Training the Next Generation of Statistical Scientists

Eileen C. KING

George Cobb presents a provocative paper titled "Mere Renovation is Too Little Too Late: We Need to Rethink our Undergraduate Curriculum From the Ground Up." He discusses the need for a total revamp to the undergraduate program. Essentials to this renovation of the curriculum include the following: "exploit context, embrace computation, seek depth, flatten prerequisites, and teach through research."

I have been a statistician involved in team science since the early 1980's and have seen a major evolution of our profession. As a manager in industry, I determined the correct mix of educational training for our employees which included BS/BA, MS, and PhD prepared statisticians. At any point in time, approximately 30% of our employees were BS/BA prepared, many of whom continued their education while still working. The BS/BA statisticians would be very active and highly regarded members of multi-functional project teams and collaborated with data managers, physicians, senior statisticians and other team members on study design, data collection and cleaning, data analysis, and data presentation and interpretation.

In order to prepare our undergraduates to contribute to their maximum potential, a new approach to the educational process needs to be undertaken. In the sections that follow, I propose one method for preparing undergraduate students for a rewarding career as a statistical scientist that incorporates the imperatives laid out by Dr. Cobb and is in close alignment with the recently issued "ASA Guidelines for Undergraduate Programs in Statistical Sciences" (*http://www.amstat.org/education/curriculumguidelines.cfm*. I also discuss the need for providing opportunities for undergraduates to advance their educational training throughout their career by having access to top statistics programs without the need to relocate or leave their current position.

Training the BS/BA Statistical Scientists

Recruitment of top talent needs to happen at the high school level and within the AP calculus, statistics, computer science, and science classes. These students enter their post-high school educational process well prepared to become successful statistical scientists. Students expressing an interest in statistical science as an undergraduate major should be required to identify 1–2 areas of collaboration by the end of their first semester, e.g. biology, psychology, medicine, business. To facilitate this requirement, the following two-step process is recommended: 1) each student would be required to attend a seminar series with speakers from other divisions who present on their research so the student can identify a collaborating division, and 2) each student would write a proposal and make a presentation on his/her interest in working with the specified division. This requirement provides the first of many opportunities for the development of effective writing and presentation skills.

At the beginning of the second semester, students would then be paired with an allied department for their collaboration experience whereby they will be involved in working on projects with researchers in those areas for their entire undergraduate program (Cobb: "Exploit context—use research as a vehicle for teaching statistics"). Students could have the opportunity to change areas after their first two years but it is important that students learn the scientific area in which they consult so that they can be effective. A faculty member and/or advanced MS or PhD student within the statistics department must provide technical oversight and training to the undergraduate student. Based on this initial assignment, the curriculum for the student can be developed; hence, "context dictates content" [Cobb]. For example, if the collaboration requires knowledge of approaches for analyzing "big data" such as in the business world or informatics, then courses specific to this area can be taken (e.g. multivariate statistics, exploratory data analysis, data visualization).

Of course, all students will need to take a set of core courses including statistical theory. However, the method, prerequisites and timing of these courses within the curriculum should take into account the individual student's progression through the research experience [Cobb: flatten prerequisites]. This approach will provide the context for learning [Cobb: "practice usually leads and theory follows."] Technology today allows for intuitive ways of learning important theoretical concepts through simulation and algorithmic methods as opposed to "formulas that you can easily compute by hand" [Cobb: Embrace computation]. For example, concepts of p-values and confidence intervals are still necessary for making inferences, especially since our scientific colleagues rely on them (although I often question if they understand them) but the approach for constructing and interpreting can be made more intuitive [Cobb: Seek depth]. Sample size calculations through simulations should be a mainstay. Undergraduates must know statistical programming languages very well (e.g. SAS, R, etc) to be marketable. Structured programming skills and the importance of producing well-documented and validated programs that lead to reproducible results is imperative. Undergraduate students must also master the "soft skills" such as good oral and written communication skills, effective teamwork and collaboration, and the capability of making effective presentations. The proposed program will provide ample opportunities for development of these skills.

Internships after the sophomore and junior years should be available and strongly encouraged for the students pursuing a BS/BA in statistics. The proposed program structure provides early opportunities to develop the skills required for effective collaboration within multi-functional teams. As a result, they will be prepared for these internships and will add value for the

Online discussion of "Mere Renovation is Too Little Too Late: We Need to Rethink Our Undergraduate Curriculum From the Ground Up," by George Cobb, *The American Statistician*, 69. Eileen C. King, Research Associate Professor, Cincinnati Children's Hospital Medical Center, Cincinnati, OH (Email: Eileen.King@cchmc.org).

sponsoring organizations.

Challenges to Implementation

I recognize there will be many challenges to this approach. First, in order to provide the opportunity for students to pair with a department upon entry into the program, the college and/or university must have access to active research programs within or external to their institution. Another challenge is the availability of faculty within the statistics departments to provide oversight to these undergraduate students. The faculty member must be proficient in collaborative research with other departments in order to be an effective mentor for these students. The biggest challenge will be the willingness of other departments to provide research opportunities for undergraduate statistics students. These departments must see the reward of participation through effective and value added statistical collaborations that lead to publication and increased funding.

Attracting top talent that can develop into highly effective statistical collaborators is a barrier. It is my perception that students typically enter their undergraduate programs with the desire to "change the world." It is important that these students see the impact they can make as a statistical scientist on producing high quality research that can, in fact, "change the world" for the better. They must experience "team science" from the beginning of their educational program.

Supporting the Career of the BS/BA Trained Statistical Scientist

Part of the renovation should also include consideration of the entire professional career for BS/BA trained statistical scientists including opportunities for obtaining advanced degrees from highly regarded programs. It is common that these individuals will, after a few years, want to further their education without leaving a position that is fulfilling, provides a good income, and oftentimes provides money for advanced education. There are many opportunities through distance learning for students to obtain a MS degree. However, there are few, if any, options available for students who wish to continue the educational process by obtaining a PhD degree from a top program without relocating to the specific university. Today's technology allows for synchronous and asynchronous learning and effective collaborations with thesis advisors at all levels of the educational process.

Conclusion

The curriculum for undergraduates in statistical science must be renovated. I have proposed one approach that augments Dr. Cobb's discussion with the aim of attracting top talent into our programs and produces well-prepared and "in demand" BS/BA graduates.