Educational Evaluation and Policy Analysis
Summer 2003, Vol. 25, No. 2, pp. 119–142

Resources, Instruction, and Research

David K. Cohen
Stephen W. Raudenbush
Deborah Loewenberg Ball
University of Michigan

Many researchers who study the relations between school resources and student achievement have worked from a causal model, which typically is implicit. In this model, some resource or set of resources is the causal variable and student achievement is the outcome. In a few recent, more nuanced versions, resource effects depend on intervening influences on their use. We argue for a model in which the key causal agents are situated in instruction; achievement is their outcome. Conventional resources can enable or constrain the causal agents in instruction, thus moderating their impact on student achievement. Because these causal agents interact in ways that are unlikely to be sorted out by multivariate analysis of naturalistic data, experimental trials of distinctive instructional systems are more likely to offer solid evidence on instructional effects.

Keywords: *experiments, instructional effects, research and policy, school effects*

For most of the history of U.S. public schools, conventional educational resources were seen as the key to making schools work. Educators, parents, and policymakers acted as though they assumed that money, curriculum materials, facilities, and their regulation, caused learning. Many still seem to assume that, as they write about the "effects" of class size or expenditures on learning. The phrasing implies that resources carry "capacity." Regulation has been thought to work by steering resources and thus capacity, within and among educational organizations; the idea is that ability grouping or segregation influence achievement by influencing access to resources. These assumptions made school improvement seem straightforward: allocate more resources or regulate schools' allocation of them.

Access to schooling does affect outcomes. Students learn algebra in classrooms, not on the street. High school students who study in academically more demanding curricula learn more than students in less demanding curricula, even when students' earlier achievement is taken into account. Disadvantaged students' achievement may fall off when they do not attend school in the summer (Alexander, Entwisle, & Olson, 2001). But several decades of research suggest that access itself does not cause learning. Researchers report that schools and teachers with the same resources do different things, with different results for learning.

Earlier versions of this study were published in Boruch, R. & Mosteller, F. (Eds.), *Evidence Matters: Randomized Trials in Education Research.* Brookings Institution, 2002 and in the Center for Teaching Policy's working paper series (http://depts.washington.edu). Seattle: University of Washington. We thank Simona Goldin for extraordinary research assistance, and Jere Brophy, Anthony Bryk, Jeremy Finn, Fred Goffree, Henry Levin, Richard Murnane, Will Oonk, Annemarie Palincsar, Jeremy Roschelle, and Alan Ruby for helpful comments. Grants from The Pew Charitable Trusts and the Carnegie Corporation of New York to Michigan State University and The University of Michigan, from The Office of Educational Research and Improvement (U.S. Department of Education) to The Consortium For Policy Research in Education (CPRE) and The Center For the Study of Teaching and Policy, and The Atlantic Philanthropies, to CPRE, supported various elements of the research. None of these people or agencies are responsible for the ideas in this article.

The differences depend on the use of resources; access creates opportunities for resource use, but resources are only used by those who work in instruction.

Some would say that is obvious, but if it is obvious in principle, it has not been so in practice. Consider the almost exclusive focus on resource provision, in 150 years of education policy; only in the last few decades has there been any attention at all to use. Consider as well research on the effects of resources, in which use also has received little attention. In most cases, the form of analysis seems to assume an unmediated relationship between resources and learning. If the importance of resource use is obvious, it remains to be understood. Hence we reconsider the role of resources in instruction. We discuss research, which illuminates the nature of resource use. We sketch a theoretical view of instruction that focuses on resources and their use, and discuss evidence on class size to illustrate these ideas. We then consider the implications of our ideas for research on the effects of resources.

Relating Resources to Outcomes

Educational resources, conventionally conceived, refer to money or the things that money buys, including books, buildings, libraries, teachers' formal qualifications, and more. There is a great deal of data on such resources, partly because it is required for official reporting, but those requirements were based on the view that the measures of resources were valid measures of educational quality. The underlying assumption was that learning depended on such resources. Yet four decades of research on the effects of resources raised basic questions about that assumption. They began with Project Talent, with *Equality of Educational Opportunity Survey* (Coleman, et al., 1996), and with *Inequality* by a research group led by Christopher Jencks (Jencks, et al., 1972). To nearly everyone's surprise, conventional resources were weakly related to student performance. Differences among school libraries, teacher experience and education, expenditures, science labs, and other resources had weak or no associations with differences among school average student achievement. Despite large differences in average achievement among schools, and especially troubling differences between schools that enrolled the children of affluent and poor parents, differences in the educational resources that most people thought significant were weakly related to differences in student performance among schools. The most powerful predictors of school-to-school differences in average student performance were school-average parents' educational and social backgrounds in contrast to, which resources had trivial effects. Schools with more conventional resources did not have substantially higher performance, once students' social and economic background was taken into account.

This was often taken to mean that schools did not "make a difference," an idea that some conservatives embraced to attack liberal social policy, and that some liberals rejected to defend it. But the research was limited: it asked not whether schools made a difference, but whether some seemed to foster more learning than others, given knowledge of school-average student social and educational background and resources. Researchers found that differences in schools' aggregate achievement were not much related to differences in their aggregate resources.

Most studies since then, prominently including the meta-analyses of Eric Hanushek, supported Coleman and Jencks (Hanushek, 1981, 1989). But some recently revived claims for conventional resources that Coleman and Jencks had reported to be ineffective. Larry Hedges and his colleagues reanalyzed scores of studies using a different approach to meta-analysis than Hanushek, and found that money made a modest difference to student scores (Hedges, Laine, & Greenwald, 1994). The Tennessee class size experiment showed that some students' learning benefited from dramatic class size reductions (Finn & Achilles, 1990; Mosteller, 1995). These reports diverge from the research of Coleman and Jencks, but it has not been clear what accounts for the divergence. Researchers who analyzed the STAR data from Tennessee disagree about why class size made a difference, and no convincing theoretical frame has been offered (Blatchford, Moriarty, Edmonds, & Martin, 1992).

Coleman and Jencks' research was a watershed. Debate about schools previously had focused on resource access and allocation, not results, partly because the latter often were tacitly assumed to be implied in the former. Coleman and Jencks' work called that connection into question, and it soon was difficult to take conventional resources as measures of educational quality, or to assume that adding resources would reliably affect student performance. There was much consternation and accusations, but in the wake of many hard words, several streams of new, more detailed, and often creative work began to illuminate the issues that Coleman and Jencks had broken open.

New Views of Instruction

In response to that shift, and often in deliberate opposition to the work of Coleman and Jencks, several researchers tried to decode educational quality, to discern what made instruction work. One group sought to figure out whether some teaching was more effective and, if so, why. They probed instructional processes and the resources used therein. Though they did not say that they were studying resources, their work offers clues to how resources are related to school outcomes. In one summary of the evidence, more effective teachers were significantly different from that of their peers, at least as judged by students' gains on standardized tests. More effective teachers planned carefully, used appropriate materials, made the goals clear to students, maintained a brisk pace, checked student work regularly, and taught material again if students had trouble. They used class time well and had coherent strategies for instruction. They believed that their students could learn and that they had a large responsibility to help. These teachers deployed resources that helped students to learn, but the qualities that we just summarized were not resources that could be captured well in measures of teachers' formal qualifications, or their schools' expenditures (Cooley & Leinhardt, 1978; Brophy & Good, 1986).

Other researchers brought a similar perspective to studies of schools. They sought to distinguish more and less effective schools, and to identify what caused the difference. To do so they probed connections between schools' collective characteristics and student performance. Faculty in unusually effective schools appeared to share a vision of the purposes of instruction (Edmonds, 1984; Rutter et al., 1979; Rosenholtz, 1985). They agreed that schools' purpose was to promote student learning, that it was their responsibility to help students to learn, and that all students had real capacity to learn. Teachers in such schools had stronger commitment to students' academic success, and their principals helped to create and sustain these beliefs and practices (Edmonds, 1984; 1979). A sophisticated study in this line, which focused especially on Catholic high schools, found that teachers in more effective schools were more likely to have a shared commitment to their students' academic success, to have strong collegial relations, and to believe they were obliged to help students learn (Bryk, Lee, & Holland, 1993). A large study of "restructuring" schools reached similar conclusions (Newmann & Wehlage 1995). An extensive program of research on teachers' academic community in high schools reported strong relationships between teachers' community and sense of collective responsibility for students' work on the one hand and students' academic performance on the other (McLaughlin & Talbert, 2001). These characteristics of schools and departments could be seen as personal and social resources—human and social capital—that were mobilized in some schools but not others.

A third line of inquiry probed teachers' and students' interactions over specific content, and offered finer clues to the role resources play in instruction. Researchers tried to map the domains that lay between such gross influences as the time that teachers and students spent on the one hand, and what students learned on the other. They reported that time alone was not consequential (Cooley & Leinhardt, 1978). Only when the nature of academic tasks was taken into account were effects on learning observed. Teachers' task definition and students' task enactment were the key influences, and students' performance of instructional tasks mediated between teachers' task setting and students' learning. One could see this work as an effort to track the paths by which several resources—curriculum materials and teachers' knowledge chief among them—were used in instructional actions that affected learning (Leinhardt, Zigmond, & Cooley, 1981). Two other researchers identified the practices that distinguished more and less effective readers, and taught them to teachers who in turn taught them to students (Palincsar & Brown, 1984). When effectively taught, the practices improved students' study in reading, and their reading achievement (Englert, et. al., 1991).

A different line of work showed that learners' resources could be as crucial as teachers', by demonstrating that learners' attributions about intelligence and learning play a key role in classroom work and learning (Dweck, 1986, 1988). Children who saw intelligence as fixed tended to avoid intellectual challenges that might publicly reveal wrong answers, but children who thought that intelligence was influenced by effort sought out and used those challenges. Children could be taught to change their attributions; when those who saw intelligence as fixed were taught that it could be influenced by effort, they increased effort and made better use of teachers' feedback. They learned how to study better, in part, by thinking differently about the resources they brought to instruction.

These studies show that some scholars' interest moved from conventional resources, like money, teacher qualifications, and facilities, to particular instructional practices and organizational arrangements, and the actions, knowledge, and culture that they entail. If practice-embedded knowledge and action affect learning, then teachers' and students' knowledge and actions also are resources. These personal resources mediate between the conventional resources that schools and school systems deploy on the one hand, and learning accomplishment on the other. Many researchers treat teaching as though it directly provoked learning, but in the work summarized here, effective teaching encouraged and closely supported what students did in instruction, and students' work helped them to learn, or not. Teaching is portrayed as activities that enable students to use materials, tasks, and other resources more or less well. Much instruction that researchers had associated with individual teachers' work also turned out to have collective features; it was shaped by teachers' work together, by leadership, and by the organizations and cultures in which students and teachers worked.

The effects of resources depend on both access and use: students and teachers cannot use resources they don't have, but the resources they do have are not self-acting. Simply collecting a stock of conventional resources cannot create educational quality, for quality does not arise simply from such attributes. If resource effects depend on their use, then modeling the effects requires a theory of instruction, for that is where most resources are used. Understanding instruction poses a serious challenge to causal inference about the effects of resources, but we begin by sketching some elements of a theory.

Resources and Instruction

Instruction consists of interactions among teachers and students around content, in environments. The interactions occur in distance learning, small groups in classrooms, informal groups, tutorials, and large lectures. "Interaction" refers to no particular form of discourse but to teachers' and students' connected work, extending through, days, weeks, and months. Instruction evolves as tasks develop and lead to others, as students' engagement and understanding waxes and wanes, and organization changes (Lampert, 2001). Instruction is a stream, not an event, and it flows in and draws on environments—including other teachers and students, school leaders, parents, professions, local districts, state agencies, and test and text publishers. This view of instruction has roots in early 19th century ideas, yet many researchers and practitioners still refer to teaching as though it was something done by teachers to learners.

To illustrate, we sketched a hypothetical 2nd-grade mathematics class. The school district recently made mathematics a priority for improvement, and adopted a new text series. Every teacher received a complete set of the materials, and, to support the initiative, the district mathematics coordinator organized ten professional development sessions for teachers. Several elementary school principals devoted meetings to math instruction. A district committee also developed a map, indexing the goals and benchmarks of the new texts to the state tests, to help teachers connect them. Yet the school board is divided on the program and recently cut the math coordinator position from full to three-quarters time, to fund computer support at the middle school.

To continue from the day before, the teacher offers a conventional subtraction problem:

$$\begin{array}{r} 72 \\ -\ 28 \\ \hline \end{array}$$

But her focus is less conventional. She asks students to do more than calculate the answer. She instructs them to copy the problem and to use base-ten blocks to model the numbers and processes precisely, as well as to figure out and justify the answer.[1] She walks around, observes students' work, and sees that some are having trouble using the base-ten blocks. One boy is trying to count out 72 using only small cubes. Another is not using the blocks, but is meticulously drawing 72 hash marks on his paper. And though several have efficiently modeled 72 with seven rods and two cubes, some are also modeling 28 while others are trying to take 28 away from 72.

The teacher opens discussion by asking students for the answer to the problem. She gets four different answers: 46, 100, 56, and 44. Recording each answer on the board, she says, "Let's figure out which one is right." Danya's hand goes up: "I can show it with my blocks. It should be 100." "Nooo," call out several children. "Let's see what Danya does," replies the teacher, knowing that Danya made an error common for students at this level—adding instead of subtracting. But the teacher thinks that working through and reasoning mathematically about the solution in public, and exposing the mix-up, will be useful for Danya's and her classmates' learning.

Danya goes to the overhead and carefully lays out seven rods and two little cubes and then two rods and eight little cubes. "Notice how Danya has represented 72," the teacher points out, stepping in to review an important concept. "She uses as many rods as she can then the rest with cubes. Can someone explain how what she has done with the blocks matches how we *write* 72?" Several children explain: "The seven rods go with the 7, for seven tens, or 70. And the two little cubes go with the 2, for two ones. So it is seventy plus two," asserts Guina.

Danya then pushes all the blocks together: she counts the eight and two little cubes as ten, and trades them for a rod. As she starts to count the ten rods, "Ten, twenty, thirty . . ." she realizes her error. "I was *adding,*" she announces, ruefully. "The answer is not 100." "Good work, Danya," says the teacher, crossing 100 off the list of proposed answers. "Being able to figure out when an answer is not right, and why, is important in mathematics. Would someone else like to try to show which is the right solution?"

"It's 44," Katie announces confidently, and "I can show it." At the overhead, she correctly represents 72 with seven rods and two cubes, and then trades in one of the rods for ten more cubes, resulting in six rods and twelve cubes. She quickly removes two rods and eight cubes. "See? It's 44." Several children nod. "What do you think, Danya?" asks the teacher. Danya nods. "Mmm-hmm. I agree."

"Ruben, can you show with numbers what Katie did with the blocks?" asks the teacher calling on a boy who has been sitting slumped over in his chair. "I didn't do it," he says softly. "My dad told me that only babies use blocks to do math." "Blocks are not for babies," replies the teacher. "Using the blocks shows that you can *explain* what we are doing. Who wants to show it with this problem? Then we will try another."

James comes up to the board and writes the conventional subtraction procedure:

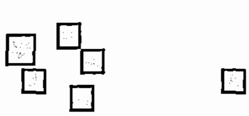
$$\begin{array}{r} {}^{6}\,{}^{12} \\ \not{7}2 \\ -\ \ 28 \\ \hline 44 \end{array}$$

He explains that the 12 is the twelve cubes that Katie had after she traded in one of the rods, and that then she had only six rods left, not seven. Then he explains that he took away eight cubes from the twelve, leaving four, and two rods from the six, leaving four. "So that shows 44, like Katie showed, and this is how we write it."

"Nice job, James," comments the teacher. "So now we see that 56 is not the right answer, and not 46 either. Does anyone know how someone would get one of these answers?" "By forgetting to regroup," calls out Lucy. "By forgetting to cross out the *tens,*" shouts Leon. "Good, okay, so these are mistakes that children make sometimes," says the teacher. "Let's try a couple more." She puts two new problems on the board, strategically selected by the curriculum developers to focus the students on the decision of whether or not a problem requires regrouping:

$$\begin{array}{r} 51 \\ -19 \\ \hline \end{array} \qquad \begin{array}{r} 59 \\ -11 \\ \hline \end{array}$$

"I want you to work alone on this for a few minutes, and this time and this time, I would like to see everyone showing it with the blocks and also in writing, okay?" While the children start copying the problem onto their notepads, she stoops down beside Ruben to help him get started. She knows that his father is upset with the new math program and is worried that this is affecting Ruben's work. She works with him to set up the first number. He sits, immobile. Then he slowly gets five little cubes and one little cube, separate, and lays them out next to each other in two groups:



The teacher asks him what number the blocks show. "Fifty-one," he begins, and then says: "No, I guess this is only *six*." "Terrific. So do you see how we can show 51 with the fewest blocks?" asks the teacher. Ruben pauses, and begins slowly to pull out some of the rods, and counts, "Ten, twenty, thirty, forty, fifty," he murmurs. Pulling over one little cube, he continues, "Fifty-*one*." "That's it, you got it now" says the teacher. "What's next?" "I need to trade in!" he exclaims. The teacher asks Ruben to do it. He carefully takes one of the rods and trades it in for ten little cubes. "Can you record what you did now?" directs his teacher. He crosses off the 5 and writes 4, and crosses off the 1 and writes 11. The teacher tells him to finish the problem.

The teacher sees the principal at the door and beckons her in. She asks how Ruben is doing, and reports that his father complained that his son was not getting enough math skills, that they spend too much time working with blocks and toys and not enough time doing mathematics. The teacher suggests meeting to explain the work. "Perhaps we should meet with more parents, since he is not the only one," the principal says.

In this example, what we casually call teaching is not what teachers do, say, or think, though that is what many researchers have studied and many innovators have tried to change. Teaching is what teachers do, say, and think with learners, concerning content, in particular organizations and other environments, in time. Teaching is a collection of practices, including pedagogy, learning, instructional design, and managing organization.[2] There are more practitioners than teachers, more practices than pedagogy, and the environments of teaching and learning are implicated in the interactions. These ideas are roughly depicted in Figure 1.

Resources are used as teachers design lessons, set tasks, interpret students' work, and manage time and activity. To do so teachers and learners must operate in several domains: they must hold and use knowledge, coordinate instruction, mobilize incentives for performance, and manage environments. The domains are not always distinct in practice, but it is more convenient to treat them separately for analysis.

Knowledge Use

The effects of resources depend partly on knowledge. The best materials are of little use if teachers cannot turn them to advantage in framing tasks, or if students cannot use them to engage the tasks. Ample school budgets can have no constructive effect on learning if they are not used to hire good teachers and enable them to work effectively. Observers would report that such

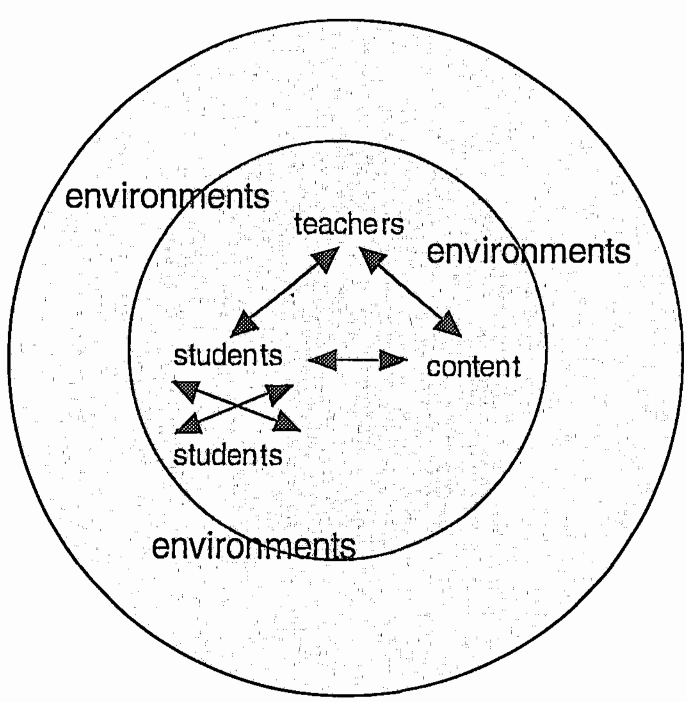


FIGURE 1. *Instruction as interaction.*

schools had rich resources, but that the potential to affect instruction was unrealized (Odden & Busch, 1998).

Knowledge counts in several ways. Teachers who know a subject well, and know how to make it accessible to learners, will be more likely to make good use of a mathematics text, to use it to frame tasks productively and use students' work well, than teachers who don't know the subject, or know it but not how to open it to learners. Our teacher used knowledge of subtraction and modeling to focus students' work on key topics. She knew that having students use the blocks merely to obtain an answer was not sufficient, and so emphasized correspondence among the numbers, their representation in the base-ten blocks, the operation of subtraction, and the symbolic form of the problem. She also made error analysis, a staple of good mathematical work, part of instruction, rather than treating errors as shameful. The teacher attended to Ruben's resistance, and used her knowledge to open the content and help him use the task. She used her mathematical insight to exploit both the arithmetic and the representational dimensions of the problem. Had she seen this as only a matter either of getting answers or of using manipulatives, the lesson would have unfolded differently. Yet there is no way that teachers and students can know enough to always use resources optimally, for that would require omniscience. Nor can those who design materials or other resources know enough to fashion them best for all uses, for that would require perfect foresight.

Similarly, students who have learned to reflect on their ideas, listen carefully, and express themselves clearly are likely to make better use of materials, teachers, and other students' work. They also are likely to make it easier for other students and teachers to use their work. How students and teachers organize their interactions also shapes access to resources: if they work in classrooms that support the respectful expression, explanation, and scrutiny of ideas, they are likely to generate more usable material for instruction, and have more resources to use, than classrooms in which teachers do most of the talking, students work in isolation, and errors are shamed.

This domain of practice includes an extraordinary array: knowledge of academic subjects, practices of learning and teaching in which such knowledge is managed and the organization and management of instruction. In our example, the classroom was structured so that children could discuss their work. The teacher managed in part by listening attentively and helping students to attend to each others' ideas. Students' achievement thus depends in part on how deftly teachers probe and understand their work; the strengths or disadvantages they "bring" are partly a matter of what teachers can see and hear, and how skillfully they respond. In another class, with a teacher who could use mathematics less well, Ruben might have been seen as "unmotivated," offered an incentive to engage, and might not have gotten the help he needed to make progress. What reformers term instructional "capacity" is not a fixed attribute of teachers, students, or materials, but a variable feature of interaction among them.

As things presently stand, teacher quality is less well predicted by formal qualifications than by more direct indicators of teachers' knowledge, which probably also proxies for their ability to make pedagogically fruitful use of materials and students' work (Ferguson, 1991). The current weak validity of formal qualifications is an artifact of existing professional education, in which intending teachers are not well educated in conformance with sound standards of academic performance. Instead they are sketchily educated in conformance with very general standards that are weakly related to teaching performance and academic learning. If teachers were better educated in conformance with academic performance standards, formal certification would be more tightly related to performance, and qualifications would be better proxies for teaching proficiency.

Coordinate Instruction

The use of resources also depends on coordination in instruction. One dimension of coordination concerns teachers' and students' work on content. The teacher in our example worked on subtraction problems, but if students were fiddling with their pencils, passing notes, or drawing, they would be less likely to learn from the tasks she posed. If several students were absent the day before, they would be less likely to know what was being discussed, and to learn from the problems. Even if everyone worked on the problems, and knew what to do, the teacher might addressed them algorithmically while the curriculum supported the development of mathematical ideas. If she addressed the work as the text did but not probed students' ideas, she would not have known how they understood it. Instruction occurs in time, which opens up another dimension of coordination. How do these subtraction problems connect with each student's work tomorrow, and what is to happen in two weeks? Learning depends on students and teachers making bits of lesson work develop and connect, yet there are always absences, memory lapses, and inattention to contend with (Lampert, 2001).

Since instruction consists of more or less complex interactions among teachers, learners, and content, there are many opportunities for uncoordination. Other dimensions of coordination concern pacing across time, relations among classrooms within grades, among successive grades, and between work in schools and external guidance for instruction, as from academic standards and assessments. Coordinating instruction thus depends on making connections among teachers' and students' ideas, among students' ideas, among both over time, and between both and elements in the environment. These things depend on teachers' knowledge of content, on how it is represented, on learners' understanding, on agents in the environment, and on the will to make fruitful connections. Coordinating instruction in these senses also depends on social resources that build trust, support the collection and analysis of evidence about teaching and learning, and enable communication about the evidence.

Each dimension presents potential sub-problems. If students and teachers do not focus on the same task, learning is likely to suffer. If students' work is not paced to maintain cognitive demand, students may be overwhelmed or bored, and learning will suffer. If steps are not taken to coordinate in these and other ways—including how work is organized in periods, days, and years, and student mobility within and among schools—instruction is likely to be less effective. If our teacher merely assigned problems and collected and graded papers, as has been typical, the lesson would have been far different. Instead she coordinated in many ways, including her trouble-shooting work with Ruben, her use of base-ten blocks, and increasing difficulty from the first problem to the second two problems. Using resources effectively depends on such coordination.

Mobilize Incentives

It takes effort to teach and learn, and that often creates friction within learners and among them. Incentives are required to mobilize effort to overcome that friction, and creating such incentives is a third instructional domain. Teachers have incentives to exert themselves and press for ambitious work, for their professional success depends on learners' success—or on explaining why learners could not succeed. Learners also have reasons to work hard, for it can satisfy their curiosity and wish to learn, enhance their sense of competence, and enable them to meet teachers' and parents' hopes. But teachers and learners also have incentives to do less ambitious work, for friction and effort increase as learners encounter more difficulty, as do uncertainty, risk of failure, and chances to disappoint themselves and others. The subtraction problems our teacher used are of this latter sort, for the representation and explanation can be challenging for second graders. Teachers who frame such work are more likely to encounter learners' resistance, frustration, and failure, even if greater success beckons. Learners and teachers who do less ambitious work reduce these problems and increase the chance of some success. Teachers and students face a dilemma that is stitched into the work: should they aim low, accepting modest results in return for some success, or aim high, risking resistance and failure in hope of more impressive accomplishments for learners and teachers? To teach and learn is to manage that conflict.

In our example, the teacher posed a challenging task when she asked her students to represent a conventional subtraction problem using concrete materials, and to make a careful correspondence between their work with the blocks and the steps of the procedure. Her introduction of the task with a review problem allowed students to get started and build confidence and to be clear about what she wanted them to do. Her invitation to resolve discrepancies in the proposed answers, using the blocks and explanation, and to do so as a performance in front of the class, focused the students on precision and meaning, and engaged them in a challenge. Mobilizing incentives to learn and teach is not simply a matter of "motivating" students and teachers, but also of using knowledge and skill to situate incentives to work hard in specific academic tasks, and using performance in the tasks to motivate engaged work. A basketball coach who was keen for his players to win, and good at cheering them on, but knew little about offensive playmaking, would be as unlikely to produce winning results as a history teacher who was keen for her students to learn, but knew little about the historical record, or how to set suitably engaging and challenging academic tasks.

Manage Environments

Instruction is situated in what often are depicted as external influences, including other teachers, school leaders, parents, district policies, state requirements, and more. But if these things do in some sense exist outside instruction, they also appear within it. A fourth domain of instruction is managing such elements of the environment. When teachers and students deal with problems of coordination, resource use, and incentives, they do so in and with environments. Teachers and students are more likely to exert themselves if schools are linked to institutions of higher education or firms that offer strong incentives for ambitious performance, for students and teachers import the incentives. Teachers and students whose principals urge ambitious work will be more likely to do it, while equally able colleagues in schools whose principals prefer less ambitious performance will be less likely. Again, those working inside instruction import external guidance (Bishop & Mane, 1998; Bishop, 1998, Skrla & Scheurich, 2001). Teachers and students work with the classroom manifestations of such influences; though they have little leverage at their source, they can notice or ignore them, capitalize on them or leave them unused.

Coordination also is less difficult in environments that offer coherent guidance for instruction. As the teacher in our example worked on mathematical concepts and skills, she also dealt with parents' views of the new math curriculum. She learned that Ruben and perhaps others were doing poorly in part because their parents disparaged the work. She also knew of many signals about instruction, including upper-grade teachers' expectations, her principal's exhortations to make sure all students develop basic skills, the district's investment in the new curriculum that focused on concepts, and state tests that rewarded speed and accuracy. The United States has had distinctively incoherent guidance for instruction, which makes it more difficult to coordinate within instruction. Standards-based reform has sought to order the confusion, but it has not reduced the proliferation of guidance in many states, and may have increased it, as new guidance that calls for coherence has been laid on many earlier layers of less coherent guidance.

Teachers and students shape environments by what they notice and how they respond, but environments shape attention and response. If school leaders place a high priority on improving disadvantaged students' work, teachers are more likely to engage that task. If leaders go further, by offering teachers opportunities to learn how to improve, it is more likely that teachers will constructively deal with student disadvantage. How educators manage environments is influenced both by the clarity and authority of priorities, and by teachers' and learners' attention, will, and knowledge. The more knowledgeable and skillful teachers are, the more likely they will make productive use of signals from the environment, but the more inchoate the environments, the more difficult it is for even the best teachers to make such use of them.

Many researchers treat environments and practice as separate; they view teachers and learners as technical workers inside practice, and environments as outside influences. Researchers and educators often portray economic and social differences among families as external causes of differences in student performance, yet those differences only count as they become active inside instruction, as learners import elements of the environment and teachers interpret them. Students and teachers are delegates from environments beyond the technical and professional world, yet they are the key agents in that world. The designers and publishers of materials also frame content to manage the environments of instruction, as when texts intended for sale in Southern states fail to mention evolution. Teaching and learning are not simply internal technical work that external environments influence, for teachers and learners work, inside instruction, with and on elements of what is conventionally thought to lie beyond practice.

Resources Reconsidered

Our analysis distinguished among types of resources, and offered a view of causality. Conventional resources include teachers' formal qualifications, books, facilities, class size, and time. Personal resources include practitioners' will, skill, and knowledge. Environmental and social resources include state guidance for instruction, academic norms, professional leadership, and family support. Each type counts. Students in classes of 35 probably have less access to teachers' time and expertise than those in classes of 15. Students with outmoded texts probably have access to less substantial content than those with up-to-date books. Students in less developed nations, with uneducated teachers and few books have access to fewer resources than those in industrialized nations with better-educated teachers and more books.

Yet conventional resources only count as they enter instruction, and that happens only as they are noticed and used. Some sorts of resources seem more immediately usable than others: books and other materials are close to the center of instruction, but class size is not as close to the center, nor are school facilities.[3] Their effects all depend on teachers' and students' personal resources: their knowledge, skill, and will, but some sorts of resources are easier to use than others. Students often learn from materials on their own, with no mediation by teachers. Environmental and social resources like leadership are most likely to influence learning by influencing what more immediate resources teachers and students notice and use. Leaders can support or deflect common academic priorities: the principal in our example engaged the school with mathematics by making it a focus of conversation and attending to parents' views. If teachers' and students' use of resources is central to instruction, the chief means by which actors in their environment can influence use is by focusing their attention and improving their capabilities as users.

If we are roughly correct, it is quite unlikely that naturalistic research could, by itself, yield valid inferences about the use of conventional resources. The pathways are too intricate, and the ways in which teachers and learners capitalize on the available resources, or compensate for their lack, are extremely complex. We argue, a bit further on, that the interactive nature of classroom life is likely to obscure causal relationships among resources, users, and outcomes. At the same time, studies of how conventional resources are used can lead to important *hypotheses* about the effects those resources might have under varied approaches to instruction. We discuss the evidence on class size in the next section, to illustrate this point. Such hypotheses would be an excellent contribution, but only if they could be tested systematically. We consider below how such hypotheses could be generated and tested concerning resources and instruction.

Class Size

The evidence on class size reduction (CSR) offers an opportunity to apply and develop the ideas sketched above. It is a distinctive type of resource, not as close to the center of instruction as materials, but not as remote as many environmental influences. Hence it is more usable than the latter sort of resource, and less usable than the former sort. Research on class size has accumulated for decades, including small experiments, studies of naturally occurring variation, and two recent large studies: the experiment in Tennessee (STAR), and the evaluation of the initial years of Wisconsin's CSR program (SAGE). Despite differences in their designs, STAR and SAGE had comparable effects, which are consistent with those that Gene Glass and his colleagues published in their earlier meta-analyses (Glass et al., 1982). Finn and Achilles (1999) summarized the STAR results: "On average, students in small classes evidenced superior academic performance to those in other conditions . . . The effects were always attributable to the difference between the average performance of small classes and that of other class types . . . The benefits were substantially greater for minority students or students attending inner-city schools in each year of the study . . . [and] was also statistically significant for all school subjects in every subsequent year . . ." (Finn & Achilles, 1999, pp. 98-99).

These findings are relatively uncontested, but differences ". . . arise over the implications . . . One interpretation, shared by Achilles, Bain, Finn, Mosteller, and others, is that the STAR findings confirm the intuition of most teachers: Children perform significantly better in classes with fewer students . . . class size reduction is expensive but . . . less expensive than . . . allowing students to fall behind in school." Others question the size but not the direction of the effects, and hold ". . . that the costs of the class size reductions far outweigh the small achievement gains" (Ritter & Boruch, 1999).[4]

Our interest is in the dynamics of class size effects; we want to use the research to illuminate resource use, and that requires comparisons among types of use. Yet most of the research focused on average effects, not on the distribution of effects or sources of variation in that distribution. A few studies report on the distribution of effects, reporting that their size and direction varies. Alan Krueger reported that two-thirds of the ". . . small-class effects are positive, while one-third are negative.... Thus, some schools are more adept at translating smaller classes into student achievement..." (Krueger, 1999, p. 526), Eric Hanushek found fewer positive effects: in 79 STAR schools which had classes for each experimental condition (small, regular, regular with aide), small classes outperformed regular classes in 40 schools. In the remaining 39, students in small classes made gains that were equal to or less than those of students in regular sized classes, hence there was not equally effective use of smaller classes.[5]

What distinguished classrooms in which there were positive effects from those in which there were not? The answers could illuminate resource use, and might suggest ways to enhance the effects of class size reduction or other resources. When Gene Glass and his colleagues undertook their studies, they found substantial positive effects on students' academic performance, and discussed possible causal mechanisms. They wrote that: "Class size has no magical, unmediated effect on student achievement. Instead, it influences what the teacher does, his or her manner with the students, and what the students themselves do or are allowed to do. These differences in classroom process in turn influence outcome measures like student achievement, student attitudes, and teacher morale. It is essential to study and understand this full sequence of events. *A class size reduction provides an opportunity for improvements in classroom processes. Teachers can take advantage of this opportunity in different ways and to different degrees.* [emphasis added]" (Glass et al., 1982, p. 67)[6]

Several researchers investigated how teachers or students ".... take advantage of this opportunity," with some useful results. Yet none seemed to frame the inquiries with theories that pointed to likely explanations, and several seem to have expected spontaneous change. Hence few researchers were in a position to test hypotheses about how this resource was used. But two groups did use STAR to investigate the dynamics of teachers' and students' response. Finn and Achilles did a Grade 4 follow-up of STAR classrooms, collecting test scores and behavioral data,[7] and concluded that students' behavior changed much more than teachers'. "... The key to the benefits of small classes is increased student engagement in learning ... every student is on the firing line. It is difficult or impossible to withdraw from teaching-learning interactions ..." (Finn and Achilles, 1999, p. 103.)

The importance of group size seems undeniable, but if it is "... difficult or impossible" for students to hide in smaller classes, how to explain the small classes which did not outperform large ones? The answer seems to turn on students' and/or teachers' use of the resource, either because students who did not gain as much from small classes were more advantaged and thus less sensitive to the resource change, or because teachers made it more difficult for some students to improve their use of this resource, or both.[8] In either event, if differential use seems critical, published studies offer no evidence with which to pursue the point.

Were there evidence, our theory suggests at least three hypotheses about students' and teachers' resource use. Students could make better use of themselves and materials with CSR, capitalizing on fewer distractions to attend to their own work more effectively, or to spend more time on it, or both. Or they could make better use of teachers, using the teachers' greater availability to gain more attention than in large classes. Or teachers could press students to do either or both of these things. In the former case, students would use the changed social situation to improve their opportunities to learn, and they would be the key causal agent. In the second, the effects would depend on students trying to use their teachers, and on teachers responding constructively, so causality would be joint. In the third, teachers' initiative would drive either or both of the first two mechanisms.[9] Either the second or the third entail some change in teaching, whether in the allocation of time, pressure brought to bear on students, or both.[10] The mechanisms are not mutually exclusive, but different. Stating the hypotheses, which we would have been less likely to do without our theory, also shows that if class size reduction changes opportunities to teach and learn, they are not irresistible; students and teachers must use the opportunities, and that requires will and knowledge.

Carolyn Evertson and John Folger also used STAR to probe the dynamics of use, and their report could bear out all three hypotheses. They studied math and reading lessons in 52 2nd-grade classrooms, and wrote that "... in mathematics, students in small classes initiated more contacts with the teacher for purposes of clarification, giving answers to questions that were open to the whole class and contacting the teacher privately for help [hypothesis 2]. In reading, more students were on task, fewer students were off task, and students spent less time waiting for the next assignment" [hypotheses 1 and perhaps 3,]. They also wrote that "... there are predictable differences in class processes that follow simply from the numbers: students are more visible; each student is more likely to get a turn more often during class lessons; students don't have to wait as long for help; student initiate more contacts with teachers." (Evertson & Folger, 1989, pp. 9–10).

These actuarial differences are clear in the numbers, but nothing follows "simply." As Glass et al wrote, they are opportunities; CSR became effective only as it was used by teachers and students. In another report on STAR, Evertson and Randoph tried to probe change in teaching in response to CSR. To their surprise, it changed little or not at all: "... instructional formats observed in these [2nd-and 3rd-grade STAR] classrooms showed little variation ... tried-and-true methods of reading and mathematics instruction [prevailed]." (Evertson & Randoph, 1990, p. 98.) But "change in teaching" might mean several things. It could mean that students' or teachers' the allocation of time changed; it could mean that there were more student initiated requests for help that teachers responded to; or it could mean that there was more teacher pressure on some students to work; or it could mean some combination of these. None of these changes entailed new pedagogy or content. The study suggests that higher scores were due to the use of added time and social space. If so, teachers and students turned to instructional mechanisms that were easy to use, not to unfamiliar practices that would have been more difficult.

Finn and Achilles also concluded, in an analysis of 4th-grade classrooms, that "... teachers do not, *de facto,* alter their primary teaching strategies. Small classes are academically superior not because they encourage new approaches to instruction, but because teachers can engage in more (perhaps *enough*) of the basic strategies they have been using all along. More profound changes occur in students' participation in learning, including students who would be unwilling to participate if they were part of a larger class ... When class sizes are reduced, the pressure is increased for each student to participate in learning, and every student becomes more salient to the teacher. As a result, there is more instructional contact, and student learning behaviors are improved." Poor and minority students benefit most because "... disengagement is found more commonly among minority or low-income students ..." (Finn & Achilles, 1998, p. 103)[11]

This feature of the response to STAR was also found in Wisconsin's SAGE program. This non-experimental program began in 1996–97, and grew rapidly. Researchers evaluated the initial two years, comparing the performance of students in classes of 15 or fewer and a single teacher and in classes of 30 students and two teachers, with that of comparison students in classes of 30 students and one teacher. Researchers reported test score gains of approximately the same size as those in STAR, in both of the first two groups when compared with the third (Molnar, et. al., 1999). They also reported dynamics that parallel those in STAR. SAGE teachers in both treatment conditions reported that they spent less time on discipline and class management, more time with individual students, more time covering the same content as in previous years, but little other change in instruction (Molnar, et al., 1999; Zahoric, 1999). Apart from "... slight increase in hands-on activities ... the dominant mode of teaching remains direct instruction. Teachers continue to structure, manage, and pace all activities. The teacher gives information, asks questions, praises correct responses, and controls interactions with students in other ways. The students are largely passive in that their role is to listen and to follow the teacher's directions." (Zahoric, 1999, p. 52.)

On this view, CSR offers students and teachers more access to each other. On average, both take it, and scores improve. Yet this is not the end of the explanatory line, for time and attention are not the only influences on resource use. Our theory holds that curricula and environments may influence instruction, and Evertson and Randolph note that STAR occurred in a state with a vigorous basic skills program. Tennessee Basic Skills First (TBSF) included a system of tests that focused on the TBSF curriculum: "The state designed objectives and an assessment system for the basic skills curriculum (language arts and math). Local school districts are required to either use the state's system or design their own which meets state guidelines ... Skills are measured by frequent objective tests. The demands of the program ... encourage adherence and few deviations ... While class size was manipulated in this study, outcome measures were not. Student achievement was still measured on standardized tests tightly tied to this basic-skills oriented approach to reading and math." (Evertson and Randolph, 1990, p. 101.)[12]

This suggests another hypothesis about resource use in STAR: pervasive and potent state pressure for basic skills inhibited change in content and pedagogy. Content was prescribed, and there were incentives to conform. Evertson and Randolph wrote that the basic skills system permeated classrooms. "Where this type of recall-oriented performance is closely linked to what will be tested, there may be no need or encouragement for teachers to change . . . to more complex or multitask settings." (Evertson & Randolph, 1990, p. 101). It would be easy to conclude that TBSF the reason that pedagogy changed little in STAR. Evertson and Randolph used it to explain why professional education for teaching smaller classes, which they offered teachers, had no discernable effect. Yet, if TBSF probably influenced teaching, giving so much weight to that environmental influence would be justified only if there were other studies that reported substantial change in teaching in response to class size reduction, absent programs like TBSF, and several other studies reveal little or no change in pedagogical approaches, even when programs like TBSF did not operate (Betts, et al., 1999; Rice, 1999). No studies report substantial change in pedagogy owing to class size reduction, so explanations for stability and change must take account of more than TBSF.

Our frame leads us to several hypotheses about why teaching might change slowly or not at all with class size reduction. Each derives from our analysis of the domains in which teachers must work. One is that many teachers lacked the skill and knowledge to make more complex changes, and another is that they lacked the incentives. We expect that both played a role in teachers' and students' use of the added resource. Teachers must weigh incentives to push students and themselves to produce more learning and thus more professional success, against disincentives for such push arising from greater effort, difficulty, problems, failure, and student resistance.[13] Substantial class size reduction offers most teachers and many students ways to manage this problem at a rather low cost: teachers can increase the likelihood of success for students and themselves with little or no increase in instructional effort, simply by allocating the same amount of instruction, for the same content, to many fewer students. Smaller classes also make it easier for teachers to coordinate instruction, and to use knowledge they already know how to use, to better advantage; if so, they would be likely to actually reduce the effort teachers had to make to achieve satisfying results. Students would gain an increased probability of success, satisfying their own wishes to learn, their teachers' and parents' desires, or both, simply by taking more advantage of their teachers' availability, reduced distraction, or both.

These ideas fit the published evidence. It would be unwise to conclude that teachers and students in STAR were only responding to changes in group size and time, or that the changes were either "simple" or nearly unavoidable. Even if there had been no pressure for basic skills, we would expect no different distribution of teachers' instructional responses. Modest change in that distribution would occur if there were strong environmental pressures for different content and pedagogy, and more change would occur if, in addition, teachers had substantial help in learning how to use the CSR resource differently, in ways that made it possible for them to solve problems of incentives, coordination, and resource use. If reduced class size prompts change in instruction, it will do so within the parameters suggested by our account of work in the four central domains of teaching and learning.[14]

Class size reduction changes the resources available, but the effects on teaching and learning depend on how teachers, students, and others in their vicinity use the resource. We expect classes in which no improvement occurred to have teachers or students who either saw no promise of improvement in the reduction, or were unwilling or unable to take advantage of it. That would hold for other resources that teachers bring to instruction, including their content knowledge. Consider the finding that teachers with higher test scores, or who know more about a subject, have students with higher scores in that subject (Coleman, et al., 1966; Jencks, et al., 1972; Ferguson, 1991). If we view teachers as people who stand and deliver, those who know more have more to deliver. But the very words are deceptive, because knowledge is not self-enacting: if students benefit from teachers' content knowledge, it is either because students are able to use the knowledge without benefit of teachers' instructional skill, or because teachers with more content knowledge can, on average, put that knowledge to better use in teaching. The two sorts of knowledge differ. Many teachers who know mathematics can barely use it to teach, as many university students know. Knowledge of a subject is a necessary condition of such use, for teachers cannot use knowledge they do not have, but it is not sufficient.

More generally, any mix of personal and environmental resources opens some possibilities for and contains some limits on the use of any conventional resource. Given any particular set of personal and environmental resources, research might show that added conventional resources appeared to independently affect learners' accomplishments, other things being equal. But that apparent independent effect actually would express an interaction among personal, environmental, and conventional resources. The instructional effects of conventional resources depend on their usability, their use by the agents of instruction, and the environments in which they work. When added conventional resources appear to directly affect learning, it is because they are usable, because teachers and students know how to use them, and because environments enabled or did not impede their use. The effects of conventional resource increments should be taken to imply these relationships. Moreover, these effects express an average over many differences in use. The potential effect of something like the Tennessee class size experiment could be greater than the average, perhaps much greater, if teachers and students who made weak or modest use of it were taught to do better. Experiments, which estimate only the average effects of resources and present no evidence on their use, severely limit what can be learned.

If these ideas are correct, then when added resources lie outside the range of teachers' and students' knowledge, norms, and incentives, they will have no discernable effect.[15] A hypothetical legislature might mandate that teachers use innovative content standards to engage learners in more creative and demanding work. The legislature might even provide money to write and disseminate the standards and support discussion of them. Yet research on the effects of such a policy would probably show that the new resources had no average positive effects on students' learning, for the policy would have required most teachers and students to work well beyond their skills, knowledge, and will, without providing opportunities and incentives for them to learn much more.

When research fails to find effects for particular conventional resources, it should not be seen as evidence that such resources are ineffective until several other explanations are ruled out. One is that teachers and learners did not know how to use the resource; ruling that out would require research on the effects of teaching them to use it. Another possibility is that the change was not enough to enable significantly better use, given extant practices, knowledge, and norms. The evidence on class size suggests that only large reductions permit changes in teaching and learning, given teachers' and students' extant knowledge and motivation.[16] A third possibility is that some conventional resources are not salient to the learning in question. Science laboratories might bear on science learning, but it would be surprising if they were linked to reading. Building two gymnasiums, three pairs of bathrooms, and a larger playground for every school could have many good effects, but they seem unlikely to be academic effects.

Instructional Interaction and Research

Our theoretical frame makes interaction between teachers and students over content central to instruction, and portrays teachers and students as interdependent actors: teachers' effectiveness depends partly on how well they can use students' ideas and initiatives, and students' effectiveness depends partly on how well they can use the tasks their teachers set, the comments their teachers make, etc... How teachers and students use resources like class size depends on their work together. None of this is news to those who observe instruction, but the interdependence of which we write poses a major challenge for research on resource effects. Teachers and learners are thinking beings, and they use each other and other resources based on judgments about which resources to use, how, with whom, and to what end. They base these judgments on what they know and believe about themselves, one another, and the content. Some teachers judge with great care and seek evidence with which they might revise, but others judge quickly and with little care. In either event, teachers calibrate instruction to their view of students' capabilities, and their own capabilities to teach. Schools formalize such calibration in ability groups, grade retention or promotion, and related practices, which allocate resources within classes and schools, and even, within the same student, across subjects. Students also make judgments about instruction and calibrate their use of resources to their estimates of teachers' and parents' expectations, and their own preferences. In our example, Ruben navigated between his father's disdain for the math program and his teacher's intervention. Ruben was less prepared for class, and less interested, since his father told him that the work was not worth doing. Ruben's teacher had to believe he could do the work, convince him that it was worth doing, and help him to use what he knew to do it.

One cannot imagine instruction without such calibration—even computer-based instruction includes it. But if teachers adjust instruction within and among students on the basis of their estimates of students' capabilities, there will be significant differences in the resources that teachers use, or the ways they use them, with individual students among subjects, among students within classes, and among classes within schools. If so, how can researchers identify, observe, and measure the resources that are used in instruction? If teachers adjust the tasks they assign and the materials they use, correct estimates of resource allocation and effects would depend on valid evidence of use. That would depend on teachers knowing and articulating what they did, and having the time and inclination to do so, or on researchers' valid observation of teachers' reports and practices, or both. Such evidence would not be easy to define or collect, especially since teachers often adjust their own knowledge, skill, and will as they apply them. How can teachers be aware of such things? If they are, how can researchers learn about them? Lacking valid and reliable evidence on such matters, how could we make valid inferences about the effects of resources on learning? Nonexperimental studies of resource effects on student outcomes that fail to take account of how teachers adjust instruction in light of their judgments about students will likely misestimate and confound the resources used, those merely present, and their effects.

Recent research on "dynamic treatment regimes" in medicine and psychotherapy illuminates this matter (Robins, Greenland, & Hu, 1999). In such regimes, the treatment is calibrated to the current status of the patient, on the basis of an assessment of the patient's condition. The regime consists of one set of rules for assessing those to be treated and another set for assigning interventions to them. One can arrive at strong causal inferences about the effects of any such regime, if those treated are randomly assigned to alternative regimes. Weaker causal inferences can be based on quasi-experimental comparisons of regimes. However, within a given treatment regime, which is where nonexperimental research often operates, it appears impossible to make a meaningful causal inference about the association between treatment and response. This is because in the continuing adjustment process, treatments are as much a consequence of the patient's current condition as the cause of a subsequent condition.[17] Research in medicine and psychotherapy shows that a regression of responses on treatments, controlling initial status, will not give a reasonable estimate of a treatment effect.[18] It also suggests that the effects of interactive treatment regimes can only be accurately evaluated if: (a) there are different regimes that (b) consist of well-explicated rules for assigning treatments, given particular statuses, and (c) the regimes vary across patients treated.

Education is not medicine, and few educational interventions come close to the precision of many in medicine. But if teachers calibrate instruction to their views of student ability, one could make accurate causal inference about instructional effects only by reconceiving and redesigning instruction as a regime, or system, and comparing it with different systems. "Regime" refers not to authoritarian prescriptions, but to systematic approaches to instruction, in which the desired outcomes are specified and observed, and in which the intended outcomes are rationally related to consistent methods of producing those outcomes. The key feature of such systems of instruction is not the detail of specification but consistency between instructional ends and means, and across instruction among learners and among teachers.

Conventional resources are not a system of instruction, for they cause nothing. They are used or not used in particular systems of instruction. Resource effects depend both on their availability and on their use within those systems. The central focus in research on resources therefore should be the instruction in which resources are used—and how they are used, and to what effect—not the resources alone. One key reason for this is that resources and their effects are likely to vary among the instructional systems in which they are used. A text that focused almost entirely on phonemic awareness in an instructional system that was designed for whole-language instruction probably would be used differently, and have different effects, than it would in an instructional system that was designed for phonemic awareness.

This line of reasoning has both theoretical appeal and several vexing aspects. The continuing adjustment of resources within instruction calls into question the interpretation of a vast body of correlational research on relations between discrete instructional behavior and student outcomes, including many of the studies that we discussed earlier (Brophy, 1988). If research on dynamic treatment regimes may not require randomized experiments, randomization is optimal for causal inference. That suggests more caution about causal inference from nonexperimental evidence, a narrower role for survey research than has recently been the case in education, and a larger role for experimental and quasi-experimental research. But if such studies offer a better grip on causality, they are more difficult to design, instrument, and carry out, and more costly.

New Designs for Research

We have discussed two significantly different perspectives on resources. In the inherited, dominant perspective, conventional resources are treated as if they were active agents of instruction, and the key problem is to identify and then deploy the resource mix most likely to improve learning. In a more recent and still developing perspective, teachers and students, and features of their environments are the active agents in instruction, and the key problem is to identify and mobilize the knowledge, practices, and incentives that will enable them to best use themselves and other resources. One perspective is grounded in established habits of thought and politics, while the other is grounded in studies that probe how schooling works.

There has been movement between these views. In the last decade or two some policymakers began to revise views of resources, partly in response to research on instruction. Standards-based reform is premised on the view that schools' learning goals should be clarified, and resources used to achieve those goals. Standards and accountability are seen as important because they would influence resource use. Officials in some districts and states have encouraged schools to focus on improving the use of resources (Odden & Busch, 1998). There has been growing interest in more direct measures of teaching quality and improving teachers' knowledge through professional development, rather than relying on course titles and degrees. There is growing attention to resource use, and the conditions that influence it.

Much action and debate nonetheless focus on class size, teachers' qualifications, facilities and equipment, and budgets, for these have long been the indicators of school quality. They are plainly visible, and it is easier to observe and quantify dollars and class size than teachers' knowledge of mathematics or their skill in using students' work. It is easier to manipulate dollars and class size than knowledge and skill, and they are much easier to associate with the taxes citizens pay. Policymakers' and school managers' actions often are contested, and require justification to tax payers and voters that generate demands for evidence politicians turned to research for help. Specialists in education, economics, politics and sociology have increasingly occupied themselves with the effects of schooling, and they attend chiefly to the resources that play a part in policy and argument. Data are relatively easy to come by, have face validity, and interest policymakers. Policymakers and researchers can most easily deal with the resources that are least directly related to students' learning, while policymakers and managers can least easily deal with the resources that are most directly related to learning. Improved research could help to bring the two closer together.

The Frame

The overarching research question cannot be "Do resources matter"? No deliberate attempt to learn or teach is conceivable in the absence of conventional resources, and there is ample evidence that teaching is causally related to learning. The overarching question must be: "What resources matter, how, and under what circumstances?"

One key circumstance is the desired result. The question can only be answered once an educational goal, and a strategy to achieve it, have been adopted and spelled out. Thus a better question is: "What do educators need to do a particular job?" Putting it that way helps to make clear that the answer would depend strongly on what "the job" was, that is, what is to be taught and learned. Conventional resources do not follow from defining an instructional goal, for the instruction required to achieve any given goal might be done with somewhat less skilled teachers, somewhat larger classes, or a somewhat smaller budget for materials and equipment.

Hence the first question should be: "What instructional approach, aimed at what instructional goals, is sufficient to insure that students achieve those goals?" A second question follows: "What resources are required to implement this instructional approach"? The research question should not be the one that most researchers concerned with school effects have asked, namely, "How do the available resources affect learning?" Since resources can enable and constrain particular instructional aims and methods, the second question often would be followed by a third: "Is it possible to achieve the same or similar results with a different mix of resources?" Educational policy and practice inevitably involve negotiation among goals, instructional means, and resources, and research should weigh the consequences of varies resource constraints within instructional approaches. It is, however, illogical to conceive of resources as the "cause" and learning as the outcome. Systems of instruction are the cause, and resources are facilitators or inhibitors of teaching and learning.

Most researchers have placed conventional resources at the center of inquiry, and tried to identify how each affects performance, or what the best mix is. We propose instead to place teaching and learning at the center of inquiry, and to design research that helps to identify the resources that best support particular goals. It may not seem novel to write that the effects of conventional resources depend on how they are used, but the change would be a revolution of sorts, for it assumes that resources are means, and can only work in relation to instructional ends. To accept that is to bring a kind of theory of relativity to the study of resource effects, for one can only conceive the effect of resources in relation to a specified aim and a strategy to achieve it. Building a new lab may be essential to one approach to science instruction but irrelevant to another. Class size probably is salient to literacy instruction if it entails frequent, high-quality feedback on student writing and serious class discussion of the writing, but that approach also requires literate, motivated teachers. Class size might be less important to other educational aims. Research on resources would be more fruitful if it was grounded in conjectures about such relationships and evidence on the conjectures.

Active and Passive Research Programs

Programs of research on instructional resources should focus on well defined systems. One example might be a program carefully designed to improve reading in the primary grades, which links curriculum and teaching of phonemic awareness, text recognition, and comprehension, to specific assessments in those areas. Such systems of instruction would have several critical features. One is outcome measures that would require students to present the academic performances that the instruction is designed to help them learn. Another is the optimal features of the treatment that is intended to produce the outcomes, including more or less elaborated versions of the academic tasks that were central to the regime, and optimal versions of the instructional media needed to enact the tasks. A third feature would be optimal descriptions of the teaching that is intended to help students use the tasks and materials to produce the desired performances, including descriptions of how teachers would be expected to deal with students' responses to the tasks.

Such systems would require much more consistency in instruction than has been common in the United States. For without such consistency it would be impossible, within a system, either to validly estimate its effects or to systematically vary some resource constraints while holding other elements constant. There are very different ways to achieve consistency: instruction could be relatively tightly scripted at one extreme, while at another, communities of practice could be built around agreed-upon elements of instruction, using intensive communication among teachers about examples of students' and teachers' work to develop and learn shared criteria of quality and methods of instruction. In the former case, consistency would be created by teachers closely following detailed directions, while in the latter it would be created by developing professional knowledge and norms around a skeleton of objectives and tasks. Instruction within regimes could be consistent in either case, but the means to achieve consistency, and quite likely the content of instruction itself, would vary. Combinations of the two methods and others also could achieve the required consistency.

Though any such system would contain articulate rules that regulated or characterized instruction, there could be enormous variability in the range of instructional behavior that are governed by such rules. Teaching and learning school subjects are ill-structured domains, and even in the most constrained regimes, rules could not cover anything like the entire range of instruction. A great deal must be left to teachers and students to deal with on the spot, and, in devising regimes, those who would change instruction would have to decide on the features of instruction to which they would attend, and those that they would ignore. Some systems might focus on a very constrained domain, like word recognition or multiplication, while others might focus on broader domains like reading comprehension or place value.

Research on these systems would address two rather different sorts of questions. A first line of work should probe the effects they have for students on its central outcomes, when resources are plentiful. A second line could test the effects of such regimes under various resource constraints, which also could allow various modifications of the regime that enable its enactment under different conditions.[19] Pursuing the two lines of work for any regime would yield evidence about its effects under a variety of resource conditions, including those that might be optimal. Pursuing both lines of research for regimes that share outcomes, wholly or in part, would yield evidence about their robustness, generalizability, and cost effectiveness. As each was tested and modified, the research program would reveal the resources needed, as well as the ways in which they must be coordinated to produce effects, given the regime. This active research agenda does more than passively discern the effects of extant resource configurations; it seeks valid causal inferences about specific instructional designs.

This agenda would give priority to research on designed systems of instruction, and thus would require excellent programs of development, field testing, and revision. A focus on regimes also would imply a high priority on experimental and quasi-experimental tests under varied resources constraints. Our principle of relativity also means that there could be neither "regime-free" answers to questions about levels, combinations, and coordination of resources, nor "regime-free" studies of their effects.

Given our analysis of mutual adjustment within instruction, consistent regimes seem the most reasonable way to probe causal relationships between resources and learning. It would not be useful for researchers to attempt to disassemble regimes into their components and do conventional research on their effects. The work that we propose would improve understanding of the complex relationships within teaching and learning and open up opportunities for more coherent research on instruction and its effects.

Our approach also would tend to redefine the opportunities for, and limits on, the sorts of passive research that have become conventional in the study of instructional effects. Observational or survey research typically use existing achievement tests, which do not offer a student outcome measure that expresses the aims of a specific, well-conceived approach to instruction. A deliberately designed regime will set clear instructional goals, and research on it would require outcome measures that assess achievement of these goals. Broad-purpose achievement tests would be used within the agenda we propose, since it would be important to know how new regimes bear on more traditionally defined success. But conventional assessments are unlikely to capture the proximal outcomes of a well-defined system of instruction.

Even if that problem was solved, passive inquiry cannot yeild strong evidence on the effects of instructional systems in best-case situations or under resource constraints. For instruction is a system of interaction in which students and teachers continually mutually adjust, so it would be extraordinarily difficult either to uncover and delineate how a given resource is used, or to distinguish well defined regimes. Existing instructional arrangements and resource use emerged through historical processes of negotiation and accommodation. To the extent that regimes occur naturally, they have developed in part to cope with the resource constraints of given settings. It would be difficult or impossible to answer the question "What resources are essential, given the regime," because of mutual adjustments around existing resources. Instructional practice within a given setting tends to involve a mix of individualized adaptations, and in natural conditions there would be little chance to hold the regime constant and vary resources. If only survey data are available, it is essential to measure student background and school context and statistically adjust for them in models that relate instruction to outcomes, but that can tell us little about what would happen if instruction were deliberately modified. Ethnographies and surveys can reveal how teachers and students think and act within a setting, but they cannot reveal how things would change with new regimes and resources. Deliberately developed regimes would.

Passive inquiry would play several roles in the approach that we sketched. Large-scale surveys could roughly estimate the range of instructional approaches, and related resource availability, within regimes. Paired with ethnographies, surveys might enable researchers to discern the extent to which anything like coherent systems of instruction occur "naturally," and, if they do, to identify them. Carefully focused ethnography could clarify the configuration and operational features of existing regimes. Ethnography also can be invaluable to those who design instruction, for "... designers can deepen their understandings of and therefore broaden their abilities to describe their own regimes when they closely observe and interrogate expert teachers who implement them (because the teachers will raise questions about gaps in the current description, do some things in ways that are different from and often better than what the regime called for, and add things that the developers may want to adopt as useful elaborations)" (Brophy, 1988, p. 15). Paired survey and ethnographic research could illuminate what students know and can do over a range of naturally occurring instruction and settings, and focus attention on where educational effort might be directed. Studies of school facilities and resource allocation would be useful, mostly for how they enable or constrain the implementation of well-defined instructional regimes.

Though we argue that instruction is so interactive as to preclude treating resources as individual variables, we do not argue that instructional systems must remain black boxes. The design of instructional systems would require extensive learning about how instructional systems work, both in their development and in practice, under optimal and various sub-oprimal conditions. Micro-ethnographies would be helpful in both areas, and could suggest ways to design instruction for specific sub-groups. Research that clarified the internal dynamics of instructional system would be especially useful in comparing dynamics across resource variations within regimes.

The approach sketched here contains important roles for active and passive research. One justification for the former is that it would give an explicit definition to regimes and resources, thus creating a basis for valid causal inference. Another is that it would create a useful context for survey and ethnographic research, which currently float largely free of knowledge-building frameworks. Extant instruction reflects accommodations to currently available resource levels, to views of student background, and to prior achievement. Valid causal inference about the effects of instruction or resources is extremely elusive in such webs of mutual accommodation. Economists would describe this as a situation in which the causal variable of interest is "endogenous," that is, determined in part by current levels of outcomes and other unobservable factors that lead educators to make choices and compromises. That makes it very difficult to separate effects of causal variables from those of a host of other factors, observed and not. Naturalistic survey and ethnographic research can help to advance understanding in several important areas, but they are not well suited to producing defensible causal conclusions. Doing that requires causal variables to be made "exogenous," i.e., varied independent of confounding factors.

The best way to do that is through deliberate, well-defined interventions, to which schools or classrooms are assigned randomly. Such assignment of students to regimes may be feasible in some instances, but we anticipate that schools or classrooms could more often be so assigned.[20] Interpreting the results of such experiments would require carefully designed research on the dynamics of instruction within regimes. Active and passive research would be interdependent: without ethnographic research on instructional dynamics, it would be difficult or impossible to grasp the role or importance of various influences on instruction, and so to interpret the experimental evidence. The results of passive research programs also could help to generate ideas for regimes and resource allocation within them, as well as helping to explicate the dynamics of well-developed instructional systems. Active programs of research, in which deliberate interventions vary resources in relation to well-articulated regimes, are at the heart of our proposal, but so are well-designed programs of passive research.

Some skeptics expect that administrators, teachers, or parents will flatly resist randomized assignment to alternative instructional approaches. While there surely will be resistance to such studies in some instances, there may also be strong incentives to participate. Participation reveals a public commitment to school improvement, and will typically bring new resources to participants. Moreover, school districts are currently under pressure to adopt innovative approaches; tying these to sound research will often be appealing. We expect those who support instructional innovation to offer incentives to participate in serious tests of innovations. In some cases, random assignment to novel regimes will not be practical. In these cases, users will select the regimes, and sound quasi-experimental design will be essential. Researchers must identify and control selection biases, and causal inference will be more tentative (Cook & Campbell, 1979). But in both true and quasi-experiments, study of the dynamics of instruction would be essential to illuminate the role and importance of various influences on instruction.

Effective instructional systems will only result from systematic development, including research that makes it possible to specify the required resources, and their relation to specific aims. In addition to the benefits already discussed, the development of such systems would tend to support more precise and common professional conversation. It also would make it possible to rigorously evaluate alternative regimes relative to common goals, to evaluate claims about the effects of particular levels or combinations of resources, within regimes, and to enable evaluation of varied versions, for which resource requirements differed. Such research is essential both to the more applied task of learning how to improve schooling, and to the more basic task of defining educative resources and learning how they are used in instruction.

Conclusion

The research program that we have sketched is not a design for all educational research, but for inquiries that focus on resource and instructional effects. We proposed a dramatic shift, from a research frame that gives priority to conventional resources and asks how they affect learning, to one that gives priority to coherent systems of instruction and asks how resources are used within them. One key premise is that because resources become active when used in mutual instructional adjustment, they are unlikely to have a fixed instructional value. Their value is likely to depend on the uses to which they are put, which in turn depends on the ends and means of instruction. To understand the nature and effects of resources, researchers must focus on how instructional ends and means are defined, and on what resources are crucial to them. Thus we have proposed designing coherent instructional regimes, submitting those regimes to tests of their effectiveness, and assessing how resource constraints modify their effectiveness. Within this framework, a variety of passive forms of research, including surveys and ethnographies, can play important roles.

Our proposal has complementary benefits and costs. On the one hand, our picture of instruction as a system of interactive mutual adjustment complicates understanding of the dynamics of teaching and learning, and of the ways in which resources influence them. In such a system, the value of resources is likely to depend on the ways they are used. That raises fundamental questions about how validly conventional research can tease out the causal influence of particular resources, across a great variety of schools and classrooms. That may unsettle many researchers, but it seems inescapable if our account of instruction is roughly right. On the other hand, our account offers a theoretical frame for research on instructional and resource effects that builds on several decades of work, that opens up promising research agendas, and that creates opportunities to lodge active and passive research within mutually reinforcing knowledge-building structures.

Some might argue that these agendas would be insufficient to illuminate policy makers' decisions. To know what resources are optimal for a given approach in mathematics tells us little about what resources are needed in general. Small classes might be needed to enact a given approach in literacy, but teachers' subject matter preparation, rather than small classes, might be the crucial ingredient in teaching an effective math curriculum to the same grade. Small classes taught by knowledgeable teachers may not be fiscally feasible. Varied studies would tend to send mixed signals about how many teachers to hire and what qualifications to require, but that is just what extant research has done. One chief task of a coherent educational research program would be to make just such trade-offs visible, based on sound empirical study. If our approach is correct, policymakers would be well advised to adopt more complex approaches to resource allocation, that capitalize on the role of resource use.

Others might argue that developing such an agenda is infeasible because well-specified instructional regimes could not be devised or because experiments could not be done, or because the entire enterprise would be too costly. Yet the last decade's work in reading at The National Institute for Child Health and Development, and in some whole-school reforms, show that carefully designed systems of instruction can be created, and that experimental research is possible (Cook, Hunt, & Murphy, 1999). The work that we propose would be difficult, but if educators and researchers took the ideas seriously, a great deal that seems difficult today could soon be feasible. One reason for our confidence is the rising demand for solid evidence on the effects of such interventions; there is likely to be a market for just such work. Another is the recent experience with reading research and some whole-school reform models. Still another is the success of several seemingly impossible experiments in health care, housing, and welfare. And another is the growth evidence-based medicine, which faced similar problems.

The sorts of research and instructional design that we have sketched would take careful planning and clear priorities, for such work must be strategic. Researchers cannot investigate everything, or even half of what they might wish to. We aim to investigate a few key issues well. Even a modestly designed program would require a broad and energetic constituency that included public and nonpublic supporters and more capable management than educational research has had. It also would require a level of federal commitment to scientific research management that was closer to health research than to education. The result could yeild a stream of sound evidence on the resources required to attain a variety of educational aims, and inform thought and debate about the aims of schooling and levels of investment in education. In time it could inform debate and decisions, and might even close out unfruitful arguments as well as highlight new problems. But it could not prescribe decisions about resources, for those require interaction among people and groups whose authority to decide is civic rather than scientific, and who often differ. Research on instructional resources could inform but not replace a broad discussion about schools and their improvement.

Notes

[1] Base-ten blocks are wooden blocks designed to model place value: a small cube, a rod composed of ten small cubes, a flat square composed of a hundred of the small cubes or ten of the rods, and a block made of ten of the flat squares, or a thousand little cubes. In this case, the teacher is using the little cubes as ones and the rods as tens.

[2] For convenience, we often refer in what follows to "instruction," in which we include this clump of practices, rather than either teaching alone, or the more clumsy "teaching and learning with materials."

[3] We are indebted to Jere Brophy (personal communication, Nov 8, 2002), for this and several other important points.

[4] See Hanushek, (1999). Several analysts agree with him that when the size of achievement gains attributable to class size reduction are compared with the costs, CSR is shown to be very expensive. See Slavin, (1990), and Levin, Glass, and Meister, (1984).

[5] Hanushek, (1999) p. 157. Some differences between the two analyses may be due to differences in the analyses; Krueger's are based on a pooled within-school sample, but Hanushek's is not.

[6] Given Glass et al's analysis, it is puzzling that Finn and Achilles write: "... dozens of earlier studies" did not help to clarify "... the classroom processes that distinguish small from large classes" (Finn & Achilles, 1990 p. 102). Yet Glass et al. discuss nearly all of the mechanisms taken up by Finn and Achilles and other studies. Slavin, (1990), reanalyzed Glass's data, and reported that much of the effect was due to tutorials, some of which had no academic content.

[7] Teachers "... rated each pupil who had been in STAR on the ... Student Participation Questionnaire ... [which assesses specific learning behaviors ... judged by educators to be important ... The instrument yields reliable, valid measures of the effort students allot to learning, initiative taking in the classroom, and ... disruptive or inattentive-withdrawn behavior. Finn and Achilles, (1999 pp. 101–02).

[8] Jeremy Finn (personal communication, 6/17/02), reports unpublished analyses showing that the size and direction of the effects are inversely tied to student SES; the more advantaged students were, the smaller the positive effects on achievement.

[9] Finn and Achilles, (1999) prefer this explanation.

[10] All teachers are willing or able to do so; see Blatchford, et. al (2002).

[11] Blatchford, et. al (2002), report similar results in a British study; students in small classes got more of teachers' time, and teachers were more satisfied, but there was no report of dramatic change in teaching.

[12] The TBSF tests were used in the primary grades, but the CTBS and Stanford Achievement Test revealed similar results; Finn and Achilles (1990).

[13] Ehrenberg, Brewer, Gamoran, and Willms (2001) argue that the institutionalized nature of schooling, in which pedagogy is decoupled from organization, is part of the explanation, and technical constraints like required texts are another.

[14] Mitchell, Beach, and Badarak (1989) argue that the advantage of small classes may be the result of nonrandom assignment of very low achieving students to large classrooms (alternatively, the nonrandom assignment of higher achieving students to small classes). Their models of classroom processes are consistent with the skewing of large class teaching to the weakest

students, which, in turn, could result in slower learning for all. The presence of more low-achieving students in large classes than would have been expected by chance from the overall sample suggests possible migration of more able students to smaller classes; if so, this would be a serious threat to the internal validity of the STAR experiment. We have found no way to check on this possibility. Even if Mitchell et al. are correct, from our perspective they only offer a different causal mechanism for explaining the use of small class size. For a very thoughtful discussion of the STAR experiment on this and other points, see Goldstein and Blatchford (1998).

[15] This probably is what Eric Hanushek (1989, 1996) means, when he argues that teachers mediate the effects of resources.

[16] Jere Brophy refers to these as "threshold effects" (personal communication, November 2002), and notes that they probably hold for range of resources.

[17] The causal effect of treatment A relative to treatment B is defined as the difference between the potential outcomes of a student under A or B. Causal inferences are thus meaningful only when a student is potentially exposable to alternative treatments (Holland, 1986). If a regime were perfectly and thus strictly enacted, only one possible "treatment" would be conceivable at any moment for a particular student: there would be a treatment A but no treatment B! This means that such a student would have only one potential outcome, so that no causal effect could be defined. In practical terms, this means that, within a perfectly enforced regime, it is not possible to find or even imagine a student who is similar to the student of interest but who receives a different dose. It is quite feasible, however, to define the causal effects of *alternative regimes* and to design studies that can estimate those causal effects.

[18] The renowned "pygmalion" experiments (Rosenthal & Jacobson, 1968) in classrooms can be read as a scheme to change the instruction (read "treatment") that teachers offered students by experimentally inflating evidence on students' IQs (Raudenbush 1984).

[19] Medical researchers distinguish between efficacy and effectiveness trials. Efficacy trials establish whether a treatment can have significant positive effects in carefully controlled settings with plentiful resources; they attempt to estimate a maximum potential effect. Effectiveness trials take efficacious treatments to the field where implementation is more challenging and resources more constrained. Our recommended approach to studies of educational improvement is similar, with the added proviso that effectiveness trials deliberately vary key resources so that their effects can be rigorously evaluated.

[20] For example, in evaluating whole-school reform efforts, one might envision assigning schools rather than students randomly to treatments. This could be done ethically if many schools expressed interest in adopting a reform but resources allowed implementation in only a restricted number of sites at any one time. Intervenors could promise all schools the opportunity to participate, but the timing of participation would be decided via lottery. A randomized "wait-list" control group of schools would then be available. If instructional systems were instead chosen by school or classroom, a strong effort to explicate the regime and associated resources would be in order, as would tentative causal inference.

References

Alexander, K., Entwisle, D., & Olson, L. (2001). Schools, achievement, and inequality: A seasonal perspective. *Educational Evaluation and Policy Analysis, 23*(2), 71–191.

Betts, J. R., & Skolnik, J. L., (2001). The behavioral effects of variation in class size: The case of math teachers *Educational Evaluation and Policy Analysis, 23*(2), 193–215.

Bishop, J. (1998). *The effect of curriculum-based external exit exam systems on student achievement.* Consortium For Policy Research In Education. Philadelphia, PA: University of Pennsylvania.

Bishop, J., & Mane, F. (1998, November). The New York state reform strategy: Incentive effects of minimum competency exam graduation requirements. Paper presented at the National Invitational Conference on "Education in Cities: What Works and What Doesn't," Wisconsin.

Blatchford, P. Moriarty, V., Edmonds, S., & Martin, C., (2002). Relationships between class size and teaching: A multimethod analysis of British infant schools *American Educational Research Journal, 39*(1), 101–132.

Brophy, J. (1988). Research on teacher effects: Uses and abuses. *The Elementary School Journal, 89,* 3–22.

Brophy, J., & Good, T. (1986). Teacher behavior and student achievement. In M. C. Wittrock (Ed.), *Handbook of research on teaching, third ed.,* (pp. 328–375). MacMillian Publishing.

Bryk, A., Lee V., & Holland, P. (1993). *Catholic schools and the common good.* Cambridge, MA: Harvard University Press.

Coleman, J., Campbell, E., Hobsen, C., McPartland, J., Mood, A., Weinfeld, F., & York, R., (1966). *Equality of educational opportunity survey.* Washington, DC: U.S. Government Printing Office.

Cook, T. D., & Campbell, D. (1979). *Quasi-experimentation.* New York: Rand McNally.

Cook, T. D., Hunt, H. D., & Murphy, R. F. (1999). *Comer's school development program in Chicago: A theory-based evaluation.* Unpublished paper, Chicago, IL: Northwestern University.

Cooley, W., & Leinhardt, (1978). *Instructional dimensions study.* Pittsburgh, PA: Learning Research And Development Center, University of Pittsburgh.

Dewey, J. (1905). *The child and the curriculum, and school and society.* Chicago, IL: University of Chicago.

Dweck, C. S. (1986). Motivational processes affecting learning. *American Psychologist, 5,* 1179–1187.

Dweck, C. S. (1988). Children's thinking about traits: Implications for judgments of the self and others. *Child Development, 69*(2), 391–403.

Edmonds, R. (1979). Effective schools for the urban poor. *Educational Leadership, 37,* 15–27.

Edmonds, R. (1984). School effects and teacher effects. *Social Policy, 15,* 37–39.

Englert, C. S., Raphael, T. E., Anderson, L. M., Anthony, H. M., & Stevens, D. D. (1991). Making strategies and self-talk visible: Writing instruction in regular and special education classrooms. *American Educational Research Journal, 28*(2), 337–372.

Ehrenberg, R., Brewer, D., Gamoran, A., & Willms, D., (2001). Class size and student achievement. *Psychological Science In The Public Interest, 2*(1), 1–30.

Evertson, C. M., & Folger, J. K. (1989, March). *Small class, large class: What do teachers do differently?* Paper presented at the annual meeting of the American Educational Research Association, San Francisco, CA.

Evertson, C. M., & Randoph. (1990). Teaching practices and class size: A new look at an old issue. *Peabody Journal Of Education* 85–105.

Ferguson, R. F. (1991). Paying for public education: New evidence on how and why money matters. *Harvard Journal on Legislation, 28,* 465–498.

Finn, J. D., & Achilles, C. (1990). Answers and questions about class size: A statewide experiment. *American Educational Research Journal, 27,* 557–77.

Finn, J. D., & Achilles, C. M. (1999). Tennessee's class size study: Findings, implications, misconceptions. *Educational Evaluation and Policy Analysis, 21*(2), 97–109.

Folger, J., (1989). Lessons For Class Size Policy And Research, *Peabody Journal Of Education, 67*(1) 123–132.

Goldstein, H., & Blatchford, P. (1998) Class size and educational achievement: A review of methodology with particular reference to study design. *British Educational Research Journal, 24*(3), 255–268.

Hanushek, E. A. (1981). Throwing money at schools. *Journal of Policy Analysis and Management, 1,* 19–41.

Hanushek, E. A. (1989). The impact of differential expenditures on school performance. *Educational Researcher, 18*(4), 45–65.

Hanushek, E. A., (1996). School resources and student performance. In G. Burtless (Ed.), *Does money matter?,* (pp.) Washington, DC: Brookings Institution.

Hanushek, E. A., (1999) Some findings from an independent investigation of the Tennessee STAR Experiment, and from other investigations of class size effects. *Educational Evaluation and Policy Analysis, 21*(2), 143–163.

Hedges, L. V., Laine, R. D., & Greenwald, R. (1994). Does money matter? A meta-analysis of studies of the effects of differential school inputs on student outcomes (An exchange: Part I). *Educational Researcher, 23*(3), 5–14.

Holland, P. (1986). Statistics and causal inference. *Journal of the American Statistical Association, 81*(396), 945–960.

Jencks, C., Smith, M., Acland, H., Bone, M., Cohen, D., Gintis, H., Heyns, B., & Michelson, S. (1972). *Inequality.* New York, Basic Books.

Krueger, A. (1999). Experimental effects of education production functions, *Quarterly Journal Of Economics, 114*(2).

Lampert, M. M., (2001). *Teaching problems and the problems of teaching.* New Haven, CT: Yale University Press.

Leinhardt, G., Zigmond, N., & Cooley, W., Reading instruction and its effects. *American Educational Research Journal, 18*(3), 343–361.

Levin, H., Glass, G., & Meister, G., (1984), Cost-effectiveness of educational interventions. Report # 84-A11, Stanford, CA: Stanford Center For Educational Research.

McLaughlin, M. & Talbert, J. (1999). *Professional community and the work of high school teaching.* Chicago, IL: University of Chicago Press.

Mitchell, D., Beach, S., & Badarak, G. (1989). Modelling the relationship between achievement and class size: A re-analysis Of the Tennessee Project Star Data. *Peabody Journal Of Education* 34–74

Molnar, A., Smith, P. Zahoric, J., Parker, A., Hallbach, A., & Ehrle, K. (1999). Evaluating the SAGE Program: A pilot program in targeted pupil-teacher reduction in Wisconsin. *Educational Evaluation and Policy Analysis, 21*(2), 165–177.

Mosteller, F. (1995). The Tennessee study of class size in the early school grades. *Future of Children, 5*(2), 113–127.

Newmann, F. M., & Wehlage, G. G. (1995). *Successful school restructuring: A report to the public and educators.* Center on Organization and Restructuring of Schools, University of Wisconsin-Madison: U.S. Department of Education, OERI. R117Q00005-95.

Odden, A, & Busch, C., (1998). *Financing schools for high performance: Strategies for improving the use of educational resources.* San Fransisco, CA: Jossey-Bass.

Palinscar, A. M., (1986). The role of dialogue in providing scaffolded instruction, *Educational Psychologist, 21,* 73–98.

Palinscar, A. M., & Brown, A., (1984). Reciprocal teaching of comprehension-fostering and comprehension-monitoring activities, *Cognition and Instruction, I,* 117–175.

Raudenbush, S. W. (1984). Magnitude of teacher expectancy effects on pupil IQ as a function of the credibility of expectancy induction: A synthesis of findings from 18 experiments. *Journal of Educational Psychology, 76*(1), 85–97.

Rice, J. (1999). The impact of class size on instructional strategies and the use of time in high school mathematics and science courses pp. 215–29.

Ritter, G. W., & Boruch, R. F. (1999). The political and institutional origins of a randomized controlled trial on elementary school class size: Tennessee's Project STAR. *Educational Evaluation and Policy Analysis, 21*(2), 111–126.

Robins, J. M., Greenland, S., & Hu, F. (1999). Estimation of the causal effect of a time-varying exposure on the marginal mean of a repeated binary outcome. *Journal of the American Statistical Association, 94*(447), 687–701.

Rosenholz, S. J. (1985). Effective schools: Interpreting the evidence. *American Journal of Education, 93,* 352–388.

Rosenthal, R., & Jacobson, L. (1968). *Pygmalion in the classroom.* New York: Holt, Rinehart and Winston.

Rubin, D. B. (1974). Estimating causal effects of treatments in randomized and nonrandomized studies. *Journal of Educational Psychology, 66,* 688–701.

Rutter, M., Maughan, B., Mortimore, P., Ousten, M., & Smith, A. (1979). *Fifteen thousand hours: Secondary schools and their effects on children.* Cambridge, MA: Harvard University Press.

Skrla, L., Scheurich, J. J. (2001). Displacing deficit thinking in school district leadership. *Education and Urban Society, 33*(3), 235–262.

Slavin, R., (1990). Class size and student achievement: Is smaller better? *Contemporary Education, 62*(1), 6–12.

Slavin, R., (1990). Class size and student achievement: Small effects of small classes. *Educational Psychologist, 24,* 99–110.

Zahoric, J. (1999). Reduced class size leads to individualized instruction. *Educational Leadership, 57*(1), 50–53.

Authors

DAVID K. COHEN is John Dewey Professor of Education, Walter A. Annenberg Professor of Public Policy University of Michigan School of Education 610 East University, 4109 SEB; DKCohen@umich.edu. His aeras of specialization are education policy, the relations between policy and practice, and school improvement.

STEPHEN W. RAUDENBUSH is Professor of Education and Statistics, School of Education, and Senior Research Scientist, Institute For Social Research, University of Michigan 610 East University, 4109 SEB; His areas of specialization are.

DEBORAH LOWENBERG BALL is Thurnau Professor of Education, School of Education, University of Michigan, 610 East University 4109 SEB; dball@umich.edu; Her areas of specialization are teacher education and mathematics educaton.

Manuscript Received June 20, 2002
Revision Received January 20, 2003
Accepted June 4, 2003