EDUCATIONAL INNOVATION AND THE PROBLEM OF SCALE*

David K. Cohen and Deborah Loewenberg Ball
School of Education, and
School of Public Policy
The University of Michigan
Ann Arbor, Michigan
January, 2006

*To be published in 2006 by Rowman and Littlefield, in a volume Titled Scale Up In Principle. An earlier draft of this paper was presented at a 2003 meeting ("Conceptualizing Scale-Up"), sponsored by the Data Research and Development Center of The University of Chicago, and supported by the National Science Foundation, in connection with the Interagency Research Initiative (IERI). The ideas in the paper owe a good deal to our work with Karen Gates, Joshua Glazer, Simona Goldin, and Donald Peurach, all colleagues in the Study Of Instructional Improvement at The University of Michigan. We owe special thanks to Donald Peurach and Simona Goldin for the improvements that they made to this draft.

OVERVIEW

The U.S. has had five decades of increasing pressure for school reform, beginning before the Soviet Sputniks and continuing today. Researchers appear to believe that none of these reforms were implemented at scale – i.e., widely and well.[1] Despite this view, most authors write as though it was reasonable to expect innovations to be adopted widely and well; when studies report innovation failure, they usually do so in a disappointed tone.[2] Yet, like love affairs, many innovations consist more of ideas and hopes than of the carefully designed details of daily operations that often are required to make appreciable change. Our scrutiny of the literature revealed no discussions of what might reasonably be expected from innovation in education, let alone what might be expected, given the nature of schooling in the U.S. We found only three studies that sought to discern broad patterns of innovation, adoption, or implementation.[3]

The contemporary presure for major national improvement of student learning has created an appetite for better knowledge about innovation. We sketch what is known and believed about innovation, offer some conjectures that merit investigation, and consider strategies for successful innovation.

SOME KEY IDEAS

Begin at the beginning: what is an innovation? It is a departure from current practice – deliberate or not, originating in or outside of practice, which is novel. Innovations include novel practices, tools or technologies, and knowledge and ideas. In some cases there are clear distinctions between "designers" and "users" of innovations, as when a textbook publisher markets a new text to a state, and teachers and students use the books. In other cases, designers and users are one in the same, as when teachers devise their own innovations. On this inclusive view there are several sorts of innovators, including practitioners, agencies that sponsor and govern schools, and organizations external to practice that urge, design, and market changes. These innovators often are associated with different sorts of innovations, but all have two things in common: they urge a departure from conventional practice, and both design and use require accommodation with the environment. When designers consider what sort of innovation may appeal, or work, or help, they weigh such accommodation among many other things. If practitioners decide to use an innovation, they are likely to consider and take advantage of such accommodations.

What is scale? The answer depends partly on the innovation, and partly on how we

---

[1] Michael Fullan, who has done the most extensive work in this field, argues that view in The Meaning Of Educational Change, 1982.
[2] Stake and Easley, 1987. Welch, 1979.
[3] The longest-running study is Everett Rogers' Diffusion of Innovations, 2003, which deals, even in its later editions, only glancingly with implementation. The most comprehensive is Fullan, op cit. The other study is Paul Berman and Milbrey McLaughlin, 1979, which considered several federally sponsored programs.

judge scale. One could judge it in terms of adoption, or in terms of use, or in terms of intended use. Adoption is always a different matter than use, but as innovations become more complex, the difference grows. In some cases – texts, curricula, and tests, for instance – one might plausibly contend that the extent of scale can be decided quantitatively, by the number of schools that adopt it. To require a new textbook series in New York City, and get it to all teachers, is to achieve a sort of scale, since all or nearly all of the texts will be used somehow. But that tells us little about use, because it depends on many other factors than adoption. In other cases – Comprehensive School Reform Designs (CSRD), for instance – adoption is a much weaker proxy for assessing the extent of scale, because these designs cannot be used in any meaningful sense unless there is extensive qualitative change in the schools that adopt them.

In the latter case, the very meaning of scale is distinctive, because its qualitative elements can be as important as, or more important than quantitative elements. How deeply an innovation permeates practice may be as important to an assessment of its success at scale as how many sites adopt it, and innovations differ in how much is needed for them to deeply affect practice. For instance, many whole school designs require building school infrastructure for professional education, supervision, and leadership that will enable use. Success For All and other CSRD designs contain such infrastructure as key elements of the designs, and they are accompanied by unprecedented external assistance. In contrast, though textbooks contain directions for their use by teachers, most are designed to be used with little direction or external assistance. Whole school reforms, in contrast, contain elaborate and often carefully staged guidance for use; so central are these directions that they often are indistinguishable from "the innovation."

If this analysis is roughly correct, there can be no single definition of scale, nor any single metric to judge it, for the tasks and problems of implementation vary with types of innovation.

EXPLAINING FAILURE

We take up these explanations because they have been the focus of much prominent research on innovation, and thus offer a quick view of the terrain. There are at least three schools of thought.

> Innovations fail because they were badly designed. For some researchers, this means that innovations fit poorly with practice.[4] Others argue that they do not make room for practitioners to adapt them to practice.[5] Still others argue that the key design flaw in innovations is that they do not make room for teachers to learn how to use them. In these lines of work, innovations are the villains and schools are the

[4] Larry Cuban, 1993.
[5] Berman and McLaughlin, op. cit.

victims, for the problems arise in the poor design of innovations. [6]

In a second sort of work, schools are the villains and innovations the victims. Researchers argue that innovations fail because schools and school systems offer few or no incentives to change practice[7], or that educators resist change[8], or that schools are so rigidly organized that they leave no room for innovation,[9] or that schools are institutionalized organizations which effectively buffer the technical core from environmental pressures.[10]

In a third way of framing the issues, researchers claim that few innovations succeed because few were robust treatments that addressed problems that seriously concerned practitioners. When robust designs that addressed serious problems of practice were tried, they reached scale.[11]

In a fourth set of views, the problems of innovation center in environments. One argument focuses on the complexity and fragmentation of educational environments, which promote openness to proposals for change at the same time as they inhibit consistency of adoption and implementation among units and over time.[12] Another argument holds that the openness of U.S. educational environments enables the emergence of many educational innovations at all levels, but the school improvement industry tends to inhibit the growth, maturation, and institutionalization of organizations that develop and sponsor innovations.[13]

KNOWLEDGE ABOUT INNOVATION

Each account has some appeal, but how might we judge their relative merits? We begin to explore that by asking what we know about three rudimentary matters. Who adopts innovations? What sorts of innovations do we consider? In what environments do they occur? The answers to these questions will shed a bit of light on the answers to another: what can reasonably be expected concerning the scale of innovation? The answer to that could help us to evaluate the explanations of presumed innovation failure, sketched above. It might also suggest some conjectures about innovation and the scale of change that researchers could test.

ADOPTERS. Though there has been little research on what or who adopts

---

[6] Seymour Sarason was the first to write about this problem, in 1971. A more general version of this perspective is offered by Cohen and Levinthal, 1990.
[7] Eric Hanushek, 1989. Milton Freidman, 1962.
[8] Dan Lortie, 1975
[9] Sarason, op. cit.; Lortie, op. cit.
[10] John Meyer and Brian Rowan, 1977
[11] Cohen and Ball, 2003.
[12] Cohen, D.K., and Spillane J., 1993.
[13] Rowan, B., 2002.

innovations, we do know a few important things. One is that the size and organization of the adopting units vary. Adopters include individual professionals, individual schools, districts, states, and federal agencies. Some adoptions bridge these units. The most commonly studied adopters are teachers and schools; there are fewer studies of districts, states, or federal education agencies. Yet the very considerable differences among adopters suggest a question: would explanations for the actions of one sort of adopter be likely to hold for others – i.e., would we expect explanations for the adoption of teacher-initiated innovations to be similar to explanations for the adoption of state-sponsored innovations? That seems unlikely. In addition, it seems likely that innovations that are to be adopted and used by individual teachers would have greater difficulty spreading than innovations that must be adopted by states. We found no studies that probe these issues.

The relations between innovators and adopters also vary quite dramatically. In an appreciable number of cases, the adopter also is the designer, as in the case of many teacher-devised innovations, and for some LEA and state-devised innovations. In such cases, the meaning of "adoption" is different than when considering innovations that are designed by people and organizations that are external to the intended adopters. For when adopters and users are not also the innovation designers, a variety of personal, professional, and organizational differences are likely to come into play around adoption and use. It also seems likely that implementation would differ between the two types. We found no analysis of these matters.

Adopters' resources vary. The U.S. is distinctive among developed nations for great inequalities in the educational, human, and fiscal resources to which adopters have access. That is true among states, among local educational authorities (LEAs) within states, and among teachers within LEAs. . If schools, LEAs, and states are at all like other organizations, we would expect such resource differences to influence the design, adoption, and use of innovations. We would not expect teachers in Mississippi delta schools to either initiate or adopt innovations at the same rate or with the same facility as teachers in Evanston Illinois or Shaker Heights Ohio. Moreover, we would not expect explanations for innovation to be the same across those differences. These variations suggest differences in the potential for scale, and in the requirements for achieving it. Studies of innovation infrequently mention these matters.

Whether devised by practitioners or external agents, the idea of "scaling up" an innovation brings with it the problem of "transfer": getting the innovation from one place to another. Researchers conceptualized that in different ways. Some portrayed it as agentless "diffusion". Others saw it as a matter of uni-directional "broadcasting". Still others represented it as a fairly straightforward process of bi-lateral communication. When considering scale-up, the challenges of both framing the processes of transfer and of devising ways to deal with it hold for the practitioner-designer who seeks to spread the innovation to others, and for the external-designer who seeks the same. As research and experience accumulated, these earlier conceptualizations of "transfer" turned out to be weak, to be focused on not-so-complex innovations, and thus made "scaling up" appear to be pretty transparent. There also was little attention to the use of innovations and their implementation.

INNOVATIONS differ. One source of variation is sponsorship: some are created and sponsored by individual practitioners, some by LEAs, others by state and federal governments, and still others by agencies outside of schools or government, including publishers, civic leaders, advocacy groups, and professional associations. Some are multiply sponsored, as with privately published texts that gain state adoption. Such differences are likely to create variation in the comparative advantage for scale of some innovations. We found no studies that probe these differences.

Another difference among innovations concerns the sector of schooling at which they aim: do they concern finance, instruction, budgeting, assessment, professional licensing, or hiring? Each is a distinct operational domain, with distinctive constraints and opportunities. Each bears on others, but none is overarching. It seems reasonable to expect that the adoption and use of innovations in some of these domains would be explained differently than in others. For instance, the influences on innovations in budgeting are likely to differ, at least in part, from those in professional licensing and hiring. We found no studies that investigate these differences.

Still another difference among innovations centers on whether they occur as the result of standard operating procedures. If they do, they are more likely to be routinely adopted at a large scale. Examples include state-approved text books, state academic standards, and state test programs. In these cases, adoption is either compulsory or very nearly so. Many other innovations are not due to standard operating procedures. Adoption thus is more voluntary, and typically there are no penalties for non-adoption. Innovations of the first type are more likely to achieve some version of scale broadly and quickly, while those of the second type are less likely to do so. Teacher-made innovations would face rather different problems of scale-up – including communication, resources, and authority -- than state-made innovations. We found no studies that investigate these matters.

Most research on innovation concerns instruction, and a substantial fraction concern efforts to promote ambitious instruction, including NSF and other curricula, some state standards-based reforms, and Progressive education. These are a distinctive sort of innovation, for they pertain to what many researchers refer to as the “technical core” of schooling, a domain that, until very recently, has been little managed by extra-classroom authorities. It is difficult to imagine why patterns of or explanations for the creation, adoption, and use of instructional innovations would resemble patterns of or explanations for the creation, adoption, and use of changes in school schedules or school finance, for the phenomena to be described and explained, and the units of analysis, are quite different. We found no research that addresses these differences.

One of the most distinctive differences among instructional innovations is in their ambitions for academic work, and thus the demands that they make on practice. Some call for sharp departures from conventional practice, because they seek fundamental re-orientation of teaching and learning. Others have more modest aims. As Eugene Bardach pointed out in a different context, innovation creates incompetence by proposing to retire familiar practices in favor of novel and less familiar practices.[14] More ambitious innovations cause or imply more incompetence among practitioners.

Innovation at scale — or implementation — is a function of design, adoption, and use. Two features of innovation design can affect implementation by influencing practitioners' capability to learn new practices and thus overcome incompetence. One, which we call "elaboration," deals with the detail with which a reform is developed; the second, "scaffolding," is the degree to which the innovation includes a design for and other means of learning to carry it out. We consider these in turn, beginning with elaboration: innovation designs are plans for what is to be done, and those plans can be more or less rich and detailed.[15] Every innovation is elaborated to some degree. Some consist primarily of goals, principles and visions that suggest broad directions. Others include curriculum, examples of intended learning and teaching, and tests. Full elaboration is impossible, since even modest departures from conventional practice are too complex to exhaustively plan, and even routine classes include unpredictable events. Even very elaborated designs leave room for much invention, as implementers reorganize practice around them. Some things are best worked out in practice, and much elaboration can only occur in practice. Elaborating a design is more or less of a beginning, and always entails trade-offs between designers' view of what is essential, and limits on what can be spelled out. Given the inevitable incompleteness of any innovation, what is elaborated matters, for that influences both the quality of implementation and the distribution of work between designers and implementers.

Elaboration is sometimes treated as a matter of preference; some say that innovations should not constrain autonomy and invention, while others contend that little would happen without detailed direction. But to elaborate innovation designs is not necessarily to constrain, for different sorts of elaboration are possible. An intellectually ambitious elementary mathematics curriculum that was designed to encourage students' active engagement in mathematical reasoning and problem solving might be elaborated only in terms of its broad objectives. Or it could be further elaborated in terms both of its objectives and its main mathematical themes. Or it could be more elaborated in terms of its objectives, main mathematical themes, and the specific types of instructional activities. It could be even more elaborated in terms of all those things, and extensive examples of desired student performance. Or it could be differently elaborated in lists of

[14] Bardach writes: "incompetence is not, of course, a trait like having brown eyes. It is a description of a relationship between a task or function in the situation up to a given, though ultimately arbitrary, standard of some sort…it may not in any meaningful sense be 'his fault' that he is incompetent to perform certain tasks – it may well be the fault of the people who assigned him the task in the first place" 1977, p. 126.
[15] Elaboration modifies design; it refers to specificity of detail in design.

topics, the order of coverage, suggested tasks for each topic, and even scripts of teacher-student interactions. Each of these elaborations of the innovation offers users different information, detail, and guidance. More elaboration in this case offers users more resources, thus enabling — not constraining — them.

From one angle, extensive elaboration seems essential to illuminate an innovation's requirements for use, to alert designers and implementers to work to be done, and to reveal potential problems. Less-elaborated designs would not only be less useful but even self-defeating, for they tacitly delegate large amounts of invention to implementers, increasing the probability they would interpret interventions as versions of conventional practice, since the designs offer little guidance for anything else, and conventional practice is both familiar to and understood by implementers. Limiting instructional designs to general principles does not reveal much about their operation to designers or users. It is likely to increase the chances of confusion, the load on implementers to invent the design, and to yield generally weak and variable implementation. That, in turn, would be likely to erode the coherence and identity of innovations. From this perspective, less elaborated innovations constrain implementers more, for they limit their capacity for real change.

Other things being equal, more elaborated designs are likely to improve implementation by providing operating information about innovations. But other things are not equal, for more elaboration takes more time, forethought, and knowledge, and increases work, time, and cost. Designs that are thinly elaborated require less of innovators, and leave more to implementers' choices. Many educators may prefer such designs, for they intrude less. Despite their greater usefulness, more elaborated designs may discourage or turn away adopters. In addition, greater elaboration would almost certainly require more of designers as well, for elaboration depends on greater experience with and learning from implementing an innovation, in order to make sound decisions about what and how to elaborate. That implies a recursive process of innovation design and development, systematic formative research in the development process, and more time, human resources, and money. Extensive elaboration can turn innovation into a very complex enterprise, partly because it tends to expand the innovator's role, pushing it toward more involvement with and responsibility for implementation.

Another feature of innovation design that can influence the quality of implementation, by influencing the capability to learn new practices, is how well scaffolded they are. By scaffolding we mean the materials and social processes that can support, or scaffold learning. Innovations can only be implemented as they are apprehended and used by teachers and learners; all influences on instruction are mediated through their ideas, norms, and practices. The more innovations depart from conventional practice, the more new ideas, beliefs, norms and practices teachers and students would have to learn, and the more implementation would depend on that learning.

All innovations are scaffolded to some degree, for all include some means to acquaint implementers with new practices. These may include brief or extended professional education, parent education, examples of adoption processes, and video materials that depict the knowledge and skills required for use.[16] The extent of scaffolding varies greatly among innovators: some leave it to implementers to devise their own opportunities to learn, while others devise and offer extensive opportunities. Scaffolding can offer rich guidance with few constraints on invention and autonomy, or be quite structured and relatively constraining, with many combinations in between. An intellectually ambitious elementary mathematics curriculum could be scaffolded in several different ways, which parallel the options for elaboration. If it was elaborated only in terms of its objectives, scaffolding would likely be restricted to creating opportunities to learn about those objectives. If it was elaborated in terms of its objectives and the main mathematical themes of representing, reasoning, communicating, and solving problems, scaffolding could include opportunities to learn about the themes and how they compared with those in other mathematics curricula. If the innovation was elaborated in terms of its objectives, main mathematical themes, and types of activities, scaffolding could match this, adding opportunities to probe examples of the activities, objectives and themes.[17]

The first sort of scaffolding offers implementers modest opportunities to learn, and would impose only modest burdens on innovators; nearly everything would be left to implementers. The second offers slightly more opportunities to learn and comparably increased burden for innovators, and the third a bit more. Still, all three delegate most learning to implementers, and do not greatly burden innovators. Most teachers would not learn much, either about the innovation's aims and designs, or about how to extend and adapt the designs.

Other things being equal, more scaffolded designs would improve implementation by providing more opportunities for adopters to learn how to use the innovation. But other things are not equal, for more scaffolding takes more time, forethought, and money, and increases work, time, and costs. Extensive scaffolding would require extensive elaboration, and the combination would turn innovation into an even more complex enterprise, and place greater demands on designers. Greater scaffolding also would depend on greater experience with and learning from implementation, to inform decisions about what to scaffold, and how. That implies a recursive process of innovation development, more systematic formative research in the process of development, and more time, human resources, and money. Work of this sort would also revise the innovator's role, pushing it toward more involvement with and responsibility for implementation.

---

[16] Scaffolding is not the flip side of elaboration. For example, a direct instruction scheme could be extensively elaborated, but have thin and mechanistic materials for professionals to learn to use the approach.

[17] Elaboration with little scaffolding leaves it to implementers to learn how to bring the intervention to reality. Scaffolding with little elaboration provides implementers with resources for improvement without much guidance as to how they could be used or toward what ends.

At the same time, more scaffolded interventions complicate potential implementers' decisions about adoption, since they would make plain the need for extended relations with innovators or a third party, to teach and learn the intervention. Less scaffolded designs leave more to adopters interpretation and choice. Many educators prefer that, partly because it intrudes less. Hence, more scaffolded designs could discourage adopters, which could deter designers. Because they demand so much more capability and effort, more scaffolded designs also may not appeal to some designers.

This discussion reveals that our analysis of innovation design, use, and implementation stands in sharp contrast to several earlier approaches. To the extent that innovations require substantial elaboration and scaffolding, simpler notions of "diffusion" and "broadcasting" would not be helpful in understanding the requirements of scale. In Addition, the differences among innovations reveals a curious pattern: there is little or no descriptive evidence on the most rudimentary patterns of innovation, or how these might influence implementation, yet researchers have offered broad explanations for why innovation fails. How can such sweeping explanations be valid if we know rather little about the most elementary patterns of innovation?

ENVIRONMENTS in which innovation, adoption, and implementation occur also vary. Consider the difference between California, which has relatively weak state standards and accountability, and Texas, which has relatively strong standards and accountability. These and other features of the environment seem likely to bear on the formation, spread, and implementation of innovations, in part because they would be likely to shape the incentives for innovation design, adoption, and implementation. Yet we found only one study of what these features of the environment may be, or how they exert influence.[18] We do, however, know a few important things about the environments in which educational change occurs.

States, LEAs, schools and teachers work in a sector in which, until very recently, neither professionals nor schools or school systems could fail and go out of business, unless the student population vanished. Whatever incentives for change there were, they were not tied to the success of schools or professionals working in them; whether students learned or not had no influence on whether schools existed or teachers held their jobs. If there were incentives to innovate, they could not have been linked to the educational success of the organization or professionals in it.

Until very recently there also have been weak standards of professional performance in schools. The extant standards concerned teachers' course-taking, their experience, and their students' decorum. These things aside, there was no common basis for professionals to judge their academic work. There also was very weak knowledge about teaching and learning that was grounded in common academic standards. If there were any incentives or standards for innovation, they could not have been strongly and generally related to commonly held criteria for professional academic work.

[18] Grissmer, Flanagan, Kawata, and Williamson, 2000.

If practitioners wished to change teaching or learning, then, it was likely to be for personal and somewhat idiosyncratic reasons, or for local political reasons. It was not likely to occur because there were strong and broad professional or organizational incentives to change, or because there were strong and salient standards against which performance was judged and the need for change estimated. At the same time, the issue-attention cycle in education policy grew increasingly short in the last five decades; there were many more innovations, but many came and went quite quickly, as they often lacked the features necessary for implementation, and educators learned that they had few incentives to take most innovations seriously.

Until very recently, most authority for instruction in the U.S. was devolved to classroom teachers. State and local school agencies had very modest strength in teaching, learning, curriculum, and school improvement. Following Jefferson, Jackson, and classical liberal economic ideas, state governments were loath to build their own strength, and eager to delegate problem-solving to localities or the private sector. Local educational authorities also were reluctant to build central capability. State and local education agencies did not develop deep capability in the technical core, and were thinly staffed in those domains. For instance, when California began a very ambitious set of innovations in mathematics teaching in the middle and late 1980s, the state education agency employed fewer than three full-time professionals concerned with mathematics to monitor and support all the schools and districts in that huge state. Test and text publishers often have had more capability in curriculum than most state or local school agencies. Managers in schools and central offices, and governing boards took only very general responsibility for instruction, had little knowledge of it, and had no common criteria with which to review or change it. This offered classroom teachers less guidance than in most nations, and more room for personal idiosyncrasies with respect to both their practice and changes in it.

The education professions have not taken a deep interest in the quality of professional work. Teacher unions are politically influential on bread and butter issues, but until the AFT's recent history, no organizations that represented education professionals sought to define good practice, to investigate practice, or to improve practice. Or rather, their conception of practice improvement consisted of more spending and more professional education. The higher education institutions that offer professional education likewise took little or no responsibility for defining standards of practice, took little or no role in trying to improve it, and took a similarly weak role in regulating or informing practice with research. Researchers did studies, and some were informative, but neither professional organizations nor higher education institutions attempted to use research or professional knowledge to set standards for practice, to judge its quality, or to improve practice or education for practice.

Public education thus lacked anything approaching strong internal standards for the quality of practice, strong professional pressure to align practice with such standards, or substantial capability, in government or education, to evaluate or improve practice. The environments of public education offered little support for the implementation of innovations that made even moderately serious demands on practitioners. The innovations most likely to be widely adopted and sustained in these circumstances were those that made few demands on users, for the greater the demands, the less likely that users or innovators could find resources to support improvement in practice. The lack of connection between the political and professional environment, on the one hand, and educational innovation on the other, does not create resources for improvement, and thus impedes reform.

The preceding discussion deals mainly with the way that environments bear on innovation users, yet those same environments bear on innovators. As they developed new products, they had to choose. They could accommodate to the environment either by developing modest innovations that required little capability to be used effectively, or they could develop more ambitious innovations, that required appreciable capability, but delegate that problem to users or some other agents. The first course of action would increase the chances for adoption and effective use. Though the second might yield broad adoption, it would reduce the chances of effective use. The burdens of development would be greater in the second case, but neither would saddle developers with the task of designing and operating systems to scaffold use. Alternatively, innovators could challenge the environment by developing more ambitious innovations that required appreciable capability, and by designing and operating systems to scaffold use. Until very recently, there have been few cases of the third sort, at least in part because the tasks of development and support are so formidable.

CONJECTURES

Several conjectures follow from this analysis. They derive from research, but we offer them as suggestions for further research. First, under the circumstances sketched above, it would be quite unlikely for innovations that were devised in schools and classrooms to be adopted at any scale. Practitioners devise many innovations, but extensively devolved authority for instruction, the lack of common standards of practice, and the lack of incentives for substantial change inhibit their diffusion. Even if such innovations were well-designed, and many educators wanted to change, the incentives, communications networks, common standards of practice and vocabulary that would support diffusion, adoption, and use are absent. There is no social system that could turn practitioners' inventiveness to educational advantage.

Innovations that were more broadly sponsored – i.e., by state education agencies, commercial publishers, or national and regional professional groups -- would be more likely to be widely adopted. Judgments about the quality of use would depend on the nature of the innovation. For some, including commercial texts, fidelity in use seems a questionable concept, because there is little evidence that faithful use is intended. For many texts have been increasingly stuffed with material, on the assumption that teachers would pick and choose. For others like CSR designs, for which fidelity was intended, implementation with fidelity to the developers' design and intentions would be quite difficult, because devolved authority for instruction, the lack of common standards of practice and incentives for substantial change would discourage high quality use, even if innovations were well-designed, and educators wanted to change.

Second, one would expect innovation design to accommodate to the circumstances that we sketched above. One such accommodation is a broad menu of choices, from which adopters could select according to their inclinations; textbooks and other curriculum materials have developed in this direction, including greater and greater amounts of material. Another accommodation would be to define innovation in rather spare terms – i.e., as a set of "principles" – that offer implementers room to redefine them to suit their situations and inclinations. Several widely adopted innovations, including school restructuring and the Coalition of Essential Schools among others, took this path. Still another accommodation would be to tacitly delegate most or all responsibility for implementation to users, with little guidance about use. Many innovative curricula and several whole school designs (Core Knowledge is a leading candidate), have taken this path. Such accommodation tends to blur the nature of the innovation, making it difficult to identify what fidelity might mean. That would have survival value in public education, for it fits with the absence of standards of practice, and fragmented school governance.

Third, we expect innovative activity at every level of education, but typically sketchy implementation. Decentralized decision-making, the absence of much extended professional learning, and the lack of strong professions all offer educators great latitude to make their own innovations, and to make their own sense of others' innovations. The environment offers many points of access, there is little overarching direction, there are many weak priorities and little institutional memory. These conditions offer entrepreneurial innovators and educators many opportunities to design products that deal very partially with instruction, to do little to anticipate implementation, and thus, even when there is broad adoption, to expect variable and often weak use in practice.

At the same time (fourth), we expect that innovations that arise from social and cultural rather than educational changes could achieve broad adoption and use. For weak professionalism, popular governance and strong non-government agencies mean that schools are quite permeable to their environments. One example is the effects that prosperity and anti-poverty programs appear to have on the school achievement of Hispanic and African American students: there was a dramatic reduction of racial differences in average achievement in reading and math on NAEP, between the early middle 1970s and the mid-1980s.[19] Another example was the simultaneous spread of more relaxed classroom conduct and organization and the demise of strict discipline and formal recitation. These changes spread widely and rapidly between the 1920s and the early 1970s, beginning in primary schools between the world wars and spreading in secondary schools in the 1960s. The change arose partly from the growing popularity

[19] O'Day and Smith, 1993.

of child-centered family life, and a growing aversion to strict discipline, rigid feeding schedules, early toilet training, and the view that children should be seen and not heard. Broad social change outside schools provided most of the infrastructure needed for broad change in schools, as teachers and students arrived with the new values and behavior already in hand, in schools that had habits but few strong professional norms.

Although several researchers argue that innovations are more likely to succeed when they do not call for serious departures from teaching and learning.[20] But we have found several examples of substantial innovation in instruction which were adopted quickly and broadly. They suggest a fifth conjecture –- namely that instructional innovations which depart sharply from conventional practice can reach impressive scale, if they fit with the circumstances identified above. If innovations promise to solve problems that educators see as urgent, are relatively usable, and require only modest professional learning, they have a good chance of broad adoption and use. Graded elementary schools and graded texts were sharp departures from established practice in the 1880s and 1890s, but a flood of students made it increasingly difficult for teachers to manage with the existing non-graded materials and classes. These innovations helped schools to accommodate unprecedented waves of students, in schools that had meager funding. Schools and school systems could re-tool themselves for an age of batch processing.

Just a few decades later, standardized tests, elementary school ability grouping, and differentiated secondary curricula offered ways for the newly graded elementary schools and rapidly growing high schools to accommodate a continuing avalanche of students. They differentiated instruction, screened seemingly weak students out of many classes, and defended these changes with science. In the 1950s, secondary schools invented and adopted several demanding academic curricula and courses, in response to continued enrollment growth from working class and disadvantaged students, and rising pressure to improve academic content. The Soviet Sputniks provided a stimulus and a rationale for sweeping changes in secondary curricula.[21] These were conditions in which dramatic innovation in instruction could scale up rapidly.

These conjectures suggest a remarkably open and active system of schooling, in which substantial innovation in instruction can occur, either when external social change infiltrates schools or when innovations satisfy specific conditions in and around schools. A final conjecture, then, is that broad adoption and quality implementation of instructional innovations become increasingly less likely, as innovations depart further from conventional practice. The "normal" state of affairs has been frequent but modest innovation, accompanied by variable but generally weak implementation.

STRATEGIES FOR INNOVATION

The last twenty years have seen escalating ambitions for instruction: state standards-

---

[20] One of us included; see D.K. Cohen, 1987.
[21] Powell, Farrar, and Cohen, 1985, Chapter 5.

based reform, Goals 2000, Bill Clinton's Improving America's Schools Act, the Comprehensive School Reform Designs program, and George W. Bush's No Child Left Behind all urge higher academic standards and more school improvement. Whatever their differences and difficulties, these policies and programs press more aggressive and ambitious innovation on schools. Our analysis implies a few things about the prospects for such innovation, and how researchers might frame inquiry about them.

Public education presently is the scene of a collision between rapidly rising expectations for school performance on the one hand, and modest capability for the use of innovations on the other. Academically ambitious innovations are unlikely to succeed on any scale unless they deal with the sources of that collision. There are three general strategies that innovation designers and implementers could use to do that. One is to attempt to supply what is missing in the environment – i.e., to design and implement innovations so as to compensate for the problems sketched earlier in this paper. That would mean designing innovations that would: (a) offer powerful and continuing guidance for instruction, (b) compensate for weakness in the schools and professions; (c) support the formation of new professional norms, so that those at work in schools could not only orient their work to more ambitious standards, but could do so on their own; (d) include potent incentives, built into instructional guidance and professional norms, for more demanding and difficult teaching and learning, to increase the chances that the work would be engaging enough to bear the greater load of effort; and (e) reorganize instruction, school organization, and leadership to enable these changes. Innovations that included these elements would be more likely to be implemented by schools and professionals that have modest capability to improve instruction on their own. They also would hold more promise for use in environments in which only weak support for change from government and private agencies can be expected, and under conditions in which educational and political priorities often change rapidly. Such designs also would be more likely to cope with the lack of commonly accepted and understood standards concerning ambitious instruction, and in an inherited culture of innovation that offers potent incentives for quick and superficial action.

Since the best innovation design would be of little use if it could not be implemented widely and well, this approach probably would require implementation strategies that made it more likely that innovations could become self-sustaining, at scale. One element of any such strategy would be intensive, sustained support for implementation. Given limitations in the environment, the designer would be the most likely source of that support, but since that would place a considerable burden on designers, other sources of support probably would have to be devised. One possibility would be the construction, by designers and implementers, of new educational sub-systems. If they could be created and sustained, these could build social and professional solidarity, support the innovation's ambitious educational and professional values, and, by offering mutual aid, increase social capital. Such sub-systems would be situated in the larger school system and its environment, but would include means to both support improved professional performance and to buffer, or insulate, schools from that environment. One example might be networks of continuing technical and professional assistance, including tools, networks for the exchange of usable knowledge for practice, and trouble-shooting. Another might be digital archives of resources for practice. Still another might be new organizations, like some of the CSR designs, or networks of charter schools, that elaborate and scaffold responsible professional performance.

A second alternative would be for innovators and practitioners to change the environment, to make it less disruptive for ambitious efforts to improve instruction. The creation of state standards could be a step in this direction, and in some jurisdictions it seems to have had that effect. But even if the standards were exemplary, and written in ways that could be useful, they would be only a first step toward more coherent instructional guidance. It would remain to devise tests that were consistent with the standards and validly measured students' academic progress, to invent or adapt curriculum that could be turned into appropriate academic tasks, and to change professional education so that it focused on the standards, curriculum, and tests.

This is an incomplete account of the required changes, but it is an enormous set of tasks. It would require great change in federal, state and local government, in private organizations, and in public expectations for schools and government. The changes would be desirable, but innovators have not had much influence on them in the past, and we can see no reason to expect that to change in the near future.

That leaves a third alternative, which is for innovation designers to accommodate to environments. This implies designs that conflict little with environments, and have survival value in them. It also implies a sharp trade-off between detail of design and fidelity of implementation. In the environments of US schooling, more open and less elaborated designs are likely to have more survival value. These trade-offs grow more steep as innovation designs depart further from conventional practice. It is difficult to imagine that weakly elaborated designs for academically ambitious instruction would be either effectively or consistently implemented in very many schools, or, if they were implemented well, that they could be sustained for long. In the circumstances that we sketched in this paper, the designs that would best adapt to the environments of U.S. education are those that would be most likely to be weakly implemented.

THE SCALE OF SCALE

Most discussions of "scaling up" treat it as a quantitative problem, a matter of broad adoption and implementation.[22] Our analysis suggests instead that scale is as much a qualitative as a quantitative problem. All of the barriers to effective implementation that we discussed are qualitative. They include the difficulty of designing innovations that are usable for teachers who have modest professional knowledge and few common professional standards, the difficulty of addressing weaknesses of capability, and the

[22] A notable exception is H. Dickson Corbett and Bruce L. Wilson, 1998.

difficulty of devising means to manage the environment and support implementation. These are qualitative because they refer to the attributes of innovations. They do have quantitative implications, for these barriers could be much less substantial if innovators were working with ten implementers than if they were working with five hundred. To solve the problem of "scaling up" requires "scaling in" – by this we mean developing the designs and infrastructure needed to support effective use of an innovation. That, in turn, requires consideration of the problems that have made some sorts of innovation difficult, and taking these into account in deciding what to change, and how to design the means to do so. It also requires significant attention to designing the use of innovations by practitioners, in the environments in which they work.

Given the difficulty of doing such work, it is worth considering what "large scale", or "scale-up" might mean. Most accounts reference scale to the number of schools. From this angle, implementation in ten percent of schools in New York City would seem small. But if innovation designs and schools have been as weak, and environments as unruly, as we claim, effective implementation of even a modestly complex intervention in ten percent of New York City schools would be an enormous achievement. Scale is relative not just to the universe of possible implementers, but to the scope and depth of what must be done to devise and sustain change. The idea of scale thus means both less and more than has been imagined.

References

Bardach, Eugene. 1977. The implementation game: what happens after a bill becomes a law. Cambridge, Mass.: MIT Press.

Berman, P. and M. W. McLaughlin. 1979. Federal Programs Supporting Educational Change, Vol. VIII: Implementing and Sustaining Innovations. Santa Monica, Rand**:** 58.

Cohen, David K. 1987. "Educational technology, policy, and practice." Educational Evaluation and Policy Analysis, 9, (Summer): 153-170.

D.K. Cohen and James Spillane, "Policy and Practice: The Relations Between Governance and Instruction", . In G. Grant (Ed.), Review of Research in Education 18, 3-49. Washington, DC: American Educational Research Association.

Cohen, D.K., and Deborah L. Ball. 2003. "Instruction, Innovation, and the problem of scale." Unpublished draft, Ann Arbor.

Cohen, W.A., and Levinthal, D.A., "Absorptive Capacity: A New Perspective on Learning and Innovation, Administrative Science Quarterly, Vol 35, no. 1, March, 1990, pp. 128-152.

Corbett, Dickson H., and Bruce L. Wilson. 1998. "Scaling Within Rather Than Scaling Up: Implications From Students' Experiences in Reforming Urban Middle Schools." The Urban Review. Vol 30, No.4: 261 – 293.

Cuban, Larry. 1993. How teachers taught: constancy and change in American classrooms, 1890-1990. 2d ed. New York: Teachers College Press.

Friedman, Milton, with the assistance of Rose D. Friedman. 1962. Capitalism and freedom. Chicago: University of Chicago Press.

Fullan, Michael. 1982. The meaning of educational change. New York: Teachers College Press, Columbia University.

Grissmer, David W., Ann Flanagan, Jennifer H. Kawata, and Stephanie Williamson. 2000. Improving Student Achievement: What State NAEP Test Scores Tell Us. Santa Monica, Ca: RAND.

Hanushek, Eric A. 1989. "The Impact of differential expenditures on school performance." Educational Researcher (May): 45-62.

Lortie, Dan C. 1975. Schoolteacher; a sociological study. Chicago: University of Chicago Press.

Meyer, J and Rowan, B., "Institutionalized Organizations: Formal Structure As Myth And Ceremony". American Journal of Sociology, Vol 83, pp. 340-363, 1977.

O'Day, Jennifer, and Marshall S. Smith 1993. "Systemic school reform and educational opportunity." In S. H. Fuhrman (Ed.), Designing coherent education policy: Improving the system. San Francisco: Jossey Bass. 250–312

Powell, Arthur G., Eleanor Farrar, and David K. Cohen. The shopping mall high school: winners and losers in the educational marketplace. Boston: Houghton Mifflin.

Rogers, Everett M. 2003. Diffusion of innovations. 5th ed. New York: Free Press.

Rowan, B., "The Ecology of School Improvement: Notes on The School Improvement Industry in The United States", Journal of Educational Change, Vol 3, pp. 283-314, 2002.

Sarason, Seymour B. 1971. The culture of the school and the problem of change. Boston: Allyn and Bacon.

Stake, R.E. and Easley, Jr., J. A. 1987. Case studies in science education. Washington, DC: US Government Printing Office.

Welch, Wayne W. 1979. "Twenty Years of Science Curriculum Development: A Look Back." Review of Research in Education, 7. American Educational Research Association: 282-306.