data Analysis Activity

Participation Status	Survey Confidence and Skill Scores, N = 176	
	Mean[a]	Standard Deviation
Before Activity	2.239	1.409
After Activity	3.321	0.890

[a]Student's t-test: $t = 14.117$; $p < 0.01$; degrees of freedom = 175.

DOI: 10.1021/acs.jchemed.5b00710
J. Chem. Educ. XXXX, XXX, XXX−XXX

Table 3. Gain in Self-Reported Excel Skills by Students Suggesting Maintaining or Removing the Activity from the Course in Future Semesters

Self-Reported Gain among Students Suggesting...	Survey Skill Gain Scores	
	Mean[a]	Standard Deviation
Activity Be Maintained ($N = 132$)	1.197	1.066
Activity Be Removed ($N = 44$)	0.739	0.766

[a]Welch t-test: $t = 3.093$; $p < 0.01$; degrees of freedom = 102.

Moving forward, it would be useful to develop an independent measure of Excel skills to address this possible issue.

Several student concerns remained even after refinement of the activity. Students mentioned that the chemistry in the article did not relate to the labs completed in the course. This perception is likely linked to instructor choice to place this activity before the spectroscopy experiment. Going forward more effort should be made to explain the instructor's reasons for situating this activity before the relevant lab. A few students also requested more direct instruction regarding data analysis and terminology from the article. However, because the activity was intended to provide structured support for learning to face the challenges of troubleshooting data analysis software, the graduate student researcher and instructor maintain it is important to avoid step-by-step instructions. In general, student feedback was strongly positive, with several students commenting on the peer-teaching aspect that naturally emerged during the activity, as encouraged through the absence of step-by-step instructions. Student comments included, "Teaching Excel this way is good, with the group sessions so people can talk and figure things out together" and "I think it's a good assignment and a great way to connect computing/Excel with real lab work."

CONCLUSIONS

A collaborative approach to curriculum design provides a learning experience for the curriculum designers and the students in the course. By pairing with a research lab to share data analysis techniques, we found an opportunity to discuss pedagogical decisions with a researcher and authentic research considerations with undergraduates. This mentoring of graduate students to develop their research into assignments for the teaching lab opens opportunities for graduate students to gain valuable curriculum development experience. Challenging students to read complex literature and supporting them through the data analysis proved to be a successful approach to teaching data analysis techniques. Moving forward in the course, little to no support was needed in subsequent laboratory and data analysis activities for the spectroscopy, acid−base, and kinetics laboratories.

Here, we demonstrated successful integration of ongoing research into instructional laboratory courses in the context of teaching basic data analysis skills. The authors suggest the following be taken into consideration when adapting a similar exercise for a course:

1. Articles should be based on research that is appropriate for students at their particular level. The subject matter may be foreign but should build upon concepts that the students are familiar with from the course itself. The depth of information in the article may intimidate students, but their interaction with the article will be a more realistic representation of how young researchers (in contrast to experts) achieve science literacy.
2. Articles must contain an experiment (or portion thereof) amenable to an exercise that allows students to manipulate raw data. Having students interact with experimental data is an important component of the exercise, since real data will require students to use critical thinking to analyze the data and prioritize results.
3. The authors strongly encourage choosing articles from research laboratories within the department or the same academic institution. This connection is meaningful as it encourages graduate student involvement and allows students to immerse themselves in local graduate-level research. Whenever possible, face-to-face interactions between the researchers and the students are encouraged to maximize the positive elements of the exposure to the local research environment.

ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available on the ACS Publications website at DOI: 10.1021/acs.jchemed.5b00710.

Data analysis activity directions (PDF)
Writing assignment (PDF)
Survey questions from the final iteration of this activity (Spring/Summer 2015) (PDF)

AUTHOR INFORMATION

Corresponding Author

*E-mail: sjh2115@columbia.edu.

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

The authors wish to thank the students who participated in this activity.

REFERENCES

(1) Committee on Professional Training, ACS Guidelines and Evaluation Procedures for Undergraduate Bachelor's Degree Programs. 2015. http://www.acs.org/content/dam/acsorg/about/governance/committees/training/2015-acs-guidelines-for-bachelors-degree-programs.pdf (accessed Jul 2015).
(2) Zhu, J.; Kaufman, L. J. Collagen I Self-Assembly: Revealing the Developing Structures that Generate Turbidity. *Biophys. J.* **2014**, *106*, 1822−1831.
(3) Brown, P. J. Understanding Solubility through Excel Spreadsheets. *J. Chem. Educ.* **2001**, *78*, 268.
(4) Cress, N. L.; Robinson, M. A.; Corner, L.; Legge, R. L.; Ricardez-Sandoval, L. A. Problem-solving and concept integration using a computational tool in first-year undergraduate chemical engineering. Teaching Kinetics Using Excel. *Education for Chemical Engineers* **2012**, *7*, e133−e138.
(5) Loyson, P. Teaching Kinetics Using Excel. *J. Chem. Educ.* **2010**, *87*, 998−998.
(6) Wong, K. W.; Barford, J. P. Teaching Excel VBA as a problem solving tool for chemical engineering core courses. *Education for Chemical Engineers* **2010**, *5*, e72−e77.
(7) Martini, S. R.; Hartzell, C. J. Integrating Computational Chemistry into a Course in Classical Thermodynamics. *J. Chem. Educ.* **2015**, *92*, 1201−1203.

DOI: 10.1021/acs.jchemed.5b00710
J. Chem. Educ. XXXX, XXX, XXX−XXX

(8) Bennett, N. S.; Taubman, B. F. Reading Journal Articles for Comprehension Using Key Sentences: An Exercise for the Novice Research Student. *J. Chem. Educ.* **2013**, *90*, 741−744.

(9) Drake, B. D.; Acosta, G. M.; Smith, R. L. An Effective Technique for Reading Research Articles - The Japanese KENSHU Method. *J. Chem. Educ.* **1997**, *74*, 186.

(10) Tilstra, L. Using Journal Articles to Teach Writing Skills for Laboratory Reports in General Chemistry. *J. Chem. Educ.* **2001**, *78*, 762.

(11) Gawalt, E. S.; Adams, B. A Chemical Information Literacy Program for First-Year Students. *J. Chem. Educ.* **2011**, *88*, 402−407.

(12) Jensen, D.; Narske, R.; Ghinazzi, C. Beyond Chemical Literature: Developing Skills for Chemical Research Literacy. *J. Chem. Educ.* **2010**, *87*, 700−702.

(13) Roecker, L. Introducing Students to the Scientific Literature. *J. Chem. Educ.* **2007**, *84*, 1380.

(14) National Research Council, Challenges in Chemistry Graduate Education: A Workshop Summary; The National Academies Press: Washington, DC, 2012; http://www.acs.org/content/acs/en/about/governance/acs-presidential-commission-on-graduation-education-in-the-chemical-sciences.html (accessed Nov 2015).

(15) ACS Presidential Commission, Advancing Graduate Education in the Chemical Sciences. 2012; http://www.acs.org/content/dam/acsorg/about/governance/acs-commission-on-graduate-education-summary-report.pdf (accessed Jul 2015).

(16) Feldon, D. F.; Peugh, J.; Timmerman, B. E.; Maher, M. A.; Hurst, M.; Strickland, D.; Gilmore, J. A.; Stiegelmeyer, C. Graduate Students' Teaching Experiences Improve Their Methodological Research Skills. *Science* **2011**, *333*, 1037−1039.

(17) Ethington, C. A.; Pisani, A. The RA and TA experience: Impediments and benefits to graduate study. *Research in Higher Education.* **1993**, *34*, 343−354.

DOI: 10.1021/acs.jchemed.5b00710
J. Chem. Educ. XXXX, XXX, XXX−XXX