Implications of the OECD Comparative Study of Performance Standards for Educational Reform in the United States

CSE Technical Report 419

Eva L. Baker
National Center for Research on Evaluation,
Standards, and Student Testing (CRESST),
University of California, Los Angeles

November 1996

National Center for Research on Evaluation,
Standards, and Student Testing (CRESST)
Graduate School of Education & Information Studies
University of California, Los Angeles
Los Angeles, CA 90095-1522
(310) 206-1532

The work reported herein was supported in part under the Educational Research and Development Center Program cooperative agreement R117G10027 and CFDA Catalog number 84.117G as administered by the Office of Educational Research and Improvement, U.S. Department of Education.

The findings and opinions expressed in this report do not reflect the position or policies of the Office of Educational Research and Improvement or the U.S. Department of Education.

IMPLICATIONS OF THE OECD COMPARATIVE STUDY OF PERFORMANCE STANDARDS FOR EDUCATIONAL REFORM IN THE UNITED STATES

Eva L. Baker
National Center for Research on Evaluation, Standards, and Student Testing (CRESST)

The purpose of this paper is to explore the implications for educational reform in the United States of the OECD comparative study of performance standards. To provide some context for the reader, the paper briefly reprises major shifts in the intellectual underpinnings of U.S. educational reform and reports on its present prospects. The major analytical section, however, posits a functional model of reform that includes key elements thought to be necessary for successful operation of a dynamic educational system. As a special case, performance standards will be characterized as they support different elements and functions of reform as described in the country case studies provided by OECD scholars, with particular emphasis on the U.S. setting. Finally, a discussion will consider the ways the U.S. and other governments could profit from the OECD case studies of performance standards.

U.S. Background

Educational reform, for reasons of constitutional authority, tradition and continuing predisposition, arises from three levels of authority: local, state, and national sources. Unlike the systems of the majority of OECD countries, in the U.S. the national (federal) authority possesses relatively weak responsibility for the design and virtually no responsibility for the actual operation of educational systems. It is clear as well that in the last 20 years, state governments have demonstrated increased leadership in the area of educational policy, particularly on matters of school finance, teacher certification, and the expectations for school programs. Major changes to fiscal policies have resulted in states bearing an increasing share for the support of schools. As a consequence, states have been aggressive in their pursuit of policy revisions. A number of specific policies have been addressed to educational reform including the development or substantial modification of curriculum frameworks to guide desired student attainments, the

introduction of goals related to success in the workplace, the development of assessment systems at the state level to monitor or to certify student achievement in academic and work skills areas, and the creation of explicit rules for the adoption of texts and other instructional materials purchased from the commercial sector. While these policy areas are traditional features of many OECD country systems, attention to these areas in the U.S. at the state level has led to an attempt to consolidate at state centers new and stronger educational authority and represents a net shift from local control to more centralized authority over education. Given the size of some states, for instance 30 million or more citizens in California alone, policy reform of this scope is easily analogous to that of some OECD nations.

Attributes of Educational Reform

Although thoroughly discussed in the individual country reports, it is worth revisiting at a summative level the model of educational reform that appears to have international currency. First, there is a dual emphasis on improving educational quality and improving access to educational services for increasingly diverse student populations, goals that result in the well-known North American tension between excellence and equity. Second, there is a sense that educational practices and outcomes must be increasingly relevant to the demands of the societies in which they are embedded as well as to expectations to maintain international economic competitiveness. Third, the fitful history of educational reform has induced the view that educational change should be systemic; that is, it must consider in parallel key parts of the endeavor so that changes in one area will support goals and processes of other areas. For example, curriculum change should be supported by concurrent adjustments in teacher preparation. Fourth, education must be a cost-sensitive enterprise, and expenditures must be carefully justified in the light of their impact. Consequently, greater interest has been expressed in the accountability of educational systems, the ways in which they demonstrate their impact, and the ways in which relatively inexpensive sources of information can be used to improve them.

These components of educational reform are not without their critics, inside and external to the educational field. Critics from within the field note with some alarm the process of centralization and quantification, and the influence of economic metaphors in educational policy formation. There is concern that the

idea of systemic reform somehow equates to Utopian notions of education, notions bound to fail because they emphasize idealized and uniform views. Others, from outside education, are skeptical about systemic educational models, particularly those embodying new views of learning, teaching, and assessment. Failure to do a good job using old methods does not automatically build confidence in attempts to meet more complex and ambitious goals. Some worry that the moving target of educational reform is simply a device to avoid real accountability that comes with a stable system.

A Functional Model of Educational Reform in Multinational Use

The following model is offered as an overview of educational reform applicable in various countries (Figure 1). The framework is similar to those previously employed to represent descriptive models such as educational indicators. Such models rely on categories of inputs, processes, and outputs and posit a loose, causal relationship. The reform model differs from indicator models in a major way, for input, process, and output categories serve to classify functions served by reform in various countries rather than as a structure to subsume data sources. Each of the categories could be expanded almost exponentially. For example, the contextual inputs could be greatly augmented to provide details about traditions and expectations. For purposes of this analysis, these inputs

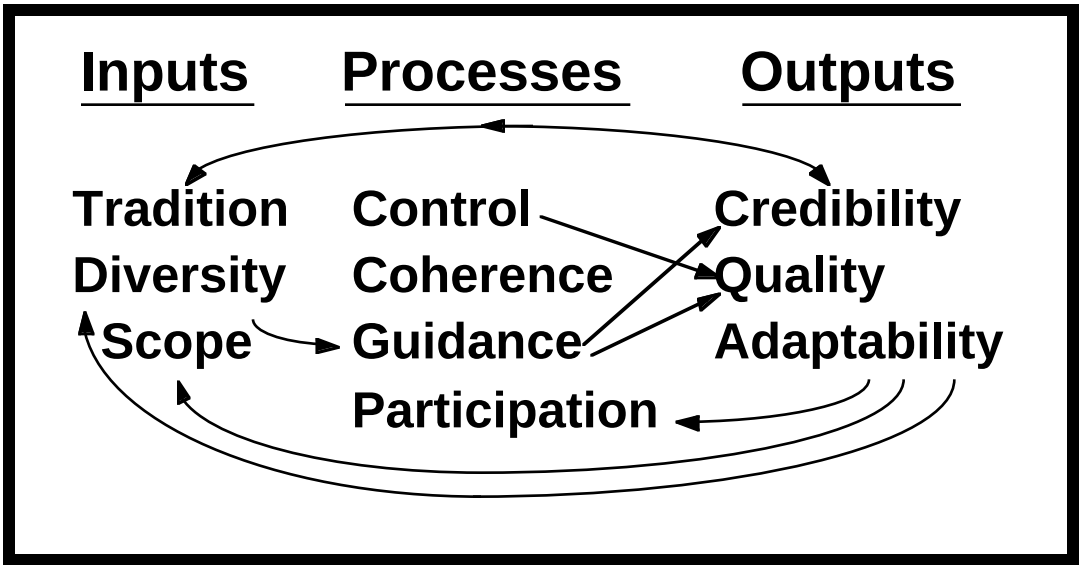

Figure 1. A functional model of educational reform in multinational use.

are presented at their present high level of abstraction so that international audiences may apply them to the relevant array of societal conditions. This general model will be explicated in the light of the problem of setting useful performance standards for educational attainment. It could be similarly applied to other educational issues, such as governance, teacher development, and curriculum development.

Performance Standards

In the U.S., the use of the term "performance standards" is subject to wide interpretation. Uses in formalized procedures of instructional development and training, for instance, in the military or in business and industry, signify the level of attainment groups of individuals need to meet. Most commonly these have been expressed as percentages of correct scores on a pool of discrete test items, although in some versions they include the goal statement of how many individuals, for instance, 80%, would need to meet the desired percentage of achievement. In other cases, where responses are more qualitative, performance standards have been set by specifying attributes that the criterion example would have to meet, such as "no spelling errors" in an essay. In the present iteration of the interpretation of the term, the use of proficiency levels or cut-scores continues for some to refer to criteria used to classify students as very successful, adequate, or unsuccessful. For other users, however, the term performance standards is used to signify the behavior students are expected to exhibit to demonstrate their achievement. Some educators consider performance standards operationalized by examples of assessment tasks students complete, such as solving problems with particular characteristics, or interpreting literature of certain types. In the international community, much broader interpretations of performance standards are made. For some, they are the descriptions of the expectations of the system, not only including what levels of attainment will be reached but related to broad distributions of educational access, for example, 10% of students will matriculate to the highest levels of postsecondary education.

In the United States, definitions of performance standards remain relatively concrete. A key question in the U.S. relates to the validity of these performance standards, however expressed, insofar as they communicate adequately the desired end-states of students at varying curricular points. This question is not surprising in the light of the quantitative and research orientation of U.S. academics.

A variety of approaches are used to establish formal, public performance standards. One common strategy to determine levels of achievement desired involves the application of judgment. Professional judgment of teachers, subject matter experts, or other academic or school-based constituencies brought together for this particular purpose is perhaps the most frequent approach, varying in details of group composition, the level of abstraction of the task, types of directions or data provided, and guidance by authority. In recent years, these standards-setting groups frequently have been expanded to include various representatives of the public, such as persons from business and other non-educational sectors, as a tactic to broaden the base of judgment and to build constituency support for the resulting product.

There have been other, more technical approaches used to augment judgmental strategies for setting performance standards. One such technique depends upon estimating the number of classifications to which students will be assigned and then predicting the likelihood of misclassifications. False-positive or false-negative probabilities can be estimated, and policy can be set in the light of which kind of error is seen to be more grievous. Another approach is to take various samples of student performance, treat them as replicates, and examine stability of classifications. Classification error obviously depends upon the quality of the measures used to generate scores as well as the levels selected for boundaries between categories.

At the heart of the discussion of performance standards is what they might be used for. Clearly there is global intent that they guide the development of the measurement system and the instruction designed to develop desired student accomplishment. There is also the desire to use them to set standards for individuals (in the case of certification of performance by examination) and institutions (in the case of determining whether schools are making adequate progress in educational reform). Just as analysts have questioned whether a single test or set of measures can be aptly used for a variety of purposes, there remains the question of whether publicly specified performance levels can simultaneously serve multiple purposes with equal validity.

Applying the Reform Model to Performance Standards

Let's consider elements in the model as they play out in the development and character of performance standards (see Figure 1). Notice that the relationships

as indicated by directional arrows are complex, and their strength will differ from country to country. In considering the developments in the U.S. with respect to the model, it will be shown that the focus on performance standards is undertaken to strengthen system processes of control, coherence, guidance, and participation, as well as to set clearer boundaries for the measurement quality of educational attainments.

Tradition

Tradition refers to the cultural values of and operational expectations for the education system. It includes the extent to which education is perceived as a primary value in the general society and how that value is distributed among sectors in the society. More refined analyses would reveal whether the value held for education is attributed to its contribution to a strong economy, to the development of individual autonomy and flexibility, or to the development of an academic elite, among other options. A second factor influenced by tradition is the perceived effectiveness of the educational system. Such perceptions are influenced by the extent to which various sectors of society assume the competence to criticize the system of education, or at least certain of its components. In the U.S., for example, it has been commonplace to criticize the productivity and impact of the precollegiate educational system, but to trumpet the quality and effects of the network of colleges and universities. This tendency is connected to another feature of national educational tradition, that is, the level of trust accorded classroom teachers. The interaction of teachers' reputations as acknowledged educational experts contributes to and is formed by the value and reputational status of the educational enterprise as a whole. In the recent past, U.S. teachers have held relatively low professional status. The quality of their training has not been highly regarded, the selectivity of the system has been reputed to be relatively low, and thus dependence upon teacher judgment as a strongly valued indicator of student attainment has dropped precipitously. The category of tradition in this model also subsumes historically important characteristics of governance, structure, and operations in educational systems—in effect, where the balance is struck between centralized and devolved responsibility, how professional and lay responsibilities interact, and how the quality of educational outcomes is typically assessed. Traditions in governance in the U.S., with the assumption of educational control sited at the local level, support the lack of codified or otherwise explicit educational standards. It is only

the move to more uniformity, articulated as the need for comparisons, national standards, and national tests, that has required the explicit statement of performance standards at the state level.

Diversity

Diversity refers simultaneously to a number of important areas. First, we must consider the range or uniformity of expectations for the schools, both in view of environments for teaching and in the light of the student competencies they intend to develop. Locally controlled systems have promulgated somewhat different educational expectations. It is also clear that systems that predominantly serve identifiable subsets of the student population many times develop different expectations and standards for those students. Systems such as those in the U.S. deal with diversity in every respect: the size of the geographical area served by the educational system; the density of population; the socioeconomic experiences of the students; the average residential period in the service area; the languages spoken in the home; the diversity (or lack thereof) of the teaching staff in terms of cultural and language background, preparation, and age; the level of financial support for the schools; the type and frequency of parental engagement in education.

A correlated issue relative to diversity is the degree to which the education system formally acknowledges its existence. Acknowledgment can take the form of the recognition and redress of differential access to educational opportunities and the nature of efforts made to extend educational options to various constituencies. To be considered here are whether accommodations and adaptations for individuals and groups are in place, what the nature is of such extensions, and, importantly, whether these adjustments are perceived as equitable. In the U.S., the public discourse has mushroomed on the topic of accommodations in both education and workplace for groups heretofore designated as disadvantaged. For instance, arguments have been strongly put forth to attenuate the compensatory assistance given to minority groups by educational systems in higher education admission decisions. While the impetus of these accommodations has been to increase fairness in the system, the racial or ethnic bases of them are criticized by sectors of the society arguing that the remedies introduce new forms of inequity. Alternatives proposed include accommodations based on economic disadvantage, an approach likely to be criticized as well. Here questions about diversity intertwine with views about uniformity. Are

performance standards intended to be identical for all students, regardless of background, motivation, and capability? Are students permitted flexibility, for instance, in the selection of areas in which they wish to become especially accomplished? Are the minimums or core areas the same for all? Are differential interpretations of performance of individuals or schools permissible? For example, if statistical adjustments are made to account for disadvantage in educational background, prior attainment, resources, or socioeconomic status, are performance standards providing the level of guidance desired? Will they be interpreted appropriately? Do they foster equity? No complete answers exist for these questions, which go to the heart of discussions of fairness.

Scope

A third category of input or contextual issue involves the scope of the education system. Scope embraces the concepts of the range of goals served by the system, its breadth and ambition. This analysis involves the elaboration of the types of educational institutions available, the ambition and extensiveness of their curriculum offerings, the extent to which priorities are evinced in the design and operation of school programs and educational institutions, and the collaboration—either active or de facto—of other public and private entities with the formal educational system in meeting its needs. Scope is closely related to resources and the extent to which educational systems can be said to be well resourced given the range of action they undertake.

The Processes of Educational Reform: Performance Standards as Strategy

Educational reform operates to support systemic functions. The functions identified include how the system operations and performance are controlled, how coherence and focus are supported, how guidance is shared, and the degree to which participation from various constituencies is desirable and productive. In this analysis, it is clear that performance standards may play important roles in these functions, a point illustrated in the U.S. case.

Control

In the educational systems of many OECD countries, centralized control allows for systematic and relatively rapid change. Control is formulated as the set of procedures, requirements, and sanctions used to manage the direction of

systems. When authority is centralized, it is relatively easy to promulgate new regulations (although differences in compliance are bound to exist). Centralized expectations about curriculum goals, about choice of instructional materials, about teacher selection, training, certification, and employment may be made clear. Variations in control include the number of recognized levels of authority and strategies to maintain the focus of their attention and response. A second dimension of control is the locus of initiation for activity. Are there mechanisms for initiation at other than top levels? Is such initiation an expected or unusual event? Control also involves the allocation of rewards and sanctions. Their assignment depends upon the clarity of understanding among system personnel about consequences of action, as well as the distribution between positive and negative sanctions invoked. Other variables related to the perception of control are the intensity and interval of consequences, and the degree to which there is explicit linkage of system performance with sanctions and rewards. Control may be located in a number of places: in government through legislation or regulation; in professional societies related to standards of state-of-the-art practice and ethics; or, more informally, in the pressure exerted by peers or community to adhere to particular standards of behavior. One clear approach to control involves the application of identified quality control mechanisms. Countries with exit examinations possess a mechanism to control educational offerings, through the creation of boards of studies and examination councils and reports of results of student examinations.

In countries with decentralized educational authorities, such as the U.S., the mechanisms existing for control may also reside in the perceived responsiveness of systems to local needs, for example, through local elections of school governance groups. Control of quality has also been typically exerted through regulation and standards of practice. Regulations may specify how much time students must engage in particular subjects, the number of courses in teacher preparation sequences, requirements to teach health and life skills in addition to standard curriculum, and rules about offering instruction in various languages. Standards of practice shift so that that teachers in the 1950s might have been judged on their lecture style, in the 1960s on their ability to deliver up-to-date subject matters and classroom management, in the 1970s on the quality of their behavioral objectives for students, in the 1980s on their ability to group children heterogeneously, and in the 1990s on their ability to teach interdisciplinary topics.

In other words, although wise practitioners and academics decry the practice, there remains the search for and temporary fixation on the "best" approach to a particular area—a quest sure to fail within each country given the individual strengths and weakness of teachers and students. As a result, techniques or methods have been promoted as the true solution to educational ills in a particular discipline or for a type of student. This focus on process has not worked because standards of practice are subject to unstable beliefs. These views are derived from a variety of sources, including experimentation, extrapolation, research and development, and novelty. This instability, borne of the relatively fleeting consensus on desirable strategies and the continuing importance of local autonomy in the U.S., has resulted in educational systems that lurch sporadically from solution to solution. As a result, a focus on process control is almost always out of date. More recently, joining most OECD countries, the U.S. has begun to strongly focus on educational attainment. In state after state in the U.S., numbers of regulations specifying classroom requirements are being reduced, with the assumption that increased freedom of action will support the accomplishment of desired goals. Nonetheless, the specter of process control remains embedded in the nature of certain curriculum goals and in the performance standards set for them. For example, in the state of California, the review of language arts goals shows that it is difficult to disentangle the support of a "whole-language curriculum" from the more general desire to have children read with comprehension and write lucidly. The very manner in which performance standards are phrased communicates preferences for educational methods, preferences that may or may not be very well justified. Thus, even the use of performance standards to transfer attention away from specification of educational process to accountable performance becomes corrupted by the infusion in their formulation of the educational methods of the moment.

Coherence

A second general and desirable feature of reform is its coherence. Reform efforts should mutually support and relate to one another in a logical way. At the broadest level, for example, teacher preparation programs should connect and support particular curriculum innovations. Indeed, the principle is even true when one looks at the system from the child's point of view. For instance, it could be argued that children should be given instruction in science that is compatible with overall approaches to problem solving to be engendered by the educational

system. Otherwise, children may have difficulty in sorting through when to use various approaches. Another formulation of coherence can be drawn from the extent to which students in the educational system share in common experiences that would provide policy makers and teachers with a clear understanding of the order and nature of their learning. In many OECD countries, national curriculum statements and monitoring present one approach to the creative preservation of coherence in education.

In decentralized systems such as those in the U.S., sources of coherence have been weak, for the most part. Because curriculum statements have not had much functional power, sources of coherence are almost always indirect. One source has been the expected experiences of teachers in preparation in higher education. Institutions of higher education develop their own teacher preparation curricula, and even the same course at the same institution may have vastly different content when taught by a different professor. So it is chimerical to think that coherence would be a product of teachers' postsecondary experience. Another potential source of coherence of experiences for students occurs through their examination processes. But because no common exit examinations exist in the U.S., this is only a potential outcome. University admissions tests, for the most part, emphasize general verbal, quantitative, and analytic ability as opposed to the mastery of particular domains of knowledge and do not provide sufficient guidance for curriculum design.

In the U.S., system coherence seems to have come in the past from three major and disjunct sources. First is the promulgation of course requirements for entry into colleges and universities. Although these differ at the margin, they in general emphasize experiences in literature, higher mathematics, laboratory science, and, to a lesser extent, foreign language. Their impact might be thought to be limited to only the college bound student, but in fact, the impact of academic standards has at least superficially affected most educational systems, without regard to their principal clients. A second source of coherence relates to the materials used in classrooms. Relatively few commercial publishers provide most of the textbooks and adjunct instructional materials for students. The design of these materials is market-driven and influenced by preferences, specifications, and adoptions of relatively few large states, as publishers are unable to adapt materials profitably for smaller groups of users. Some states and local districts provide a list of options for schools to select among, undermining in the name of

adaptation and local autonomy, common experiences derived from common materials. It is also a reasonable speculation that the reliance on technology-delivered instruction will permit adaptation to local requirements not heretofore possible because of cost. A recent and encouraging development is the expectation to evaluate instructional materials, texts, or computer-based programs in terms of the extent to which they support explicit statements of goals or standards.

It is obvious that the mere statement of performance standards does nothing in itself to ensure reflexively their coherence, particularly from the viewpoint of an individual child. In the early stages of the development of standards, it appears that it is hard to keep raging educational ambition in check, and content and performance standards have proliferated beyond that which any individual child could learn, no matter how gifted or industrious. Too many competitive, discrete educational topics are deemed essential. Worse, no strategy for sorting through and balancing desires with feasibility has emerged. This profusion of options has created another opportunity for incoherence in that, as a practical matter, public policy groups must pick and choose from among desirable outcomes those that they believe the system can foster. Often this selection is more of a political bargain than an intellectual enterprise, with topics and goals traded off in one subject for decision-making prerogative in another. Nonetheless, because it is still the case that performance standards are regarded as public documents, their explication makes it possible to note gross inconsistencies and gaps, and to create, over time, better formulations of goals and intentions. There is also the persistent and unanswerable question about the optimal sort of coherence to be desired in any educational system.

Guidance

If control is formulated as the means of managing the direction of a system, and coherence is conceived as the character of desired interrelationships and supportive opportunities, guidance consists of the form and types of information provided that are needed to generate willing compliance by participants. It is a softer, gentler form of control. Guidance provides cues for translating requirements into procedures and actions and usually permits some interpretive latitude. Curriculum handbooks, teacher manuals, and teacher development workshops are all traditional forms of providing guidance to school practitioners about the goals and approaches desirable for educational improvement. In certain countries, the use of inspectorates or quality assurance groups provides guidance

relevant either to particular purposes or programs, or in the general direction of improvement. Based on site visits and face-to-face meetings with skilled professionals, inspections may have a technical assistance and collaborative character. On the other hand, they may verge toward the control side of the model when they are seen as accreditation events.

Guidance can be inferred from observing the workings of another educational setting or system. In relatively informal ways, systems can perceive directions, strategies, and solutions to local problems. The opportunity systems provide for these outward looks, even if the field of vision is relatively restrictive, can give an index of the extent to which guided innovation is valued.

Performance standards can serve as another source of assistance for teachers and curriculum developers, provided the standards are conceived and conveyed in an appropriate fashion. One major way performance standards can help teachers is to clarify the order and nature of expectations for different-aged learners or for same-aged students at different attainment levels. If performance standards in a particular area, such as written composition, have both increasing task complexity and increasingly rigorous standards for use in judging the quality of student responses, because of either student age or the achievement level (e.g., proficient compared to advanced), then we can expect the stated progression to have clear instructional implications for students. Secondly, the details of the standards, particularly the enunciation of essential quality criteria, provide particular cues on how instruction itself should be organized. For instance, if students' writing ability is to be judged in part by their reliance on concrete illustrations of general points, then teachers would be guided to provide subtasks in instruction that involve the identification and aptness of such illustrations. One of the problems with performance standards, of course, is that they may be written in too general a form to provide specific instructional cues; on the other hand, they may be equally useless if they are expressed solely in quantitative terms, such as the percentage of children who scored at a particular scale value on a test. One potential point of difficulty is finding a way to assure that the description of the performance standard, the verbal intent, matches well with the actual test or assessment given to students, the ways students' answers are judged, and the analysis and reporting approaches adopted. At any of the points it is possible to move, perhaps only subtly, to results that actually have a very

different operational meaning than what was articulated in the performance standards statements.

Participation

A fourth characteristic of educational reform processes that differentiates OECD countries is in the area of participation in educational decision making. Participation differs in countries in terms of the involvement of constituencies, the relative dominance of particular groups, the timing and type of engagement, and lines of communication provided to support the transmission of and reaction to ideas.

The degree to which participation is seen to be desirable is influenced strongly by traditions of responsibility and authority, general satisfaction with the system's effectiveness and scope, and the diversity of the publics served by educational programs. Constituency participation in setting performance standards is also influenced by the technical character of the approach taken. For the most part, discussions of performance standards and their development have involved various members of the education constituency: teachers, subject matter experts, administrators, policy makers, and parents. In most cases, representatives of the public at large and specifically business and industry have been included. Where standards have been developed for workforce expectations, the distribution has been appropriately modified. Other groups involved in these processes have been selected because they typify important classes of student interests, for example, those with various ethnic, language, socioeconomic, gender, and racial backgrounds.

The differential dominance of these groups changes with time and with perceptions of competence and power. Teachers, although usually providing major membership to standards-setting exercises, have in general been less prominent in recent efforts, giving way to members of the public and the business sectors.

Participation ranges from functional to symbolic; both ends of the distribution have utility in moving educational agenda forward. One way to determine the extent to which participation seriously informs outcomes is to study the type of engagement and timing of participatory activity. Some groups initiate, design, review, decide, and ratify. Others provide only a few functions, and the differences in types of engagement depend upon the extent to which authority is assigned to the participating groups. Furthermore, the point of entry into the

process and the amount of time available similarly cue intent, for instance, in the case where groups are brought late into a process to respond to drafted material. Finally, it is important to determine the degree to which lines of communication are encouraged to enhance participation. Good communication permits the deeper understanding of the issues but also raises control issues related to the extent to which the participation is to be broad-based and provide a summary of widespread views, whether the representative speaks as an individual on behalf of an untapped group, and the extent to which more than one constituency is encouraged to caucus and consolidate perceptions. Participation in the statement of U.S. performance standards has shown almost every variation described above.

System Outcomes

In these times, educational systems emphasize their improvement in reaching goals by measuring performance. In the previous discussion of performance standards the role of outcome measurement and levels of attainment was described. At this point, it is important to moderate the analysis of outcomes by raising three important dimensions of their development and use: credibility, quality, and adaptability.

Credibility

The historical investment in standardized measures paired with the openness of criticism of the U.S. educational system have raised credibility questions about the measurement of educational goals. The specification of new kinds of performance standards has suggested changes in the types of measures that should be used, away from multiple-choice, commercially available tests toward more performance-based assessment. New approaches upset the stability of the system and raise questions about the appropriateness of prior beliefs. It is hard to argue briefly and nontechnically why existing measures may be no longer sufficient. Skepticism about the difficulty, fairness, and trustworthiness of new examinations has undermined in some locations the entire reform agenda. Beliefs are expressed that new types of assessments have been adopted as means to avoid rather than to strengthen accountability. Experiences in the U.S. suggest that it is essential to conduct a better analysis of important audiences to address about prospective changes in the measurement base. A focus only on the educational community, even as they are augmented by participatory groups, is much too narrow in a society where many educational institutions are viewed with

growing suspicion. Each of the identified audiences will need messages and ideas tailored so that they can understand, evaluate, and possibly support change.

One negative factor in the development of credibility has been the tight schedule for educational change. Paradoxically, credibility depends on both having enough time to educate communities and simultaneously moving rapidly enough to be regarded as an active, directed entity. No ready set of mechanisms or approaches has yet been developed to promote credibility, save in those cases where the electorate voted for massive educational change. In those cases where change is incremental and emerges over time, the need to provide clear examples of new measures is a first step. These examples should be supported by evidence of their impact, if any, in other locations, or at least testimonials by reputable experts from across the political spectrum. For example, indicating that performance-based measures have been widely used in the evaluation of business performance and of the combat readiness of military personnel provides a credible source of evidence for some.

Quality

Real credibility grows from high-quality measures. In the U.S., claims for new measures have continued to outstrip their documentation. Some technical issues related to the ability to provide individual scores that meet U.S. legal challenges to fairness and prediction remain. The fact that the scientific base of new assessments is rapidly evolving may help in the future but at the present time leads to equivocation, reliance on future promise, and only occasionally to useful recommendations.

The estimates of quality of assessments, and of the trust we place in the performance standards they are thought to represent, depend upon our concepts of validity. The various purposes for assessments and performance standards and the number of purposes assigned to any particular examination create different technical requirements. Among the many questions are the following: the size of the domain assessed, the boundaries on interdisciplinary domains, the types of cognitive demands, and the degree to which performance generalizes, as well as issues of reliability, stability, and fairness to students from various backgrounds. Each of these areas requires programmatic research in order to provide satisfactory guidance for design, use, and reporting of new measures. Will there be time and resources to develop the appropriate scientific base for new assessments

before skepticism, overpromising, and retreats to earlier measurement approaches take over? The answer, at least in the U.S., is not at all clear.

Adaptability

The final point in the model related to outcome measurement is the issue of adaptability or the extent to which outcomes are subject to change. Adaptability underlies the system's capacity to expand, contract, or change direction over time. For in the U.S., the understanding has only recently dawned that many attempts at educational reform were doomed because of the reliance on outcome measures used to assess progress and accountability that did not support the direction of desired change. Outcomes, of course, will be adapted to public goals. One question is the direction of change, whether it will push toward high levels of expectation and challenge for children and their schools or whether it will regress and constrict to fewer or lower standards. A second question involves the anticipated cycles of change. How long do systems need to be stable? What intervals can be expected for stability? Are cycles estimated by judging the point of full implementation by leading schools and systems, by a majority of educational institutions, or by the lagging institutions? On the point of outcome measurement explicitly, what happens when new forms of assessment are introduced to old trend lines used to gauge system progress over long periods of time? It is clear, for instance, that present techniques for linking disparate assessments are inadequate, and new approaches need to be conceived to permit cross-referencing among measurement methods. Is change or adaptation best served by targeting segments for change systematically? For example, if the content area of mathematics is the first to move to full implementation of curriculum and outcome measurement, should it, or a subset, be systematically revisited in a fixed interval to consider refinement, revision, or revamping? In reality, system change in the U.S. is less a feature of macroplanning than the happenstance confluence of resources, politics, and innovative ideas. In the area of measuring performance, however, it is clear that a variety of strategies can be considered for the systematic adaptation of measurement systems. Some states have adopted a substitution approach, out-with-the-old, in-with-the-new, babies and bathwater notwithstanding. Others have employed dual systems of performance-based and standardized tests, risking confusion and conflicting signals about importance. Certainly, in the next few years as expectations for system changes grow, we will learn more about how to phase in new systems and what elements of existing measures can be retained for

various intervals suited to particular purposes. One hopes balance will be a guiding principle.

Implications and Summary

Common issues face all OECD educational systems. These include the pressures to diversify their services and their clients, the importance of accountability to improve performance, the development or maintenance of credibility of the systems, and the realities of economic pressures, within the educational system itself, in the societies that support the system, and the emerging reality of world economic models. Scope, tradition, and diversity affect the manner in which systems move to meet new needs or to improve their effectiveness in accomplishing existing goals. Furthermore, all systems share common strengths and maladies, although in varying degrees. There is overlap in every system, although sometimes redundancy in goals, delivery, and measurement is desirable. Every system is subject to cycles of development and periods of high and lower public and political support. Issues of technical quality are represented and resolved in every system, as are needs for verification of system operations and outcomes. In the U.S., a concerted reform effort developed in the states and to some degree was ratified and extended by various national bodies. Reform is underway, intended to improve markedly the quality of precollegiate education. The clear goals are to develop greater coherence and stronger controls on output and to increase productivity, while simultaneously addressing student audiences of increasing diversity in background and preparation. There continues to be optimism that these goals can be met, but the outward look at OECD practices provides important grounding for our expectations.

What are the ways in which we can learn from the OECD studies? We have seen the ways that tradition, control, and outcome measure components of the reform model differently function in the participating studies. In the U.S., performance standards are fashioned in part to supply sources of control, guidance, and coherence for the system and to present an opportunity for greater public participation. Our choices in ways to learn from OECD countries are multiple. Because our contexts are so different, direct application and transfer are impossible. One option is to abstract lessons from the functioning of other systems. For example, in Germany, the press for uniformity among the states is

relatively low because the credibility of state systems is high. A second form of learning is emulation, that is, applying the essence rather than the details of particular approaches. For example, as we move in various states toward examination structures, we should learn from our British and Australian colleagues about how to develop examination systems that operate with public credibility, while in the U.S. we strive simultaneously to meet requirements imposed by our quantitative orientation, commitment to fairness, and anticipated challenges in the legal system. We can also study how change is accomplished in countries that are moving in various directions in the areas of common curriculum, examinations, and articulation of priorities. France, Sweden, and others provide cues to assist us in the selection of sectors for incremental change. Taken as a whole, the studies provide an opportunity for the U.S. to accelerate its educational change by sidestepping known difficulties and anticipating and planning for others. We have learned from OECD reports to imagine multiple uses for performance standards. We can see them not only as blueprints for curriculum design and assessments, their present formulation in the U.S., but perhaps as they are used in Spain: as consolidations of important societal values. The symbolic use of these standards and the processes through which they are developed can unify and strengthen belief in and purpose of education. The OECD country studies provide invaluable assistance to our reflection and search for strategies to improve our students' performance.