

How Random is That?

September 10, 2005

Compared to the hard rock of empirical methods, 18- to 20-year-old college students are a wet marsh of spontaneous behavior and malleable minds. In 1971, notable personality researcher Rae Carlson called students “unfinished personalities” who may fundamentally differ from non-students in a number of psychological ways. Fifteen years later, APS Fellow and Charter Member David O. Sears wrote in the *Journal of Personality and Social Psychology* that “college students are likely to have less-crystallized attitudes, less-formulated senses of self, stronger cognitive skills, stronger tendencies to comply with authority, and more unstable peer group relationships.” They change personal ideologies from lecture to lecture, scuttle to and fro as their hormones direct, wake up at six o’clock — in the evening. But despite being behavioral works-in-progress, college students remain the primary subject pool for most psychological researchers, leaving some to question whether findings from this “convenient” population can generalize to the world at large.

“The goal of psychology is to make nomothetic laws — laws that apply to all people,” said APS Fellow Lisa Feldman Barrett, Boston College. “The question is, how well can you do that when you’re sampling by convenience?”

The question is an important one, considering that in 1999, students made up 86 percent of the samples for subject-based articles appearing in the *Personality and Social Psychology Bulletin*, and 63 percent for the *Journal of Personality and Social Psychology*, according to a study led by APS Fellow Richard C. Sherman. Since its inception in 1992, the *Journal of Consumer Psychology* has included college samples in another 86 percent of its empirically based articles.

Though the numbers may seem alarming, asking why students are so widely used is like asking why breathing air is the preferred method for oxygen intake — the reasons range from the obvious to the more obvious. “They are a very convenient and captive subject pool that researchers can dip into with relative ease,” said Michael Hunter, University of Victoria. So convenient, they are commonly known as the “convenience sample,” often showing up at a researcher’s door as part of a requirement for an introductory psychology class.

The price is right, too, said APS Fellow and Charter Member Peter Killeen, Arizona State University. “They’re cheaper than white rats, and they’re more similar to the population to which we hope to generalize,” he said. “And they seldom bite.” Feldman Barrett believes that without these low-cost, easy-access samples, textbooks would be as empty as journals and her lab would be as empty as either; in such a scenario, she predicts being able to run merely a quarter of the experiments she does now.

Ironically, just about the only thing college samples have not been used to study are themselves. It is this lack of empirical confirmation that APS Fellow and Charter Member Harry Reis, editor of *Current Directions in Psychological Science*, calls the most definitive reason why students remain the default sample. “The suspicion that the literature is flawed because of reliance on college students has been

around for a long time,” said Reis, University of Rochester. “The objection is an obvious one, but I’ve never seen data, to date, showing that it is a serious problem.”

But like a formative undergrad, that might be about to change.

Emerging Evidence

Robert A. Peterson is not a psychologist, but he often plays one in the research lab. A professor in the McCombs School of Business at the University of Texas at Austin, Peterson’s research regularly crosses paths with consumer psychology and appears in publications like the *Journal of Applied Psychology and Psychology and Marketing*. Recently, he completed what he believes to be the most thorough empirical study on student samples, primarily because few such studies exist.

“How do you comprehensively analyze the results of psychological experiments? How good are data based on college student samples? These are questions we need to ask, since the vast majority of studies being published use students,” said Peterson, who has been interested in such methodological questions for 40 years. “If you look at the issue of using students there is virtually no empirical evidence supporting or challenging their use.”

Studies do exist on whether samples of convenience can produce research of significance, but many are anecdotal in nature, said Peterson, or are driven by logic and emotion. APS Charter Member Robert Dipboye and Michael Flanagan fell just short of implying that convenience samples were a serious problem in their 1979 paper, “Are Findings in the Field More Generalizable Than in the Laboratory?” Dipboye and Flanagan analyzed content from volumes of the *Journal of Applied Psychology*, *Organizational Behavior and Human Performance*, and *Personnel Psychology* to determine if field research was more generalizable than lab findings, as was the common belief. Instead, they found field research to be “as narrow as laboratory research in the actors, settings, and behaviors sampled” — hardly a glowing endorsement for either sample — and suggested using both populations whenever possible.

But few others have addressed the problem with Peterson’s scientific rigor. He and his assistants spent years scouring the literature for articles where student and non-student samples had been used in the same research. When the dust cleared, he had compared an estimated 650,000 student and non-student subjects — perhaps trying to make up for the lack of literature in one fell swoop. The result was his paper, “On the Use of College Students in Social Science Research,” and the first dent in the armor of student samples.

After analyzing 65 behavioral and psychological relationships, Peterson found that nearly one in five conclusions based on a college student may differ directionally from that of a non-student. He also found that 29 percent of the relationships differed in magnitude — the larger effect size in a pair of student and non-student studies exceeded the smaller one by a factor of two or more. Altogether, nearly half of the effect sizes observed for students and non-students “differed substantially.”

Despite presenting some of the topic’s most convincing evidence, Peterson’s conclusion was more cautionary than apocalyptic, suggesting outside replication of student-based research before any generalizations are made. “At a minimum,” he said, “research based on one sample of college students from one subject pool at one university needs to be replicated with students from a different university.” Instead of putting an end to one problem, however, the results bring attention to another — a core barrier

to generalizability: the difficulty of reaching a non-matriculating population.

So Many Barriers

Some of the ado may be about nothing. Conveniently enough, a number of disciplines seem tailor-made for the convenience sample. “In some areas — perception, memory, attention, many of the cognitive sciences — it typically doesn’t seem to matter,” said APS Fellow and Charter Member James Cutting, editor of *Psychological Science*. Whatever their behavioral idiosyncrasies, undergraduates maintain the same core neural networks, some argue. When studying attachment processes, college students might even be the ideal sample, Feldman Barrett said, since they are often under self-esteem threat and are the exact population to which such research should generalize.

Other areas in psychological science are not as lucky. As a developmental researcher, APS Fellow and Charter Member Valerie Reyna often lacks the luxury of calling upon a student sample. An undergraduate might roll out of bed and into the lab, but a child subject must be located, a parent paid (or, in more cases, convinced) to take off work and drive to the lab. “For people who do developmental psychology research, it’s a big job to set up the things necessary to do a study,” Cutting sympathized. “It’s hard to get subjects for experiments, and it’s harder to get non-college students.” Should Reyna or a colleague successfully find the child and get him or her to the lab, bureaucratic red tape makes it hard to secure a nearby parking spot. “Even then,” said Cutting, “an undergraduate might park there anyway.”

At times, Reyna’s biggest obstacles are obstinate school systems, hesitant to release information about a study to parents (forget about reaching children directly). “Some people understand that research is an important social good,” said Reyna, University of Texas at Arlington. “But some school systems do not let researchers give out information to see if someone would even potentially want to be in a study.” To tackle critical questions like the effectiveness of public health curricula in curbing risky adolescent behavior and the best ways to interview crime victims, “it is crucial to break down the barriers to locating and recruiting non-college populations,” she said. “Scientists are limited in the questions they ask because there are so many barriers in getting to outside populations.”

However difficult it is, reaching non-student populations remains important. Ask APS Member Tom Pyszczynski if reaching non-students takes a prohibitive amount of time, effort, and money, and he agrees — then says do it anyway. “When we do research with college students, we’re assuming that psychological processes are relatively universal,” said Pyszczynski, University of Colorado at Colorado Springs. “That’s an assumption — it’s not necessarily absolutely correct.”

To see how these processes are activated in different cultures, Pyszczynski often replicates a study in an entirely new population — comparing results from a convenient sample to those collected one state, or one hemisphere, away. To test his terror management theory, Pyszczynski asked middle-aged judges in Arizona to assign bond to a prostitute. Before assigning bond, some judges were primed with images of death, and others were not. According to the theory, people control a fear of death by buying into cultural belief systems — shared rules of behavioral conduct. A violator of these social rules, such as a prostitute, causes a fear of death and elicits a harsh reaction from the community. Judges reminded of death assigned prostitutes an average bond of \$450; control judges averaged only \$50.

Pyszczynski then replicated the study with college students who, lo and behold, also recommended a

higher bond when reminded of death first. In this case, however, their monetary figure depended on whether they approved or disapproved of prostitution, which varies more in students than in judges. Borrowing Pyszczynski's theory, German psychologist Randolph Oxman did the same study overseas and found that prostitution was not a great activator. At first glance, this seems at odds with both the judges' and the students' responses, until one point is made clear: prostitution is legal in Germany and thus seen as less of a moral transgression.

To Pyszczynski, these results show the importance of getting as many samples as possible, since even universal psychological processes may be activated differently depending on a type of people.

"Generalization is very important, but it's not a simple question of whether this study will yield the same results with a different person," he said. "The question is, if you translate your psychological variables to fit the subculture you're working with, would you find *equivalent* results?"

Cutting agreed that better journal submissions use many populations. One way to do this is by using Web-based research to augment laboratory findings, a method he sees happening more and more. This confronts two problems with the convenience sample: age and gender. If college students are roughly age 20, Cutting estimates, Web-based participants are five to 10 years older. In addition, it is common for two-thirds of Web samples to be male, which counteracts the predominance of females in classroom and laboratory experiments. (In 2004, nearly half the studies published in *Psychological Science* tested for gender; of these, 59 percent of the participants were female.)

"When you have a lab-based study and a Web-based study, and the results match, then that's nice," Cutting said. "One feels a bit more confident in generalizing."

This confidence is far from unanimous. Feldman Barrett acknowledged that Web samples may be more representative of the general population's age, but they carry their own limitations — namely, an above-average socioeconomic status. "You're still getting a non-representative sample on some dimension," she said. "Scientific conclusions will be drawn on the basis of studying those who can afford to access a computer with an Internet connection."

For once, somewhat refreshingly, the Internet might not provide the answers. Instead, a step forward might require revisiting a time when fundamental statistical methods ruled research.

Statistically Speaking

Feldman Barrett admits that some psychologists aren't careful enough about making generalizations — in part because they lack a sufficient understanding of statistics. "Most of us are not trained in sampling theory," she said. "There are sciences that take sampling much more seriously than we do."

Michael Hunter is trained in sampling theory, and he does take it seriously. Co-author of the 2002 paper "Sampling and Generalizability in Developmental Research," Hunter also sees this shift away from statistical methodology in psychological research, away from a time when using analysis of variance supported an experiment's causality and using multiple regression analyses helped infer generalizability.

"Today, and for some time now, psychological researchers judge generalizability not on statistical grounds but instead on methodology, educated guesses or common sense, and accumulated knowledge," Hunter said. Some of these techniques are fine to draw generalizations from, he said — sometimes even

highly effective — as long as they are based on random samples of the population in question. The problem, of course, is that college students are seldom a random sample of a university population, let alone a national one. “Obviously, this basis of generalizability does not augur well for the generalizability of some research with college students, who are a selective population and who are rarely randomly sampled.”

Fortunately, there is a way around taking a random sample from a broad population, says Peter Killeen. Unfortunately, that way is far less traveled. Instead of traditional statistical techniques, which are often inaccurate when used with samples of convenience, Killeen advocates permutation tests, which were difficult to implement before computers yet still remain untaught in this age of machines.

“Some researchers don’t use permutation methods because, while they let us make causal claims, they don’t support predictions to larger populations,” Killeen said. “But convenience samples preclude such generalizations in the first place. So, if we continue using convenience markets, such as introductory psychology classes, for our subjects, then we should at least get right with the statistical models we use to analyze their data.”

The permutation test Killeen referred to is often called the Pitman test, after Tasmanian statistician E.J.G. Pitman. According to Cliff Lunneborg, professor of statistics and psychology at the University of Washington, the Pitman test is known as a randomization test when used on convenience samples. The key to a randomization test is that it does not require a sample to come from a large population — in fact, randomization thrives on taking one random sample, such as a set of convenient university students, and randomly dividing it into two experimental groups, which can then be analyzed and concluded upon as though they had both been drawn randomly from a vast pool of participants. “Randomization tests should be much more widely taught than they are,” Lunneborg said. “Psychological researchers understand the value of randomization in controlling for experimental error. What they have not adopted is the notion that randomization provides a powerful basis for statistical inference as well.”

Whether such methods are adopted or remain empirical orphans, Hunter said there is one unequivocal way to ensure more generalizable results, whatever the sample: “The greatest gains in generalizability will come from better measurement rather than better sampling,” he said. “The best sampling scheme in the world will not overcome poor measurement.”

Cumulative Nature to Knowledge

Like a precocious freshman in an advanced seminar, who may be onto something others are not ready to hear, Peterson’s work has been met with reluctance. “Because the reviewers evaluating the submitted manuscript used college student samples, they were generally not predisposed toward the study’s findings and conclusions,” he said. “In general, I believe it is very difficult to get research published that goes against an existing methodological paradigm.” He has since had some difficulty publishing additional research on the topic, this time looking at difference in attitudes of over 3,000 students from 58 universities. Though he preferred not to have the study quoted since its publication remains uncertain, Peterson says his new findings take the old ones slightly further, strongly urging the scientific community not to generalize based on students samples, but instead to set theoretical boundaries within which further research must be performed.

Pyszczynski thinks such a parameter may be dead on. After all, science is not about answering all

questions with one study — strands of smaller generalizations can be sewn over time into a variegated tapestry of behavior. “Rather than study one population for all human beings, it’s better to come up with a conceptual model and adapt it for use with another population,” he said. “No one study makes a point that anyone should take seriously. There’s a cumulative nature to knowledge.”

If nothing else, new scrutiny on college samples emphasizes an even larger, more urgent concern in the behavioral community — the ability to fund experiments whose immediate application might be impossible, but whose theoretical contributions will lay the groundwork for improved public health. Reyna already sees this happening with aging research — an area that requires older participants. The National Institute of Aging has created incentives for people to go beyond college populations to study aging, she said. If other funding institutions made efforts to pool and pre-screen research volunteers, science could greatly expand its reach beyond the shadows of the ivory tower.

“Scientific organizations and agencies should communicate to the public how valuable their contribution to research would be, and should help create policies that remove barriers to participation,” Reyna said. “Sometimes people see research as a luxury, when in fact it’s a necessity.”

References

- Carlson, R. (1971). Where is the person in personality research? *Psychological Bulletin*, 75, 212.
- Sears, D. O. (1986). College sophomores in the laboratory: Influences of a narrow data base on social psychology’s view of human nature. *Journal of Personality and Social Psychology*, 51, 515-530.
- Sherman, R. C., Buddie, A. M., Dragan, K. L., End, C. M., & Finney, L. J. (1999). Twenty years of *PSPB*: Trends in content, design, and analysis,” *Personality and Social Psychology Bulletin*, 25, 177-187.
- Peterson, R. A. (2001). On the use of college students in social science research: Insights from a second-order meta-analysis. *Journal of Consumer Research*, 28, 250-261.
- Dipboye, R. L. (1979). Are findings in the field more generalizable than in the laboratory? *American Psychologist*, 34, 141-150.