Graduate Students’ Teaching Experiences Improve Their Methodological Research Skills

David F. Feldon,1,* James Peugh,2 Briana E. Timmerman,3 Michelle A. Maher,4,5 Melissa Hurst,4 Denise Strickland,3 Joanna A. Gilmore,6 Cindy Stiegelmeier7

Science, technology, engineering, and mathematics (STEM) graduate students are often encouraged to maximize their engagement with supervised research and minimize teaching obligations. However, the process of teaching students engaged in inquiry provides practice in the application of important research skills. Using a performance rubric, we compared the quality of methodological skills demonstrated in written research proposals for two groups of early career graduate students (those with both teaching and research responsibilities and those with only research responsibilities) at the beginning and end of an academic year. After statistically controlling for preexisting differences between groups, students who both taught and conducted research demonstrate significantly greater improvement in their abilities to generate testable hypotheses and design valid experiments. These results indicate that teaching experience can contribute substantially to the improvement of essential research skills.

A

cademic culture in doctoral research universities’ STEM (science, technology, engineering, mathematics) programs typically values research activity over teaching (1, 2). Faculty commonly believe that research activities enhance teaching quality but disbelieve that teaching similarly enhances research skills (3, 4). These beliefs influence not only the professional priorities of STEM faculty, but also the guidance given to and the expectations of their graduate students (5, 6).

Previous research in educational and cognitive psychology suggests that a beneficial relationship between teaching and research skill development can exist to the extent that they entail an overlap of cognitive processes. When teaching in a context that requires students to effectively conceptualize research and solve problems through inquiry (for example, frame testable hypotheses, design valid experiments, or draw appropriate conclusions based on data), instructors must practice these skills themselves as they reason through these problems in order to provide appropriate guidance to their students. When students are trying to solve different problems, the instructor must likewise consider the discrete goals, structure, and methods of each problem, entailing practice in the relevant cognitive skills (7). In contrast, a research assistantship in a laboratory probably provides fewer, relatively similar projects that are based on the research agenda of the lab or principal investigator. Further, many high-level research design issues are likely to be resolved without requiring the research assistant to make substantive contributions to, for example, specifying research questions or determining methodology. For graduate students new to a lab, it is likely that the funded grant proposal supporting their work was written and submitted before their arrival.

Additionally, when learners are required to articulate their reasoning processes substantial evidence indicates that they develop more elaborate and effective schemas for problem-solving that facilitate performance on both typical and new problems (8, 9). Therefore, when instructors explain their own research processes to guide their students (10) they are further reinforcing their own learning. Research assistantships do not necessarily require extensive self-explanation (11).

Several small, qualitative studies report benefits of teaching for graduate student participants’ research development. One found that 21 of 27 teaching assistants leading undergraduate labs reported positive benefits to their research skills as a result of their teaching experiences (12). Another found that 33% of research advisors supervising participants in a National Science Foundation (NSF) GK–12 program (13) directly attributed improvements in participants’ research performance to their involvement with the program (14). Likewise, a RAND Corporation study found that STEM graduate students participating in

Fig. 1. Effect of both research and teaching experiences compared with research experiences alone for STEM graduate students’ improvement in writing testable hypotheses. After statistically controlling for pre-existing differences in the quantity of prior research experience, scientific reasoning ability, and earned scores on the written research proposal at the first time point, the quality of the hypotheses proposed were significantly higher in the teaching-and-research condition (Cohen’s d = 0.58). Error bars represent 95% CIs around the adjusted means.

Supporting Online Material
www.sciencemag.org/cgi/content/full/333/6045/1034/DC1
Materials and Methods SOM Text Figs. S1 and S2 Tables S1 and S2 References 6 May 2011; accepted 11 July 2011 10.1126/science.1207998
in educational outreach frequently reported that teaching helped them to reframe their understandings of their respective science domains to explain it to their own students (15). In a larger, quantitative survey of graduate students at one university (n = 524 students), participants who served as both research assistants and teaching assistants self-reported higher subsequent conference presentation and publication rates than that of those who served in only one role (16).

What each of these studies lacks, however, is a comparison of those who served in only one role (enrolled in the first three years) graduate students with those who engaged in both teaching and research experiences (21). Participants partnering with middle school teachers of STEM content (22). We predicted that those participants who engaged in both teaching and research activities (n = 49 participants) would exhibit substantially greater improvement in certain research skills (setting proposed research in the context of its field, use of primary literature, testability of hypotheses, research and experimental design, establishing reliability and validity of measures, selection of data for analysis, analysis of data, presentation of results, basing conclusions on data, and identifying study limitations) than would those engaged solely in research activities (n = 46 participants).

Participants were enrolled as full-time graduate students in research-oriented master’s and doctoral degree programs in empirical STEM disciplines at one of three universities in the eastern United States (22). One was a large, doctoral university (undergraduate enrollment ≈ 20,000; graduate enrollment ≈ 6700), and two selectively offered research-intensive masters degrees in STEM fields. Of the two master’s institutions, one was large (undergraduate enrollment ≈ 14,000; graduate enrollment ≈ 4000), and one was small (undergraduate enrollment ≈ 8200; graduate enrollment ≈ 500). Data were collected from three annual cohorts between 2007 and 2010.

Participants submitted research proposals related to their academic focal areas in early fall. Before submission, participants were given detailed instructions to include descriptions of the relevant literature and design for their proposed research, as well as anticipated results, other potential outcomes, and the importance of these results. Participants were also given a summary of the evaluation criteria. They then revised these proposals over the course of the academic year and resubmitted them in late spring as part of their participation in the study. The team conducting the study provided no feedback to the participants between the fall and spring submissions, although participants were free to seek independent feedback from other support networks and their programs at their discretion.

Most participants reported during exit interviews that they used their proposals for an additional purpose beyond the research study, such as to meet requirements for a class, research lab, or conference proposal. This information was interpreted as a positive indicator of both ecological validity and legitimate effort invested in the task.

The research skills addressed specifically in this study were setting context for a study, framing testable hypotheses, attention to validity and reliability of methods, experimental design, appropriate selection of data for analysis, presentation of data, data analysis, basing conclusions on data, identifying limitations, and effective use of primary literature. These criteria were selected through a review of relevant literature and iterative development of criteria with STEM research faculty (20, 22). At least two raters scored each proposal, and any discrepant scores were resolved by discussion until consensus was reached (23). Raters possessed graduate degrees in relevant STEM disciplines and attained inter-rater reliability intraclass correlations of 0.6 to 0.9 when scoring participants’ research proposals before discussion.

Rubric scores were grouped into three content areas: introduction (encompassing rubric element scores for setting the work in context, use of primary literature, and testability of hypotheses), results (encompassing rubric element scores for research and experimental design, establishing reliability and validity of measures, selection of data for analysis, analysis of the data, and the presentation of the results), and discussion (encompassing rubric element scores for conclusions based on data and identifying the limitations of the study). Multivariate analyses of covariance (MANCOVAs) were conducted in Mplus Version 6.1 (Muthén and Muthén, Los Angeles, CA) to appropriately model the statistically significant correlations among the rubric scores within each of the three content areas (introduction criteria correlations, 0.44 to 0.64; results criteria correlations, 0.26 to 0.69; discussion criteria correlation, 0.29). Further, all response variable rubric scores had 1.1 to 2.0% missing data at the first time point and 14.7% missing data at the second time point. A missing values analysis (χ²(17) = 23.20, P = 0.14) showed that the missing data met the assumption for missing completely at random (MCAR) (24). However, to preserve the sample size for analysis the missing data were handled more conservatively under missing at random (MAR) (25) assumptions by using a maximum likelihood estimation algorithm robust to nonnormally distributed data (MLR) (26). Because participants were not randomly selected or assigned to conditions, several covariates were used to statistically control for pre-existing differences between the groups assessed at the first time point: quantity of participants’ prior research experience, scores on two tests of scientific reasoning, and the rubric scores from their first research proposal submission (22).

We performed testing for significant mean differences between the two independent variable groups in three steps. First, MANCOVA analyses enabled the direct statistical test of the null hypothesis that a given rubric score element mean difference (teaching and research group mean minus the mean for the research-only group) was zero. Second, the analysis of 5000 bootstrap samples of size n = 95 participants enabled the computation of 95% confidence intervals (CIs) for each rubric score mean difference. Third, Cohen’s d effect sizes were computed for all mean differences, and Monte Carlo analyses of 5000 generated data sets of size n = 95 participants enabled the determination of the number of times in 5000 samples the null hypothesis (H₀) of a zero

---

**Fig. 2.** Effect of both research and teaching experiences compared with research experiences alone for STEM graduate students’ improvement in experimental design. After statistically controlling for pre-existing differences in the quantity of prior research experience, scientific reasoning ability, and earned scores on the written research proposal at the first time point, the quality of the experimental designs proposed were significantly higher in the teaching-and-research condition (Cohen’s d = 0.63). Error bars represent 95% CIs around the adjusted means.
mean difference for all rubric score elements was rejected. Univariate statistical tests of the observed mean differences between the teaching-and-research and research-only conditions indicated significant results for the rubric score elements “testability of hypotheses” [mean difference = 0.272, P = 0.006; CI = (0.106, 0.526)] with the null hypothesis rejected in 99.3% of generated data samples (Fig. 1) and “research/experimental design” [mean difference = 0.317, P = 0.002; CI = (0.106, 0.522)] with the null hypothesis rejected in 100% of generated data samples (Fig. 2).

These findings indicate a medium effect size for teaching and research experiences’ impact on participants’ abilities to generate testable hypotheses (Cohen’s d = 0.40) and valid research designs (Cohen’s d = 0.478) in the context of written research proposals (27.4 and 32.9% nonoverlap between teaching-and-research and research-only distributions for hypotheses and experimental design, respectively) (27). Differences in overall writing quality cannot account for the observed effects because only specific skills showed differential outcomes as a function of experience type.

These data provide direct, performance-based evidence of improvement on specific research skills associated with teaching experiences that complement traditional graduate research training. As such, they hold substantial implications for both the programmatic graduate training in STEM and the challenges that universities face as they strive to meet increased demand for instruction with fewer resources. The reframing of teaching experience as a value-added component of graduate research training suggests several substantial changes for the culture and practice of graduate education in STEM disciplines. Further, if teaching becomes a more commonly supported facet of STEM graduate education then students’ instructional training and experiences would alleviate persistent concerns that current programs underprepare future STEM faculty to perform their teaching responsibilities (28, 29).

References and Notes
13. The NSF GK-12 program provides funding for graduate students in STEM disciplines so as to gain experiences teaching in K-12 classrooms during their degree programs—often by co-teaching with a full-time credentialed instructor.
15. V. L. Williams, “Merging University Students into K-12 Science Education Reform” (RAND, Santa Monica, CA, 2002).

Mutational Inactivation of STAG2 Causes Aneuploidy in Human Cancer

David A. Solomon,1,2 Taeyeon Kim,3 Laura A. Diaz-Martinez,3 Joshleal Fair,4 Abdel G. Elkahlioun,3 Brent T. Harris,4 Jeffrey A. Toretsky,4 Steven A. Rosenberg,5 Neerav Shukla,6 Marc Ladanyi,6 David A. Solomon,1 Taeyeon Kim,1 Laura A. Diaz-Martinez,3 Joshleal Fair,4 Abdel G. Elkahlioun,3 Brent T. Harris,4 Jeffrey A. Toretsky,4 Steven A. Rosenberg,5 Neerav Shukla,6 Marc Ladanyi,6 Yardena Samuels,3 David James,7 Houttao Yu,2 Jung-Sik Kim,1 Todd Waldman1*

Most cancer cells are characterized by aneuploidy, an abnormal number of chromosomes. We have identified a clue to the mechanistic origins of aneuploidy through integrative genomic analyses of human tumors. A diverse range of tumor types were found to harbor deletions or amplifications for the culture and practice of graduate education suggests several substantial changes for the culture and practice of graduate education in STEM disciplines. Further, if teaching becomes a more commonly supported facet of STEM graduate education then students’ instructional training and experiences would alleviate persistent concerns that current programs underprepare future STEM faculty to perform their teaching responsibilities (28, 29).

Pathogenesis and chromosome segregation is thought to generate the large number of genetic lesions required for a cell to undergo malignant transformation (3). It has been hypothesized that this instability is due to inactivating mutations in genes that control the mitotic checkpoint and chromosome segregation (4, 5). However, in the vast majority of human tumors the molecular basis of chromosomal instability and the aneuploidy it produces remains unknown.

To explore this question, we followed up on previous studies in which we used Affymetrix 250K single-nucleotide polymorphism (SNP) arrays to identify novel regions of amplification and deletion in human glioblastoma cell lines (6–8). In U138MG cells, we identified a region...