changes in measures such as heart rhythm and white blood cells—even in asthmatics, whose damaged lungs contained up to six times as many particles as healthy people.

The explanation may be that it's not size or chemistry alone. The ultrafines used in the Rochester study were pure carbon black, but ultrafines in the real world are likely coated with metals and organic compounds, Frampton says. (Also, the researchers may need to test people with cardiovascular disease.) Likewise, sulfates may form the core of a particle that also contains nastier compounds such as metals, or they could change the chemistry of metals so they're more soluble in blood. Larger particles may irritate and inflame airways, exacerbating the toxicity of PM constituents such as organics and metals, says Costa. And particles may have different effects in the short term and after years of exposure. "It's far more complex than trying to decide which chemicals are toxic," says toxicologist Joseph Mauderly of Lovelace Respiratory Research Institute in Albuquerque, New Mexico.

Newer experiments are seeking to use more realistic mixtures. That became possible only a few years ago when researchers invented devices that can collect ambient air from outside a lab and concentrate the particles for use in experiments. Others are looking at pollutants from a range of sources. For example, Mauderly's group at Lovelace is conducting animal studies comparing particles from diesel engines, gas engines, wood smoke, cooking, road dust, and coal to pin down which type is most toxic. HEI, meanwhile, is sponsoring epidemiologic and toxicology studies that will take advantage of a new monitoring network at 54 sites that measures a finer breakdown of the chemicals in particles, such as sulfates, elemental carbon, and trace elements, than has been gathered previously. And EPA recently launched a $30 million, 10-year study led by University of Washington researchers that tracks correlations between these finer air pollution measurements and the health of 8700 people over age 50.

Down the road, this new information should help guide regulations—for instance, if carbon particles from wood burning were the main problem, or diesel engines, EPA could specifically target those sources. Controlling only mass, as EPA does now, might actually be counterproductive. For example, if larger $PM_{2.5}$ particle levels go down but levels of ultrafines do not, "that could make things even worse," Frampton says. That's because ultrafines tend to glom onto larger $PM_{2.5}$ particles, so they don't stay in the air as long when the larger particles are around.

Time to act

Those results won't be available for years, however, and EPA is under a court order to decide whether to tighten the current $PM_{2.5}$ standard by the end of 2005. EPA scientists in January recommended that the agency consider tightening the standards from the current annual average of 15 μg/m³ to 12 to 14 μg/m³, and the daily average from 65 μg/m³ to as low as 25 μg/m³. They also suggested replacing the PM_{10} standard with a new one for particles between PM_{10} and $PM_{2.5}$ to better target coarser particles between those sizes. In April, EPA's clean air advisory board will weigh in.

$PM_{2.5}$ levels have already dropped at least 10% since 1999 due to acid rain regulations and new diesel engine standards (see sidebar, p. 1860). They will fall further thanks to additional cuts in sulfates and nitrates from coal-burning power plants through new regulations issued this month and possibly the Administration's proposed Clear Skies program. But a tighter standard could trigger additional controls in areas with the highest particle levels, such as Los Angeles and the Northeast. Environmental and health groups as well as many scientists say that, as with tobacco smoke and lung cancer, policymakers can't wait for all the scientific answers before taking action to prevent deaths from dirty air.

–JOCELYN KAISER

U.S. Education Research

Can Randomized Trials Answer The Question of What Works?

A $120 million federal initiative to improve secondary math education hopes to draw on an approach some researchers say may not be ready for the classroom

When Susan Sclafani and her colleagues in Houston, Texas, received a $1.35 million grant from the National Science Foundation (NSF) to work with secondary math and science teachers, nobody asked them to demonstrate whether the training improved student performance. "All we had to do was produce qualitative annual reports documenting what we had done," she says. Sclafani thought that wasn't nearly enough and that NSF should be more concerned about whether the project helped students learn. Now, a decade later, she's in a position to do a lot more. And that's exactly what worries many education researchers.

Prove it. The Department of Education's Susan Sclafani wants to see more experimental evaluations in math and science education.

CREDIT: U.S. DEPARTMENT OF EDUCATION

As assistant secretary for vocational and adult education at the Department of Education (ED), Sclafani is championing a $120 million initiative in secondary school mathematics that is built in part on money shifted from the same NSF directorate that funded the Houston grant. The initiative, included in President George W. Bush's 2006 budget request for ED now pending in Congress, will give preference to studies that test the effectiveness of educational interventions in the same way that medical researchers prove the efficacy of a drug. Randomized controlled trials (RCTs) of new approaches to teaching math, Sclafani says, will help school officials know what works, and they can then scale up the most promising new curricula and teaching methods. "Randomized studies are the only way to establish a causal link between educational practice and student performance," she says.

But some researchers say that such trials won't tell educators what they need to know. And they believe their discipline is too young to warrant a large investment in experimental studies. "Rushing to do RCTs is wrong-headed and bad science," says Alan Schoenfeld, a University of California, Berkeley, professor of math education and adviser to both NSF and ED. "There's a whole body of research that must be done before that."

The proposed math initiative at ED would be a competitive grants program to prepare students to take Algebra I, a gateway course for the study of higher mathematics and the sciences. Applicants will be encouraged to use RCTs and quasi-experimental designs to measure whether the reform works, Sclafani says. The initiative comes at the same time the Administration has requested a $107 million cut in NSF's $840 million Education and Human Resources (EHR) directorate. The cuts include a phasing out of NSF's portion of the $240 million Math/Science Partnership program—a joint effort with ED to improve K–12 math and science education by teaming universities with local school districts—and a 43% decrease for the foundation's division that assesses the impact of education reform efforts (*Science*, 11 February, p. 832). Sclafani says this "reallocation of education dollars" reflects the Administration's eagerness for clear answers on how to improve math and science learning across the country. That's OK with NSF Director Arden Bement, who says ED is in a better position than NSF to implement reforms nationwide.

Although NSF watchers are unhappy with the proposed cuts to the foundation's education budget, a bigger concern for some education researchers is that ED may be overselling RCTs. It's unrealistic to think that RCTs and other quasi-experimental studies will magically produce answers about what works, they say. Before comparing the performance of students in the experimental and control groups (one receives the intervention, the other doesn't), researchers must study the factors affecting any change in curriculum or teaching methods, such as group vs. individualized instruction, or working with students whose native language is not English. Answering such contextual questions, the critics say, is similar to finding out whether a medicine needs to be taken before or after meals.

"You can design an RCT only after you've done all this work up front and learned what variables really count," Schoenfeld says. ED's approach, he argues, is likely to drive researchers to skip those necessary steps and plan randomized studies without knowing why an intervention seems to work.

Department officials insist that the time is ripe and have begun funding a handful of projects drawn from 15 years of work in curriculum development and teacher training, including efforts funded by NSF. One is a study of Cognitive Tutor, a computer-based algebra course for middle school students. Another looks at a new approach to training 6th grade science teachers in Philadelphia. Diana Cordova of the department's Institute of Education Sciences predicts that within 3 years, "they will tell us with reasonable certainty if an intervention can improve student learning."

Some of the researchers conducting these studies aren't so sure, however. One hurdle is convincing a large enough sample of schools to agree to randomization. "Everybody wants to have the treatment, nobody wants to have the placebo," says Kenneth Koedinger, a psychologist at Carnegie Mellon University in Pittsburgh, Pennsylvania, who's leading the Cognitive Tutor study. Another problem is inconsistent implementation across the experimental group. Allen Ruby, a researcher at Johns Hopkins University in Baltimore, Maryland, who is conducting the Philadelphia study, says that problems at two of the three schools involved could end up masking evidence of whether the training is working.

CREDIT: COURTESY OF ELK RIDGE SCHOOL

Luck of the draw. Elk Ridge School is one of three Philadelphia, Pennsylvania, schools implementing a grade 6 reform curriculum as part of a randomized controlled trial.

Schoenfeld predicts that these and other problems will confound any analysis. "The likely findings from this study would be something like this: Sometimes it works; sometimes it doesn't; and on average, the net impact is pretty slight compared to a control group," he says. "What do you learn from such findings? Nothing." On the other hand, Schoenfeld says, a detailed analysis of how the implementation was done at each school and how teachers and students reacted to it could tell educators the conditions under which it would be most likely to work.

The still-emerging field of evaluation research needs investments in both qualitative and experimental studies, says Jere Confrey, a professor of mathematics education at Washington University in St. Louis, Missouri, who last year chaired a National Research Council report on the need to strengthen evaluations (*Science*, 11 June 2004, p. 1583). "You need content analysis to determine if a curriculum is comprehensive. You need a case study, because a randomized trial makes sense only if you know exactly what a program is and are certain that it can be implemented over the duration of the experiment," she says. Analyses are lacking on hundreds of interventions now in use, she adds.

Sclafani says she doesn't disagree with the value of contextual studies. But she says that taxpayers deserve more from their considerable investment in school reform. "NSF has supported exploratory work for a long time. There was an opportunity to collect evidence about their effectiveness, but that opportunity has been lost [because NSF didn't insist on experimental evaluations]."

Judith Ramaley, who recently stepped down as head of NSF's EHR directorate, says she's glad that ED wants to build on NSF's work in fostering innovations in math and science education by testing their performance in the classroom. "The medical model makes sense for them," says Barbara Olds, who directs NSF's evaluation division within EHR. "We think there are many fundamental questions in education that have not been answered."

ED officials are working with states to spread the gospel of experimental evaluations. Under the department's $178 million Math/Science Partnership program—the money from which has flowed directly to the states for the past 2 years—state governments have funded more than three dozen projects with a randomized or quasi-experimental study component. (None has yet yielded results.) And the department plans to do the same thing with the new math initiative.

"Teachers are telling us: 'We know what works in reading; tell us what works in math and science,'" says Sclafani. "We hope to be able to tell them that, if you do a, b, and c, you'll be sure to see results."

–YUDHIJIT BHATTACHARJEE