
1

DISCUSSING EVALUATION IN A WORLD OF DISCURSIVE STANDARDS: ASSESSING THE NIST CENTERS

Charles F. Sabel

Charles F. Sabel is Ford International Professor of Social Science at the Massachusetts Institute of Technology. He is the author of *Work and Politics* (1982) and co-author with Michael Piore of *The Second Industrial Divide* (1984) as well as numerous articles.

In the shadow of vast efforts to reform national systems of criminal, justice, welfare, and health care by omnibus legislation, other, apparently more modest, projects to rebuild some of these and other institutions are passing almost unnoticed. The premise of these alternative projects is precisely that no one and no team, however expert, can know enough about the complexity of great matters to redefine the government's place in them in one fell swoop. The alternative projects proceed, therefore, incrementally by amalgamation and federation. Existing local entities federate to form regional bodies and these federate again to form a national structure. As the services each entity produces are shaped importantly by those who use them, the amalgamation and federation of entities at every level obliges the parties to decide which particular features of each will be incorporated into the new organization; the extension of the emerging federal system is thus accompanied and guided by a continuous discussion of what is working and what is not. By these means, it is hoped, experience and experiment will reveal possibilities and detect dangers that would have escaped the attention of even the best central designers and produce in the end comprehensive change without the defects of omnibus reform.

Of these projects the most mature, and the one built most deliberately in accordance with these principles is the Manufacturing Extension Partnership (MEP) which coordinates a national network of manufacturing technology centers (MTCs). The MEP is housed in the National Institute of Standards and Technology in the U.S. Department of Commerce. There are currently 35 such centers; their number should roughly double by 1997. The oldest has been in existence since 1989. Judging the performance of these centers is as good a way as any to begin assessing the promise of the alternative project of reform. Such review thus bears on an overarching question of public policy, and would be of pressing concern even if performance assessment of the centers were not mandated by the legislation that created them.

It is to the question of how best to evaluate the performance of the MTCs that this chapter is addressed. Its central argument is that the centers

individually and as a group have already begun to elaborate a system of self-evaluation, consistent with the principles of experimentalism and federalism on which they are founded; and the way to assess the performance of the system that best satisfies the requirement of public accountability while informing self improvement is to make those implicit systems the explicit starting point for performance comparison within and among centers, and between them and their users and governing bodies.

The body of the chapter is in four parts. The first argues that it is impossible to assess the effect of the work of the centers on business performance in any straightforward way because there is no straightforward measure of business performance. The second shows that firms themselves are nonetheless adept at developing methods of disciplined comparison that direct their own efforts to improve performance even in the absence of unequivocal performance metrics or comprehensive models of efficient organization. The third argues that mastery of these disciplines has, indeed, become a precondition for doing business at all in vast areas of the U.S. and world economies, as demonstrated by the explosive demand for certification under standards that measure competence to carry on disciplined discussions of production capacities. Helping firms master these disciplines, the argument continues, has therefore become one of the central activities of the MTCs. In providing this assistance they have begun to develop the capacity to apply these same disciplines to the assessment and improvement of their own work. And the last tries to show how, beginning with these beginnings, it would be possible to devise procedures for making implicit local self assessment an instrument of public evaluation that respected regional particularity while encouraging informative exchanges among centers and between them and their users on the one hand and the public authorities with responsibility for overseeing these programs on the other.

The Peculiarities of Measures of Business Performance

The MTCs were created to improve the performance of (small and medium-size) U.S. business. A straightforward way to measure their effect would thus be, at a minimum, to capture how they influence the firms with whom they work and how such changes affect the firms' performance. Sometimes, in fact, it is possible to proceed in this way. For example, the "products" of training programs are persons with more skills than they had before the new training. Absent distortions such as employer discrimination or prohibitively high costs of matching workers to the tasks for which they are best suited, skills will be accurately valued on labor markets. The way to assess the training program is thus to compare the earnings of the program's graduates with the earnings of persons like them in all respects but for participation in the training. In practice, of course, this turns out to be easier to say than to do; but that it is doable is demonstrated by the (depressingly) consistent finding that there is no significant return to participation in many publicly funded training programs.

But, it turns out, there is nothing straightforward at all about applying this conventional method to the evaluation of the MTCs. Indeed, I believe that it is close to impossible to do so, or at least almost impossible to do so without information that can only be generated by first using other methods of assessment.

The first obstacle to the conventional approach is that, unlike the training case, it is unclear just which unit's performance is supposed to be affected by the MTC's activity, and hence whose performance should be measured. From one point of view the relevant unit is the business function or process immediately addressed by the center project. Depending on the project, signs of progress would be reduction in scrap or inventory, quicker invoicing or the like. From another, the relevant unit is the regional economy as a whole in which the center is located. Signs of progress here might be better communication among the regional economic actors or an increase in good jobs. From yet a third

viewpoint the relevant unit would be the performance of the client firm, and progress would be progress by whatever measures measure firm performance. A moment's reflection will convince you, I think, that the crucial unit of performance is indeed the firm; but a second moment of reflection, I fear, will reveal that this conclusion is of little help because there are no generally applicable, comprehensive measures of firm performance.

The case for the centrality of the firm is simply that evidence of change at higher (regional) or lower (intra-firm) levels can only be interpreted by reference to the performance of business units (which may, of course be single firms or networks of companies). MTCs are not helping firms if they help them improve things that do not improve their performance. For that reason all the testimony to improvement routinely collected by MTCs (and offered by, generally, grateful clients) and then aggregated in the form of an estimated return on investment—dollars saved per year in inventory reduction per dollar spent on the corresponding project—may provide some coarse-grained indication of user satisfaction; but it says next to nothing about how well the MTC is doing its job. Unless we know the effect of the project on the operations of the firm as a whole and how well the firm is performing, we can not know how much credence to give the firm's evaluation of the project.

Starting at the supra-firm or regional level leads to a similar result. Suppose, to simplify, we agree that the most important thing a center can do for its home economy is increase the availability of good jobs; that we agree further that (independent of what employees may think about their situation) good jobs are those paying high (family sustaining) wages and increasing their incumbents' career security by encouraging acquisition of the skills needed for employability independent of any employer's fate; and that we know how to identify jobs of this sort. Now we face the real problem: How do we determine which interventions by centers will, with high probability, lead to such jobs or protect them where they are at risk? After all, some suppliers may be demanding wage cuts as part of an investment strategy that will eventually produce jobs that meet the goodness test; others

may be sweating workers and cutting other costs to maximize revenues before liquidation; and both may have worked with a center. Unless we can say something about the features of (potentially) successful business organizations, and how to measure them, we are going to have a hard time telling these cases, and members of many other confusing pairs, apart.

This inquiry brings me to the second, more disconcerting consideration regarding the ambiguities of business performance measures. There are, of course, many such measures. Some of these are financial: return on assets, return on sales, free cash flow, appreciation of stock, ratio of the market value of shares to book value, and on and on. Others are indicators of real performance: output per unit of input of labor or capital; inventory turns per year; scrap, defect or rebuild rates; market share, rate of new product introduction, new products as a share of all current products, and on and on.

The problem is that, in the aggregate, there is no systematic connection between good performance on any of these measures and good performance on the others. There is not even a correlation between measures within the same broad classes of financial or real indicators. More precisely, there is only a correlation among the measures when firms are doing very badly. The same findings, moreover, apply among the divisions of a single diversified large corporation as to the (U.S.) economy as a whole.

This result is less surprising than it may seem at first blush. A company on the verge of bankruptcy is likely to be incurring losses by many measures, and to have large inventories of unsold goods (many of which, however, may be newly developed). Successful companies or divisions, on the contrary, are likely to be succeeding at many different things. Some will be entering new markets or increasing market share with new products; others will be doing the same with more efficient production processes; still others may be using new combinations of both strategies to the same end. All of these differences will be compounded by differences in investment strategy and related differences in investment cycles that lead to further dispersion in measured behavior.

As a general proposition, moreover, uncorrelated performance metrics seem like a reasonable finding, given the extensive decentralization of authority within firms, and from them to their suppliers currently underway, and to which we will return in a moment. It only makes sense to decentralize goal-defining authority to subunits if those subunits are increasingly operating in diverse environments. Otherwise, it would be more efficient to measure them all by one standard—one designed to capture the goodness of fit between the unique environment and the correspondingly optimal strategy—and create an incentive scheme to encourage laggards to adopt the practices of leaders. But when the units are facing different circumstances the optimal behavior for one is not necessarily optimal for the others; hence the efficiency of ranking them by different measures. The effect of this diversification of measures would be just what we observe: All happy firms, to reverse Tolstoy's famous phrase about families, are happy in their own separate ways, whereas all unhappy firms are alike in their misery.

The upshot is that the straightforward method of assessing the centers will founder on the diversity of measures of good business performance. But it would plainly be wrong to draw from this conclusion the further one that firms have no way of assessing their own progress and that of others in a fashion that can guide further efforts at improvement. Indeed the profusion of measures of business performance has gone hand in hand with the articulation of an extraordinarily general language for the assessment of operating capacities—both one's own and that of potential partners—and it is to that language and its implication for the (self-) evaluation of MTCs we turn next.

Advancing by Comparison

Even in the absence of generally applicable performance measures or organizational models world wide, firms in the last two decades have become extraordinarily adept at improving their abilities to design and manufacture products, cooperate with each other in doing so, and even

assess their respective abilities to do what they say. In all this they rely on closely related techniques rooted in the idea of continuous disciplined discussion of comparisons between current operations and alternatives. The comparisons can be between current practice and best practice (benchmarking), between current performance and target performance (continuous improvement and just-in-time production), between the advantages and disadvantages of one design solution as against others (simultaneous engineering) or even between current performance and the expectations of a hypothetical customer (ISO 9000 and like standards). The comparisons are like discussions because parties agree on what they do know—just how good current performance is and how to improve it; which design will work best—and how they will proceed together to answer open questions. (Compare contracts, where each party assumes that the other knows from the start enough to fulfill its promises and can therefore act independently.) The discussions are continuous because the partners oblige themselves to compare progress to expectations regularly, and to revise or abandon goals as experience warrants. This they do because the discussions are disciplined in the sense that the results they produce matter: Production or development stops, with dire consequences for the parties if the jointly (re)defined goals are not met. In the extreme case of just-in-time production, where parts are, in effect, produced one at a time, this is literally true: Unless the cause of defects is eliminated upon discovery it is physically impossible to continue production; and thus improvement becomes a precondition for continued production.

Juxtaposition of the world of relatively stable international markets that prevailed roughly until 1975 with the fragmenting, fluctuating markets that have come to dominate the world economy thereafter will indicate why the new disciplines have come to be of decisive importance. Such a lapidary contrast inevitably suggests that the new model of organization is as consolidated and well understood as the old, and indeed has everywhere replaced it. I know you will not be misled. Although there is broad agreement on how the new differs from the

old, and some of the general features of the new model organizations that take account of these differences, there is no agreement even in the advanced countries on how whole corporations singly and in collaborative association should be constructed and monitored in conformity with these organizational innovations. Nor, *a fortiori*, is the change from the old to this indefinite new anywhere complete. Otherwise it would be hard to understand why firms in the U.S. and elsewhere still experience enormous difficulties adjusting to the new environment. Nonetheless, juxtaposition of the two worlds does reveal how the preconditions of business dealings have changed, and in ways that profoundly affect the operation of the centers and the possibilities for usefully measuring their performance.

In a world of stable markets and product designs, design costs were a small fraction of total production costs because they could be amortized over long production runs. Design times were a matter of indifference because, however long, they were only a moment in very long product life cycles. This same stability also induced firms to produce key components on dedicated or special purpose machines, whose cost could likewise be amortized over the long production runs. As such capital equipment was good for producing any one make or model of a final product, the firm that manufactured the final product had strong incentives to own it: An outside owner, after all, might threaten to withhold production of an indispensable part or pass knowledge of how to automate its production to a competitor unless the sale price was renegotiated in its favor. Hence the vertical integration of production widely remarked as characteristic of this period.

By the same logic relations with (generally small- and medium-sized) suppliers not owned by the end producer were regulated primarily by price. Such suppliers worked to detailed specifications provided by the customers. The quality of output could vary around the average requirements without substantially inconveniencing the customer because, with stable markets, the cost of maintaining inventory reserves against defects were thought to be insubstantial. Large customers had no reason to

prefer one potential supplier over another except for price—the cheaper the better.

In the world of fluctuating markets that prevails after 1975 design costs and times do matter, and in ways that undo the logic of special-purpose machines, vertical integration, coordination of supply through price, and apprenticeship through low-cost production. As demand fluctuates and differentiates, product life cycles become shorter and shorter and products become more and more complex and shifting combinations of rapidly changing technologies. Not even the largest firms can command the expertise and investment capital necessary to keep abreast of all the developments relevant to current and anticipated production needs.

The solution with regard to design is co-development or simultaneous engineering. Instead of designing products as integral wholes in-house, firms decompose them into discrete subsystems or modules, specify the performance requirements of each, and then entrust further elaboration and design to independent firms specializing in the respective lines of work. This economizes on development costs and times in several mutually reinforcing ways. First, the customer benefits from expertise the specialist has acquired solving related problems for other clients; and pays, in effect, a pro rated share of the costs of producing such knowledge, rather than assuming them in whole. Second, the concurrent efforts of all systems suppliers to master the tasks set them produces a steady flow of suggestions about how to modify the overall design or significant parts of it. This simultaneous problem solving directs attention to solutions that might have escaped notice if design work had proceeded sequentially from the most important component to the next and the next. By the same token it reduces total design time. This strategy is just a variant of the familiar idea that conception can often be expedited by execution, as in the low-budget film device of finishing an incomplete screen play by taking cues from each day's shooting.

Analogous changes reshape the organization of production. Because production runs are short, defects in design that affect manufacturability have

to be detected before production is running, not afterwards. This means that module or part designers must exchange information about potential difficulties and ways of eliminating them with the shop floor just as they would exchange information on other aspects of design among themselves and with the designers responsible for the product as a whole. Again because production runs are short, dedicated equipment is uneconomic. What is required instead are flexible machines or assembly set-ups that can cheaply and quickly be converted from production of one part or module to another. Finally, when production runs are short and product change-overs frequent the penalties for low or fluctuating quality levels become prohibitive. The cost of maintaining reserves against defects become intolerably high as modules proliferate and product life cycles shorten. To make matters worse, the chances and costs of missing a market entirely if supply of the product is disrupted by defective parts increase correspondingly. The solution is bufferless or just-in-time production, where parts are made in effect one at a time so that defects are immediately detected and their causes eliminated as a work piece passes from station to station, ideally without ever entering inventory.

These changes lead away from vertical integration as a principle of industrial organization and toward a system of collaborative manufacturing. In the fluctuating-world system we have been describing direct ownership of (most) design and production facilities by the end-product manufacturer is not a paying proposition. Firms want to reduce their pro-rated share of development costs by having module designers acquire expertise at the expense of other customers. The more that large customers want their design partners to diversify their clientele, the less their (design-related) motives for owning them outright. In any case, a design firm that is not dependent on any single customer is, de facto, an independent firm. Moreover, the less predictable the direction of technological development—and such unpredictability is, of course, among the premises of the fluctuating economic world—the less willing a final producer will be to make a long-term investment in a particular area of technical

expertise, diversification of the expert's market aside. Similarly, the switch from dedicated to flexible machines and from buffered to unbuffered production reduces the motives for designers to own (most) manufacturing facilities. Indeed, just as end producers want to benefit from economies of scope—the greater the range of products, the cheaper the cost of extending the range—acquired by module designers through diversification, so module designers want to benefit from economies of scope in manufacturing that their producers may gain in switching set-ups frequently to accumulate many and constantly changing clients. Under these conditions, as we will now see, prices become boundary conditions on cooperation and actual coordination is conducted in the disciplined language of a new type of discursive performance standard.

Discursive Standards

Given long production runs, we just saw, suppliers' bids to produce a given part at a particular price provide sufficient criterion for selection because the customer, who provides the design, can easily judge performance, and shortfalls within wide limits are in any case tolerable. But when designs are, in contrast, fixed only after exchanges among different design specialists and between them and manufacturing facilities, prices become boundary conditions for transactions, not a sufficient mechanism for selecting partners. Prices still matter because if the selling price of a particular class of final good exceeds certain limits, the good becomes effectively unsalable in its usual market, regardless of performance characteristics. To avoid this, firms in the fluctuating economy estimate the optimal sale price of planned products—the target price—and then set further sub target prices for individual modules in relation to the overall ceiling. Violation of these limits plainly puts the planned project at risk.

But just as plainly, offers to provide designs, modules, or components at or below the target prices are, by themselves, simply not informative enough to allow the customers to choose among competing suppliers. The bid alone tells the

customer only that the supplier can very probably provide a product that meets the initial minimum performance specifications at the proposed price. What the customers need to know in addition is how likely the supplier is to develop and realize more ambitious designs or more efficient production set ups in collaboration with others, and how timely and reliable its performance will be. As the bid is effectively a promise to do something the supplier has never, in that precise form, done before, and which the customer alone cannot do or even fully define, the bid, so long as it meets the target price, is meaningless unless accompanied by a reliable assessment of the supplier's capacity to do as promised.

Here is where new standards now sweeping through the world economy come in. Conventional standards specify and thus permit measurement of various combinations of two types of performance characteristics. The first are attributes of products such as the tensile strength of various grades of steel. Such standards allow buyers and sellers to know they are talking about the same thing when they conclude a transaction for a particular product. The Deutsche Industrie Normen provide countless examples. Other standards put greater emphasis on the attributes of production processes by specifying by what procedures a particular operation is to be conducted or how the results of particular operations are to be certified or both. Welds completed in compliance with such standards, for example, will have been executed using methods fixed according to the alloys concerned and the environment to which they will be exposed, then tested for conformity with requirements; and the results of these tests will be documented as further specified by the standards. Such, for example, are the standards maintained by the American Petroleum Institute; they assure not only that buyer and seller are referring to the same things in their negotiations, but that the buyer can also expect the seller to have diligently verified the accuracy of all representations and to be able to trace, with the help of obligatory documentation, the origin of defects that somehow escaped detection.

Judged against such conventional certification procedures, the new standards appear thin, even

vacuous. They specify neither the characteristics of products nor features of production processes—nor do they even prescribe methods for documenting the latter. What they do instead is establish procedures by which firms can be certified as competent to justify, and in that sense warranted to make claims regarding their capability to perform as promised.

Under the most comprehensive and widely acknowledged of these standards, the ISO 9000 series maintained by the International Office of Standards in Geneva, Switzerland, a production facility is certified as capable of designing and delivering products and services to customers' requirements in all areas of economic endeavor. To obtain certification a firm must convince a registrar (whose authority derives from association with the standard-maintaining body) that it can respond systematically to systematic challenges to its assertions of competence. Typically, therefore, the registrar, acting as a hypothetical client, will raise such questions as: How does the firm know that it understands its customers' expectations? How does it assure itself that it can deliver the designs and products it promises to provide in accordance with those expectations? How would the firm detect and correct shortfalls in performance? Answers are judged correct to the extent that they reflect both a good fit between firms' organization, technology, and procedures and the environment of expectations within which these operate, and an understanding by the firm of how it would have to adjust to variations in that environment. In theory, therefore, the ISO standards do not require firms to adopt any particular technology, form of organization or procedures at all, although in practice some ends can be obtained by such a limited range of means that to pursue them does entail particular operational choices. What certification under the standards does do is qualify a firm as a reliable interlocutor in eventual discussion about its performance capacities. For that reason I will refer to it and others of its kind as a discursive standard. As you will have already noticed, it is precisely such warranted information about performance in relation to promises that customers in the new, volatile economy need to supplement price bids in choosing collaborators. Certification does not, of course,

replace detailed discussion of projects and assessments of the chances of their realization; but it does establish that the collaborators are likely to understand particular assertions by the other because both can assume a common understanding of criteria by which such assertions can be justified. It is, rather, the language in which such discussions are conducted.

Under these circumstances it is not surprising that a large—and growing—share of the projects undertaken by centers with firms directly or indirectly concern mastery of the new disciplines or certification under discursive standards. On the one side, firms are more and more threatened by competitors in the U.S. and abroad who do command the new knowledge, and under growing pressure from their (large) customers to seek certification under a widely acknowledged discursive standard or the customers' own variants of it as a condition of doing business. On the other, they find that the market actors that might be expected to help with adjustment—consultants, trade associations and the large firms themselves—are too encumbered by their association with the old economic order to do so: consultants because their expertise is often in just the kind of centralized inventory, technology control, or management regimes that got the large U.S. firms into so much trouble in the first place; the trade associations because their expertise is often in lobbying for tax, regulatory or tariff changes, not in advising in reconstruction; and the large firms because they may be tempted try to shift the costs of adjustment to their suppliers by insisting that the latter make changes that they themselves do not, or may be unwilling to harmonize their standards with others to which their subcontractors are subject. There is therefore a place for public authorities such as the centers to facilitate adjustment by qualifying consultants, helping assess a particular company's need for reorganization and then designing projects to meet the needs, organizing exchanges of experience among firms that engaged in these activities, and orchestrating negotiations among groups of suppliers and their customers with regard to the harmonization of standards. Indeed, to judge by

numerous but hopelessly unsystematic conversations with center field engineers and managers such activities are now a central part of what these organizations do.

If this is correct then it follows that in their routine operations the centers are now amassing a vast store of knowledge about what works and what does not in the restructuring of U.S. industry. Some of this knowledge is purely local: which firms are willing and capable pioneers; which consultants are truly reliable; which firms might best learn from participation in a common group. But some knowledge is potentially of regional or supra-regional importance. For example, the design of co-op programs with engineering schools that permit students to practice what they have learned of the new disciplines in projects where firms actually apply them. In the best case, these projects create a better organization in the firm, a job for the students, and an enduring tie between the firm and their school. Another example is how to draw users into the review and definition of services provided. At the limit, yet other parts of this knowledge are more general still, and bear on the overarching questions such as which sequences of reconstruction are likely to help firms pursue which successful strategies in international competition.

Some of the generalizable knowledge held by particular MTCs, of course, is more effective than the same kind of knowledge accumulated in others and the latter can plainly learn from the former. Some of the potentially generalizable knowledge is simply wrong. Each MTC, in isolation, thinks it is on the verge of solving a problem that, viewed in the light of the others' experience, is beyond the reach of the methods they are using. All would be better off considering alternative solutions. In short, this looks like, and it is, a situation where continuously, disciplined comparison of operation could lead to improved performance—and quite possibly produce ways to measure improvement that would facilitate exchanges of information between the centers, the users, and oversight committees in the bargain. By way of conclusion I want to say how.

Learning to Talk Together without Saying the Same Things

By itself there is nothing novel or even helpfully precise about the idea of applying the principles of continuous improvement to the MTCs themselves. After all, as we just saw, in theory it is possible to use those techniques to get better at anything one wants to do: for instance, to encourage the use of robots, an end that, in isolation, many presumably find unappealing. The bite of my argument, if it has any at all, comes from the claim that under the new conditions of competition, and in the absence of consolidated models of industrial organization, the best thing to improve by continuous improvement is the ability of the MTCs to help firms master continuous improvement—which is now the minimum condition for participation in competition. More specifically, if the argument so far bears any weight, then the problem of assessing the MTCs, a variant of the classic end-of-the-regress problem, is to develop discursive standards by which to evaluate instruction in discursive standards.

The most effective ways to do this are best discovered by the MTCs and their users together; I offer these examples purely to illustrate the range of possibilities. The first concerns the field engineers. In their work with firms and discussions with each other they learn to distinguish promising projects—those where improvement is necessary and possible. Presumably a good deal of what they know finds expression in their own criteria for selecting and evaluating projects. Comparing those criteria across the branch offices of one center, with that center's incentive structure (do engineers' incentives match their convictions?) and with the results of similar internal comparisons in other centers seems like a natural way to render explicit and assessable the quiet routines that are already guiding activity in the established institutions. It would be equally natural to include users and groups of users in such discussions, and compare their experiences and expectations with those of the MTCs' staff. Again this would make explicit

exchanges and judgments that are already invited by the parties, but without the precision that permits and is permitted by deliberate comparison.

A second type of comparison could be directed at questions of regional economic organization. Centers have developed different ways of cooperating with technical schools, research institutes and banks. As the reference above to co-op programs suggests, some of these collaborations may already fit hand in glove with project work with firms; other activities may be pro forma demonstrations of good institutional citizenship. Assessing the fit between the efforts at regional cooperation and the work with firms within and across regions is thus indispensable to applying lessons learned in one sphere and place to what is learned in the others. Here too it is natural to include external partners—the centers' institutional “customers” and “suppliers” in these comparisons from the start.

Third would be efforts to draw out of all these discussions measures that connect with what the centers are doing and how firms are performing—or show that no such connection exists. This kind of work, which is already well underway at the Industrial Technology Institute in Michigan, could provide another way of reflecting and instigating discussion between the centers and users by systematically assessing the utility of the metrics of progress agreed between them, and thus eventually providing some summary measure of how well the parties understand each other and their situation. Obviously the only way to connect the measures that emerge in discussion is the ambit of the MTCs and any of the many conventional measures of progress or efficiency is to try connecting them, and then to divine from the effort what variants of each might yield more informative results.

Finally, let me return to the related questions of federalism and the diversity of performance measures introduced at the outset and left open until now. In discussing comparisons it may have seemed as though the goal were to create a single set of measures for all MTCs. But such an effort would fly in the face of the notion of experimental federalism as a spur to institutional learning that makes the centers and reform efforts like them all

reactive in the first place; to judge, moreover, by the diversity of measures of business performance measures, it would be pointless to aim for such a result.

Nor, given the features of disciplined comparisons, should anyone want to. The method works, in the end, by obliging the parties to agree on their motivations for joint actions, and revise those motivations when the actions disappoint expectations. So understood they amount to continuous discussion about improving the fit between strategy and environment. Anyone who can demonstrate such a fit without it being captured by a particular performance measure has a prima facie case against the use of that measure as an indicator of performance. Different performance measures coexist as we saw in the case of the best corporations. And the best way to justify such diversity is for each unit to present an independently compelling account of why things in its particular setting need to be done and measured differently.

In practice, discussion of various, potential discursive standards will always be just that: discussion of which local standards should be applied nationally, which locally, and which not at all. No MTC should have to show excellence by all measures; none should be mediocre on all, either. In a sense, and as in many of the best firms, evaluation will be as much a deliberation of the measures to be applied as an application of measures. In the end, then, as you will have long realized, the public authorities and local boards to which the MTCs are accountable, and whose powers of review discipline the centers' own, will make their judgments on the basis of the quality of those deliberations.

I believe enough in the unconfoundability of the method of continuous, disciplined comparison to trust that if the MTCs are doing good for the U.S. economy, those doing can demonstrate their abilities and ability to improve. I believe enough in democracy—aware of the limits of omnibus solutions, but also of needs for public action—to trust that if the centers and their partners do convince themselves of their worth, they will convince the citizens as well.

Evaluation in a World of Discursive Standards: Discussion

Shapira. There are two implications in following the path discussed in your paper. The first concerns the centers' abilities to adopt these different measures. Do they need to be organized and endowed differently? Second, you seem to imply that NIST should move out of business of requiring uniform quarterly reports and instead set standards and let the centers pursue fulfilling these standards the way they deem best. Given NIST's traditional standard-setting role, this role seems more appropriate than the traditional uniform evaluation approach.

Sabel. The assessment procedure should deeply respect the innovative process of these centers. These assessments should oblige discussion of goal-setting without imposing evaluation. It is important that NIST should not use the evaluation system to undermine the innovativeness of the program, the condition of its success. I agree absolutely that NIST be in the certification business. ISO, for example, requires you to respond to certain questions; it does not require you to do anything. NIST has to figure out what it wants to do, and the centers should find ways to fit into that mission.

Kaufman. I don't see NIST being a certification agency for manufacturing modernization programs. To be a certifier, NIST would have to release money based on performance. Then you would lose the responsiveness of the program. There can be minimum standards set, but for NIST to set the standards, by itself is wrong. Instead, I see NIST as a seed agent along with other stakeholders, such as the states.

Russell. I don't see NIST being a vicious imposer. I see NIST respecting different forms of industrial mixes, etc. NIST has almost been too permissive.

L. Martin. In 1988 we didn't know what we didn't know. Now, what we don't know is what we'll be measuring. We're floating along without a

vision. NIST's present vision doesn't have the sort of impact that will impress the public. I would charge NIST with developing a compelling vision, then measures would flow. The 1988 act was not sufficient.

Sabel. The difference in the way NIST centers are certified and projects are selected implies that learning is going on, but the people involved in the various NIST centers haven't learned what they need to know. At some point the NIST program will either degenerate into separate entities or centralize. How can you measure the results of separate entities? There is a huge amount of information about that. The test is whether or not you show that your strategy is the best. If you can justify that, you can justify diversity in the pursuit of different goals.

Oldsman. You are pushing the burden of evaluation onto the program. I agree with your calling for a "self-consciousness" about programs. We haven't analyzed and learned. You talk about experiments. We want a lot of experiments to occur so we can learn from them, but learning requires comparable data. That is a role for NIST which, in my view, is operating the program as a consortium.

Luria. Despite the fact that field engineers are a huge repository of knowledge, most centers keep so little information about their assistance efforts that we don't know what they did or how they helped make a firm successful.

Oldsman. That is what's lacking in the NIST reporting requirements--a center describing what it is doing to help firms.

Sabel. My point is that centers should be able to respond to that question, that is, what did you do to help a firm? There should be a debate about what information centers should capture. You shouldn't exclude the knowledgeable people, you should force them to describe what they did to help a firm.

Mendelowitz. Your presentation on discursive standards is compelling, but the notion of discursive standards is not sustainable as the program expands. Previously, the program was in a pilot phase, and there was considerable closeness between NIST and the centers, but now you are

moving from a small system that can be easily managed in a direct manner, to a large system. Management of large systems is typically by quantitative indicators, however imperfect those indicators may be. You have to use those quantitative measures, even though everyone knows that what is really important is qualitative. That's the unfortunate reality of the world.

Osterman. Your argument about evaluation was powerful in the analogy between the NIST system and firm. The NIST system should evaluate itself much the way a firm tries to learn about itself and thereby improve itself. A firm ultimately has to sell products to a customer, and NIST has a customer too; that is, NIST's evaluation ultimately must satisfy Congress which will want to know if the program is creating more quality jobs. This type of evaluation is not as helpful internally, but it would be helpful to Congress.

Mendelowitz. This issue of jobs is unclear. Should the programs' desired outcome be jobs that pay more, or simply more jobs? Quality jobs may be important, however, Congress is also concerned about a program's impact on the number of jobs.

Redman. I haven't given up the ghost on performance measures. State governments are more concerned with jobs than is true at the federal level. If we don't make an effort to develop those performance measures, we stand in political peril.

Sabel. What would be terrific is if that process led to a redefinition of what Congress wants, or how Congress measures things. For the NIST centers, in the case of jobs, there is no clear idea of what the causal mechanism is. It would be very bad to have the institution, in a delicate moment of its growth and in a delicate economic condition, side-tracked by something that is not its mission.

Pounds. If our customer is not really the small and medium sized establishment (SME), but Congress, then we need to change the mission and say it is to create jobs. Our mission now is improving manufacturer competitiveness. We have to supply two sets of books to our stakeholders, because each stakeholder has a different mission, as well as do our work to help our customers, our businesses.

Sabel. If you are keeping two sets of books, and you know what's in them, that's okay. But if you're keeping two sets of books and you don't know what you are doing in either--such as what service you provided a firm--you're in trouble.