

statistics is that we can infer quite reliably the opinions of a population of hundreds of millions based on a sample of a few thousands or even few hundreds, if the sample is a probabilistic one without any serious self-selection mechanism in play. In contrast, inferences of a population of 230,000 based on a highly self-selected 408 respondents (the actual sample size of the ASA/CB study) are not something that most (any?) professional statisticians would be willing to vouch for without carefully laying out all the heavy assumptions.

As a matter of fact, even if by luck our inferences for the 230,000 population are dead on, they still provide little information about how many have been so turned off by poorly taught AP classes that they decided not to take the AP exam or touch statistics ever again.

Most critically, thinking from the causal inference perspective, even with 100% response from everyone who ever took AP Stat classes, we still need to work hard to construct an appropriate “control group”, perhaps via propensity score matching, in order to come close to assessing the real impact of AP statistics. Again, ASA/BC correctly acknowledged, repeatedly, that this lack of comparison/control groups makes it “difficult to make comparisons and impossible to draw causal inferences.”

There is of course nothing special about AP statistics in this regard; the same caution and rigor should be exercised when evaluating other AP programs, and indeed any other educational program. However, in responding to my point that it is our professional responsibility to serve as a (not “the”) police of science (Meng, 2009), one reader wrote to me that we should self-police, namely to critically examine how we statisticians invoke statistical evidence and arguments in our own work before we police others. I, of course, completely agree: only when we practice what we preach can we convince others to take our advice seriously. But most critically, even if we don’t care about how sound our arguments are or how others think of us, this is a case where we really want to know how effective the AP program is in real terms for the sake of our own future. I therefore very much appreciate the many cautions taken by ASA/CB in presenting its findings. Indeed, the full report as posted on the College Board web site suggests that the overall conclusion of the study is substantially more cautious than the impression the quotes in RPFHS might generate.

As a matter of fact, the abstract of the ASA/CB full report reads

“Taking the AP Statistics course and exam does not appear to be related to greater interest in the statistical sciences. Despite this

finding, with respect to deciding whether to take further statistics course work and majoring in statistics, students appear to feel prepared for, but not interested in, further study. There is certainly more research needed in order to make causal inferences about the issues presented in this analysis. However, it should serve as encouragement for both AP Statistics and college statistics instructors and the broader statistical community that the AP Statistics program seems to be successful in preparing students for further study and in increasing interest in statistics.”

(The abstract was dated 2/28/2009, the same month the full report was issued as Patterson, 2009; and it was downloaded on Oct. 31, 2009 from <http://professionals.collegeboard.com/data-reports-research/cb/ap-statistics-education-choices>.)

In its “Conclusion” section, which was reproduced in the *Amstat News* (May, 2009), the entire paragraph was kept except that the word “not” was removed from the first sentence, and a more qualified “not” message was appended at the very end of the above paragraph: “*It may not, however, affect students’ choice to pursue statistics as a major.*”

This contradiction between the abstract and the conclusion could simply be an oversight or could reflect a compromise perhaps between different versions/revisions to prevent misquotations of the study.¹ From a critical self-policing perspective, adding “not” or not brings to mind the quote by Mark Twain, which was allegedly switched from “Some congressmen are ...” to “Some congressmen are not ...”—the limitations of the sample are just too great to draw any statistically sound general conclusions in either direction. I therefore fully agree with ASA/CB’s emphasis that “There is certainly more research needed in order to make causal inferences about the issues presented in this analysis.”

The Great Pressure We Are Under: Deliver Quantity And Quality

The research needed is not just about determining the quality of the existing AP program, but more critically to identify, insofar as possible, the mechanisms that have led to local “good, bad and ugly” implementations of an obviously very well intended program. Such research can help our profession make informed plans as how to improve or even reform it in order to achieve our dual goals to build and sustain a strong workforce for our profession’s future and to continuously raise statistical literacy in general. We are no longer in a stage where our central goal is merely to convince the general public, at least the scientific ones, of the importance of statistics. We are now very much desired, or even

¹ There is indeed at least one obvious discrepancy or typographical error in the full report; the text mentioned several times that response rate is higher for female, whereas the appendix listed that male’s response rate is 23.7% and the female’s rate is 15.0%

feared, as I argued in Meng (2009). We now need to *deliver*, not just in terms of quantity but more importantly quality. I put more emphasis on quality, because as we are all aware, without good quality, a product will not last for long even if it is highly demanded at some point.

A local painful experience which occurred when I became department chair reminds me well of the importance of not forgetting quality when quantity is being demanded. In 2004-2005, we had an unexpected increase of enrollment by about 85 students in our most basic introductory course, Stat 100. Therefore, literally at the last minute, we needed to find additional teaching fellows (TF) to staff 5 more sections, as Harvard's policies require one TF section for every 17-20 students. We delivered the quantity, by hiring anyone who "could breathe and count"—a sarcastic phrase that sadly was not too far from the reality. The result amounted to "a mini crisis": students complained, faculty complained, and even those TFs themselves complained because they were under pressure to do a job they clearly were not qualified for, nor were they given any training.

Given the grand challenges we are facing, as listed and discussed in detail in a number of very recent articles (e.g., Brown and Kass, 2009; Meng, 2009), it would not be pure speculation to suppose that our profession would be facing similar "mini crises" or even big ones if we only focus on quantity. Of course, this is a well recognized issue by many. For example, much effort by ASA's education department, such as the "Meeting within a Meeting" at JSMs and the STEW site (all available at <http://www.amstat.org/education/>) is about improving the statistical education quality at all levels. My emphasis here is that in an assessment of a national program such as AP statistics, assessing its quality is as important, if not more so, than assessing its quantitative aspects.

The real pressure here, as in many other situations when the demand exceeds the supply, is maintaining both high quality and high quantity, or even just adequate quality and adequate quantity. And for educational endeavors, there is also often a cascade effect. As my colleague, Joe Blitzstein (the "Youtube" sensation mentioned in my op-ed), commented on an early version of this article:

"I hope such surveys would be done in such a way as to give as much information as possible that would be useful in deciding how to reform AP Stat, not just to accurately estimate the causal effect. A major difficulty is that the more "interesting" AP Stat becomes (based on

thinking/ideas rather than cookbook-style mechanics), the more difficult it becomes to find qualified teachers."

Indeed, without enough qualified teachers, even the best designed curriculum and most well intended program can do more harm than help.

Joe is exactly right that in order to deal with this pressure, our most urgent task in terms of assessment is to learn as much as possible about the mechanisms that have led to "turn on" and "turn off" and, hence, we can be strategic with the limited resources we have. This key point is also emphasized by Kari Lock, a member of my "happy team," in her comments to me:

"If we know WHAT turns people on or off, then we can keep the aspects that are turning people on and turn our efforts to fixing the aspects that are turning people off. I think that a well-conducted study about AP statistics could be very powerful and very beneficial to our profession, but I think there might be more to gain from learning how we can increase P_{ON} and decrease P_{OFF} rather than getting accurate current estimates of P_{ON} and P_{OFF} ."

We perhaps all have our educated guesses about, or even direct experiences of, what may or may not work. Surely the curriculums and teacher's quality matter greatly, as does competition from other fields—we are not the only profession on the expanding horizon. (As discussed in Meng (2009), ironically, the rapid rising and evolution of other disciplines, such as life sciences, creates both demand of and competition for us.) And we may even know, at our local levels, which of those mechanisms are likely to be the dominating force.

However at the national level, I'd venture that currently we do not have a good understanding of the interplay of various factors. Many deep questions can and should be asked even if we yet need to estimate P_{ON} and P_{OFF} , at the national level. Here are a few that immediately come to mind.

- What kind of students are more likely being turned on or turned off? What kind of students are unlikely to be affected by the quality of the AP program?
- Are students turned on or off more because of course materials or because of their delivery? Or must it be the interaction of the two to have a strong effect, in either direction?

- For those “turn-off” courses, is the problem more due to teachers’ lack of teaching skill, lack of statistical knowledge, lack of experiences of statistics practice, or lack of enthusiasm they themselves have because of their own bad learning experiences? Or must it be a combination of two or more before driving students away?
- How widely spread is the notion that AP stat is a “softer alternative” to AP math? Is this largely due to the common perception that statistics is an “easy” branch of mathematics, or due to the effectiveness of the AP math program to attract top students, or due to students’ or even school administrations’ perceptions of who is getting to teach math courses and who is teaching stat courses?
- To what extent has the AP exam itself encouraged teaching for test-taking rather than for understanding and inspiring interest? In view of the rapid re-shaping of statistics by the scientific evolution, in what ways should we revise or even revamp the AP exam and curriculum in order to inspire and reflect students’ interests?

I am sure that none of these questions are new, and that the authors of RPFHS and readers have many more questions. As an example, here is an excellent and tough question from Joe Blitzstein, again in his comments to me:

“Somewhere you might mention explicitly what I consider one of the most fundamental downsides, and the statistical difficulties of measuring it. That is what economists would call the *opportunity cost* of taking AP Stat, the alternative that the AP stat students would have otherwise taken. In high school, they would probably take another math course (AP Calculus or at least pre-calculus) in place of AP Stat. Having more math may actually help the students in their later stat courses (if any). In college, many students who took AP Stat think they already know enough stat (so don’t take any more), whereas they might have gotten a much better intro to stat from the college course (this depends on how likely an AP Stat course is to be well-taught compared with a college intro course).”

Indeed, the ASA/CB study reported that having taken the AP stat course is a relatively

important reason for the respondents not to take any college-level course in statistics. Therefore, in order to address Joe’s question and more generally the overall impact of AP statistics, we need to also examine the quality of statistical education at the college level, where the great shortage of qualified teachers is also a well-known problem, as discussed above. As a matter of fact, one may well question to what extent the “turn-off” phenomenon exists at the college level, and whether it is in any sense better than at the AP level.

Evidently, questions like those listed above are often very hard to answer, and some perhaps are never answerable completely. Minimally, it will take significant human and other resources to conduct studies to address these questions. But I’d argue that the stakes are simply too high, especially considering how many other fields are competing intensely for future talent, for us not to give our “first gateway” to future statisticians the highest priority. I therefore urge ASA to assemble the best teams our profession can offer to conduct studies of the AP program regarding its current impact and future directions and improvement, building upon the ASA/BC study. The plural “studies” is intended to point out that what is needed here is not just a single assessment study, but rather an on-going process to keep ourselves on our toes, to identify feasible ways for improvements, both short term and long term, and to ultimately offer a continuously renovated platform for “gateway” statistical education at a national level. The overall success of the AP program would be undoubtedly an effective arsenal in releasing the great pressure our profession is under, that is, to provide both high quantity and high quality future statisticians. We certainly face a great challenge here, but with challenges often come advances. My colleagues and I are well reminded of this by the fact that our current success as a model department at Harvard started from the aforementioned “mini crisis.”

An Invitation to Join a Self-Policing Unit

I, of course, need to volunteer myself for the effort I am arguing for—asking questions is important but often is not as important as answering them. The aforementioned studies obviously need to be conducted by those who have been at the forefront of the AP program and who have intimate knowledge of its history, operation, and complexity, such as the authors of RPFHS. My knowledge of AP statistics, documented in this article, is enough for me to raise questions as an interested “outsider”, but negligible for what is needed to conduct serious studies. However, I am hoping to turn my lack of knowledge or involvement into what perhaps is best depicted

by the Chinese proverb “spectators may see more of a game than players.” When such studies are in place, I would like to invite a few statisticians to join me in forming a “self-policing unit” to provide an independent check of, and critical comments on, their design, analysis, and conclusions.

Sometimes when we statisticians carry out our “police duty”, we are ridiculed as being only interested in covering our own necks (or lower). But in the case of assessing AP statistics, it is not our necks but rather our entire profession’s future well-being at stake. No study of this sort can possibly be perfect, and it is anti-scientific to be overly cautious when facing such imperfections. At the same time, we can be confident that we are exercising the right level of cautiousness only when we hold ourselves to the highest possible standard given the practical constraints. Surely inferences always come with unavoidable uncertainties, but they should be free of avoidable mistakes, with or without passion, paradox, or pressure.

Let me conclude by thanking again the authors of RPFHS for their great inspiration. We definitely have a giant elephant in the room, to borrow a common idiom, but with a connotation echoing the “four blind men and an elephant” story: we need collective wisdom and joint effort, much more than just comparing and contrasting our individual findings. We need to understand the elephant as a whole, and, metaphorically speaking, we need to move it outside the room to help to carry the heavy load and responsibility on our shoulders, as our profession becomes increasingly desired (and feared), in reality and in perception. How many of us had expected, just a year ago, that “For Today’s Graduate, Just One Word: Statistics” would be an actual headline in *NY Times* (Aug 6, 2009)? ■

Acknowledgements: I thank Alan Agresti, Joe Blitzstein, Thomas Lee and Kari Lock for their very helpful comments. All “mis-takes of passions” are of course mine.

References:

- Meng, X.-L. (2009). Desired and feared—what do we do now and in the next 50 years? *The American Statistician*. 63, 202-210.
- Brown, E and Kass, R (2009). What is Statistics? (With Discussions.) *The American Statistician*. 63, 105-123.
- Patterson, B (2009). Advanced Placement® Statistics Students’ Education Choices After High School. Research Notes RN-38, February 2009, College Board. Reproduced in part in *Amstat News*, May, 2009.

2010 Joint Statistical Meetings, Vancouver, BC, Canada
ASA • ENAR • ICSA • IISA • IMS • SSC • WNAR

ABSTRACT SUBMISSION OPEN

The 2010 JSM Program Committee invites your participation. Submit an abstract to present one of the following:

- Topic-Contributed or Contributed paper
- Topic-Contributed or Contributed poster
- Topic-Contributed panel session

Submissions accepted December 1, 2009,
to February 1, 2010

www.amstat.org/meetings/jsm/2010

Online submission is secure, convenient, and accurate. It features automated typesetting and quick transmission to the program committee.



“Statistics: A Key to Innovation in a
Data-Centric World”

July 31–August 5, 2010

New Requirements for Travelers from the United States

As of June 1, 2009, everyone traveling between the United States and Canada, Mexico, Central and South America, the Caribbean, and Bermuda by land, sea (including cruises and ferries), or air will be required to present a valid passport or other documents as determined by the Department of Homeland Security to cross the border.

U.S. residents can access the following web sites for passport and visa information:

Passport: http://travel.state.gov/passport/passport_1738.html

Visa: http://travel.state.gov/visa/visa_1750.html