From Compliance to Improvement: Accountability and Assessment in California Community Colleges

W. Norton Grubb
David Gardner Chair in Higher Education
University of California, Berkeley

Norena N. Badway
Education Administration and Higher Education Leadership
University of the Pacific

Higher Education Evaluation and Research Group
# Table of Contents

**INTRODUCTION** ............................................................................................................................................ 2

**I. Peering Into the Black Box:**
   **A Model of Institutional Effectiveness** ................................................................................................. 7

**II. A Primer:**
   **External Accountability, Internal Accountability, Outcomes, and Student Assessment** ...................... 12

**III. The Crucible of the Classroom:**
   **The Contribution of Instructors, Students, Curriculum, and Institution** ................................................. 20

**IV. For the Long Haul:**
   **A Process for Developing an Assessment System** ..................................................................................... 27

**V. What to Expect:**
   **The Lessons of Experience** .................................................................................................................. 36

**VI. Speaking Truth to Power:**
   **The College Role in Improving External Accountability** ........................................................................ 39

**ENDNOTES** .................................................................................................................................................. 41

**REFERENCES** ................................................................................................................................................. 44
Community colleges are special institutions within American higher education. They are open-door institutions, the clearest examples of inclusiveness and equity in a system that is otherwise devoted to competitive entry, exclusiveness, and elitism. They include a much greater variety of students, including older students seeking upgrade training and retraining for new careers, “experimenters” trying to decide whether post-secondary education might suit them, welfare recipients and others trying to break into the economic mainstream, as well as conventional undergraduates. Partly as a result, they have expanded their goals and missions over time so that even their teaching purpose extends to many more goals, beyond conventional academic and occupational education. These missions now include remedial/developmental education, ESL for immigrant students, non-credit education for older students and sometimes the lowest-income students, advancement training and retraining, community service, and guidance to “experimenters” trying to find a path. They pride themselves on being “teaching colleges”, in contrast to elite universities that focus on research.

These distinctive features are crucial to the public service and the community-serving dimensions of community colleges. However, they also complicate the processes of public accountability and assessment. During the past twenty years, movements for external accountability have gathered strength both in K-12 education and in higher education, both driven to some extent by *A Nation at Risk* and the effort to force educational institutions at all levels to serve their students and their publics in more effective ways. External accountability — which we define simply as those regulations and incentives from outside an educational institution that try to hold the institution (or parts of an institution, like individual departments) responsible for various dimensions of quality — has in turn led to various kinds of assessments, or the attempts to measure in both quantitative and qualitative ways dimensions of educational processes and dimensions of outcomes. But both the accountability question — for what should colleges be held accountable? — and the assessment question — what should institutions attempt to measure? — are made infinitely more complex when there are multiple potential outcomes, when student goals and preparation are so different, and when the instructional conditions of institutions vary so much — from conventional daytime classes, to intensive evening formats, to non-traditional schedules for working students, to non-credit offerings in community settings, to learning communities, on-line education, and many other innovations. Therefore the process of developing comprehensive assessments, and even what we will call (particularly in Section IV) an assessment system, requires careful attention to the special conditions of community colleges.

There’s no question that external accountability, and the need to develop both institutional and student assessments in response, raises anxiety levels. No one likes being judged by external agencies, whether these are state or federal agencies, accrediting association, rating systems like those of *U.S. News and World Report*, or the “agency” of public opinion. Accountability done poorly carries the potential
Accountability done poorly carries the potential danger of distorting educational institutions and undermining their effectiveness and equity; assessment done badly simply measures the wrong outcomes or measures them in unreliable ways, generating the wrong signals for students, faculty, administrators, and accountability agencies.

But assessment and accountability provide certain powerful advantages as well. They provide the motivation and the information necessary to make good on the promise of being “teaching colleges” or “learning colleges”. When carefully crafted, they can give community colleges appropriate credit for their roles in enhancing equity, in providing upgrade training and lifelong learning, and in fulfilling some unconventional missions of the college. When assessment systems require faculty to collectively specify their instructional goals and then to measure progress toward them, they give faculty an appropriate voice in the running of their institutions. They also promote a form of research — research into teaching, and the creation of improvements in teaching — that is widely viewed as one of the legitimate forms of research for community colleges instructors (Boyer, 1990). When systems of *external* accountability are used to create mechanisms of *internal* accountability and institutional capacity — as we argue they should be, in Section II — then they provide the foundation for widespread institutional improvement. When the effort to develop student learning outcomes leads to better understanding about how students think about college and about learning, as we argue in Section III must be done, then colleges can become more effective learning environments — rather than students and instructors misunderstanding one another.

Above all, the entire process of creating assessments has the potential for creating a very different type of educational institution — one dominated by collegial relationships, a substantial sharing of leadership and instructional responsibilities, a much better understanding of students, and a greater attention to outcomes, including the quality of learning rather than simply enrollments and credits earned— in place of the more conventional institution that we call (especially in Section I) the “default” model. So those who animate the community college — the faculty dedicated to the teaching of many non-traditional students, the administrators guarding its special role within higher education, the students searching for more personalized learning environments and support into a complex world, the adjunct faculty and community supporters who contribute their specialized expertise — have many reasons to embrace the movements for accountability and assessment that have swept both K-12 and higher education.

But here it’s crucial to confront the two faces of accountability. Under some conditions external accountability, and the assessment measures they require, generate resentment, hostility, and minimal compliance from the institutions (or departments) being assessed. Under these conditions accountability turns into a pitched battle, a battle over compliance, stressful to all and generally ineffective in improving education. Under other conditions, when institutions embrace accountability and develop their own approaches to assessment and *internal* accountability, then the process can be converted from one of mutual hostility to one of mutual support, and institutional improvement is then the goal of all participants. Of course, between the poles of resentment and compliance, on the
The goals of both external accountability and assessment (internal accountability) should always be to improve institutional quality.

We think it better to call attention to deficits of accountability systems from a position of strength — from an understanding of and experience with creating assessment systems — rather than from a blanket objection to accountability and assessment.

One hand, and cooperation and improvement, on the other, there can be many intermediate degrees — for example when some departments within a college support assessment internal accountability and others resist it, or when faculty resist it while administrators need to show compliance in order to assure funding. But the point is that compliance mode is always to be avoided, and the goals of both external accountability and assessment (internal accountability) should always be to improve institutional quality. Everything in this report is therefore designed to enhance improvement as the goal of assessment.

More specifically, accountability is more likely to lead to improvement rather than compliance under the following conditions:

- when external accountability is promoted by external agencies in the spirit of support and improvement, rather than punishment, ridicule, or blame.
- when basic institutional survival and funding are generally assured, rather than being threatened — for example by threats of closure, “reconstitution”, the firing of faculty or administrators, or withdrawing funds from ineffective institutions.
- when assessment measures are valid and reliable in their technical dimensions, rather than mis-measuring the outcomes of institutions.
- when assessment systems can be devised by institutions (or departments) themselves, or devised in cooperation with external agencies, rather than being imposed in rigid and uncompromising forms.
- when accountability and the development of assessment systems can proceed under a reasonable timetable, and over a period of time. As the New England Association of Schools and Colleges has asserted in its policy statement, “Assessment is not a one-time activity; rather, it is evolutionary, ongoing, and incremental.”
- when external agencies provide consistent information, stability in their standards and interpretation, and technical assistance, rather than promulgating shifting requirements with uncertain interpretation that individual institutions are forced to implement on their own.
- when accountability respects the academic values of colleges, and acknowledges the complexity of the educational process, rather than treating colleges only as efficient machines for enrolling and credentialing students.
- when there is a mechanism for “social learning” to take place, or learning from the successes and failures of early stages of accountability and assessment.

Some of the conditions for improvement rather than compliance are under the control of individual colleges or departments. But some are out of their control, embedded in the practices of external agencies. Sometimes, therefore, the success of accountability and assessment is in the hands of these agencies, and for this reason we have included a section on “Speaking Truth to Power”, or the necessity of colleges being able to point out the flaws of inappropriate accountability systems. But we think it better to call attention to deficits of accountability systems from a position of strength — from an understanding of and experience with creating
assessment systems — rather than from a blanket objection to accountability and assessment.

In the recent past, California’s community colleges have operated under no fewer than four accountability systems. The state accountability mechanisms have included the Partnership for Excellence (PFE), which was designed to hold the system of colleges as a whole accountable, using system-level goals such as the number of transfers, degrees, certificates, and course completions. PFE has required colleges to collect information on outcome measures, but did not tie any incentives to the improvement of individual institutions; there was only the implied threat of funding contingent on the performance of the system as a whole. In addition, the State Report Card was an effort to assess the performance of all publicly funded workforce preparation programs by examining the employment, earnings, unemployment insurance, and welfare rates of individuals who participated. The federal accountability mechanisms have included performance measures required by the Vocational and Technical Education Act, including skill attainment in both academic and occupational courses (usually measured by grades), course completion rates, employment rates, and retention in employment. Other performance measures, specified in particularly rigid ways, are required of colleges participating in the Workforce Investment Act (WIA).

The newest efforts are those of the accrediting commission, the Western Association of Schools and Colleges (WASC); its Accrediting Commission for Community and Junior Colleges (AACJC) has recently developed a series of new Standards requiring colleges to develop mechanisms of assessment for student learning outcomes (with the unfortunate acronym of SLOs), along with a Self Study Manual for institutions being reviewed for re-accreditation.

However, the efforts to collect data under the some of mechanisms of accountability seem to have been suspended. Because of the recent chaos in state government, some state efforts like the Report Card are apparently stalled as the struggle to publish reports is nearly two years behind schedule. Virtually no one in Sacramento views the Partnership for Excellence as an accountability mechanism — it has become simply a funding vehicle (Shulock and Moore, 2002, p. 36). Administrative data to meet the VTEA requirements was held up while the new governor’s administration reviewed all contracts. Few colleges have participated in WIA and therefore have not had to comply with its requirements. But almost no one thinks that the era of accountability is over. State resources are getting increasingly scarce, especially in California, and funding shortages are always likely to lead to efforts to make sure funds are wisely spent. The dissatisfaction with public education shows no signs of abating, in either K-12 or higher education, and will surely lead to further efforts at accountability (Shulock and Moore, 2002). Congress, with its predilection for accountability rather than improving institutional capacity building, has introduced a version of the Higher Education Reauthorization Act that would strengthen federal accountability. And, over time, accrediting agencies have expanded their roles from a focus on inputs and institutional processes, to efforts to assess institutional effectiveness through program review in the 1980s, and to efforts to define student learning outcomes in the 1990s.
A final reason for California community colleges to embrace assessment as a mechanism for improvement is that the development of institutional and student assessments is part of the creation of internal accountability and capacity building — and, for reasons we clarify in Section II, it is difficult to respond to external accountability requirements without a well-developed system of internal accountability. The colleges that take the tasks of assessment seriously will not only be able to demonstrate their effectiveness and their claims to being “teaching colleges”, but they will also be better positioned to respond to future forms of external accountability.

This report is an effort to sort out the various issues in accountability and assessment, concentrating on the special conditions of community colleges. Section I describes a relatively simple model of educational institutions, to clarify what different types of assessments measure, and also outlines the special conditions of community colleges as distinct from other educational institutions. Section II then develops a kind of Primer, distinguishing external accountability and different forms of assessment, clarifying the notion of internal accountability and institutional capacity, and outlining the kinds of outcome measures community colleges might consider. Section III presents a further look at the internal workings of colleges, this time at the complexity of the classroom. In particular, we present recent evidence that instructors and students often have very different conceptions of what college is all about and what learning means, differences that undermines the potential for learning. Section IV describes how the process of establishing an assessment system might be conceived, including mechanisms to make sure that assessments are used for the crucial step of improvement. Section V clarifies the many ways that assessment might lead to improvement, emphasizing the importance of institutions linking assessment with their planning and budgeting efforts. Section VI draws on the history with assessments in other states to outline what we might expect from enhancing assessment in California, clarifying the barriers to improvement-oriented assessment and the conditions that would encourage it. Finally, Section VII on “Speaking Truth to Power”, outlines the requirements for developing accountability and assessment systems that have the capacity for social learning — of being able to learn from experience as a way of improving subsequent rounds of accountability. If the purpose of accountability is improvement, then government and regulatory agencies, including accrediting commissions, must themselves be capable of improvement.
Part I: Peering Into the Black Box: A Model of Institutional Effectiveness

In this section we present a model of how educational institutions work. Its purpose is to clarify the different types of information that might be collected in something called an “assessment”, as well as to clarify how they interact in leading to, or causing, student outcomes of various kinds. The model, initially developed to illustrate the precepts of the “renewed” school finance (Grubb, Huerta, and Goe, 2004) — or the conditions under which financial resources do and do not influence valued outcomes — also leads to several insights about resource use in community colleges.

A common way to view educational institutions is the simple input-output model in Figure 1. Resources of different kinds — money, instructors with certain qualifications, libraries, science labs and occupational workshops — influence outcomes, as do the backgrounds of students including both their prior education

![Figure 1](image-url)
and their family backgrounds and resources. But the educational institution itself — the interior of the black box — is never examined very carefully, and the specific ways in which resources influence outcomes are assumed but never specified. This has been the basis of many statistical exercises in K-12 education — though rarely in higher education — trying to determine which resources are effective in enhancing outcomes, particularly test scores. Unfortunately the results have been generally discouraging, sometimes leading to the over-simplified conclusion that “money doesn’t make a difference”. And the simplicity of this model leads to two dominant strategies for improving outcomes: increasing resources, often without thinking hard about how they will be used; and selecting more able students. The second of these is contrary to the goals of most community colleges; and the first of these does not seem possible in California, at least in the short run.

Figure 2

IC instructional conditions
NIC non-instructional conditions
SA student ability to benefit
Greater effectiveness is ... a two-stage process, of first identifying those practices that enhance outcomes; and then allocating (or reallocating) resources to those practices. The endless complaints about low levels of spending in California are certainly justified, since spending per full-time student in California is among the lowest in the country; but these complaints might have more political force if they were identified with institutional practices that demonstrably improve valued outcomes.

When we open up the black box of the college in Figure 2, it becomes clear that resources do not affect outcomes directly, but rather affect them by changing the instructional conditions (IC) of a college: the qualifications and preparation of instructors; pedagogical approaches; the use of labor, workshops, projects, outside learning, and other supplements to conventional classroom instruction; the use of novel instructional methods like learning communities; innovative technology-based methods, distance methods for certain students who cannot come to a campus; the hours at which courses are given to accommodate older students; and many other conditions. But students are themselves resources in the instructional process: those who are well-prepared, highly motivated, clear about their educational goals, and without the distractions of family and employment are more likely to respond to instruction. Conversely, students requiring basic skills instruction, ESL, or some direction in life, and those in the midst of the family-work-schooling dilemma — with demands from employment and family responsibilities that often override their commitment to schooling — may not benefit at all. And of course these two elements may reinforce one another: unmotivated or distracted students who face low-quality and unsupportive teaching are likely to become dropouts. Finally, and particularly in community colleges with their non-traditional students, non-instructional conditions (NIC) are often important. These include student services like guidance and counseling, financial aid, child care, student support groups, and the range of activities that can lead to the social integration of a student into a college (Tinto, 1987). Some of these, like tutoring and learning labs, are tied to the learning that takes place in classrooms and workshops, but most of these elements enhance the non-instructional conditions of students’ lives.

From this model of how an educational institution functions, a number of important consequences follow:

- Money is not inherently effective. It is effective only if it is spent on dimensions of colleges — IC, SA, and NIC — that are themselves effective in enhancing student outcomes. Greater effectiveness is therefore a two-stage process, of first identifying those practices that enhance outcomes, and then allocating (or reallocating) resources to those practices. The endless complaints about low levels of spending in California are certainly justified, since spending per full-time student in California is among the lowest in the country, but these complaints might have more political force if they were identified with institutional practices that demonstrably improve valued outcomes.

- A college that claims to call itself a “teaching institution”, or a “learning college”, should emphasize as institutional priorities various dimensions of instructional conditions that are arguably related to levels of learning. In earlier work on teaching in community colleges, Grubb and colleagues stressed 9 dimensions: the structure of instructors’ roles; hiring practices; promotion, tenuring, and teaching evaluations; pay scales and merit pay; pre-service education; mentoring new instructors; in-service education and staff development; innovative approaches to teaching, like Writing Across the Curriculum and learning communities; and support (or the lack of
support) from administrators (Grubb and Associates, 1999). In this report we include a tenth that we have not seen in many community colleges, but which has been promoted by accrediting associations: assessment systems that systematically measure student learning. As we clarify in Section VI, under the right conditions student assessments can lead to changes in other dimensions of colleges.

• The special conditions of community colleges complicate how we understand a model like that in Figure 2. The missions of colleges and the goals of students are much more varied, and so the range of student outcomes considered must be greater than in four-year colleges. The variety of student preparation is much greater, so that measuring student ability to benefit (SA) and levels of student preparation are more difficult, but also much more important. There is much greater variation in attendance patterns, which can be interpreted as part of student ability to participate (or SA), and many more outside demands on students from employment and family responsibilities. And, most importantly of all, student attitudes toward the purpose of college and toward learning itself are crucial, and surely vary both among students and between students and instructors; but they are often detrimental to learning itself (as we show in Section III). Without understanding these attitudes, and in some cases correcting them, it is foolish for instructors to propose student learning outcomes because students may not understand why these outcomes are important.

• There is also enormous variation among community colleges themselves, ranging (just in California) from high-quality occupationally-oriented colleges like L.A. Trade-Tech, to transfer-oriented institutions in areas with poor levels of high school preparation like Laney, to suburban colleges with well-prepared middle-class students like Santa Barbara or Foothill, to rural community colleges serving very different kinds of communities and labor markets like West Hills. The notion of creating a single assessment system for all community colleges in the state — for example, by developing a state system of curriculum standards, as has been done in K-12 education, and then purporting to measure progress on achieving these standards — would probably work poorly in community colleges because goals and missions vary so much, because the kinds of occupational programs offered vary, and because having statewide “conversations” about standards would be so difficult.

For this report, then, we assume that individual colleges are the appropriate units to develop systems of assessment, rather than districts, regions, or the state as a whole. There is yet another advantage to doing this: As we argue in the next section, responding to external accountability requires creating a system of internal accountability, where instructors and administrators create a web of responsibilities with one another for the improvement of learning. These responsibilities are easier to create within a college, where individuals already share responsibility for...
students, and much more difficult to create among colleges. And so both the creation of assessment systems, and the use of accountability for improvement rather than compliance, requires developing these practices within individual colleges.

In the end, as local colleges develop systems to improve student learning outcomes, they will, at the same time, meet most reasonable federal and state compliance measures — a win-win for students, colleges, and policymakers. Further, as colleges discover effective indicators of student success and share those with other institutions and policy makers, local experiences can inform statewide policies — local and state Boards can understand the variety of compliance measures that are most likely to be both effective and efficient.
WITH THE HELP OF FIGURE 2, WE CAN NOW PROCEED TO CONCEPTUALIZE EXTERNAL ACCOUNTABILITY, INTERNAL ACCOUNTABILITY, OUTCOMES, AND DIFFERENT FORMS OF STUDENT ASSESSMENT.

EXTERNAL ACCOUNTABILITY

The forms of accountability external to colleges — those regulations and incentives that try to hold colleges (or parts of colleges, like individual departments) responsible for various dimensions of quality — have typically come from state and federal governments, though they could also come from district administrations. In addition, accreditation agencies like WASC impose other forms of external accountability, since they have the power to deny accreditation and access to student grants and loans to those institutions that fail to meet their standards.

The most popular targets for state and federal external accountability recently have been certain measures of student outcomes, specifically measures of progress through institutions — courses passed, program completion rates, transfer rates, employment rates, and the like. The WASC standards apply to several different elements in Figure 2: Standards IIIB, C, and D refer to funding and resources; IIB (student support) is part of non-instructional conditions; IIA (instructional programs) and IIIA (human resources) encompass elements of instructional conditions; and requiring a process to formulate student learning outcomes focuses on other kinds of outcomes, as do elements of IIB, institutional effectiveness. One can also imagine state or federal external accountability applied to measures of student learning. For example, California’s Academic Performance Index (API) holds all K-12 schools accountable for student performance on various tests that comprise the API. Accountability has also been applied to individual students; for example the Florida rising junior exam, the CLAST, must be passed by all students who want to complete an Associate degree or to enter the junior year of a four-year college either as transfer or “native” students. While rates of passing the CLAST are reported by colleges, only students are accountable to the CLAST.

In sum, external accountability can be applied to many of the elements in Figure 2. Whether some of these forms of external accountability make sense for community colleges is a different point, an issue of whether external accountability systems themselves have desirable properties. For example, quite apart from the many technical atrocities of the API applied to K-12 education, we would argue that holding all community colleges accountable to a uniform set of statewide learning objectives would be conceptually objectionable as well as technically difficult.

The characteristic of these different forms of external accountability is that they are established by agencies external to the college. Whether colleges themselves
play any role in shaping these forms of external accountability varies a great deal. The Chancellor’s Office has an extensive process of formal consultation before regulations are imposed, but legislation can always override this process. While the U.S. Departments of Education and Labor took a great deal of input from states when developing their accountability frameworks, the final requirements were designed for evaluating and comparing states, and not for assisting local program improvements. Complaints about the rigid form of WIA performance measures have been especially common among community colleges. And the GAO in their evaluation of WIA (GAO, 2002) clarified how useless the WIA performance measures are for state and local purposes:

Labor reports that the performance measures are not intended to be a management tool. State and local officials, therefore, must develop alternative methods if they want to assess the quality of their services so they can identify problems and improve programs in a timely way.

And so the improvement of external accountability systems — improvement in various technical characteristics, as well as improvement in the capacity of external accountability to promote improvement rather than compliance — is a question of whether social learning can take place, a question to which we return in Section VI.

**Internal Accountability and Institutional Capacity**

While much of the angst over the accountability movement has focused on external demands, a less-noticed point is that the ability of institutions to respond to external demands — or, more specifically, to respond to them through improvement rather than token compliance — requires internal accountability, and then several other measures of institutional capacity. Internal accountability can be viewed as the responsibilities of individuals within institutions to one another, the answers to the question of “to whom or to what are you responsible?” In a conventional college of isolated instructors and distant administrators, instructors may feel responsible only to their conception of a discipline, sometimes (for occupational instructors) to the requirements of external employers, sometimes to the academic or the non-academic needs of students, but rarely to others within the college. But where institutions have created processes for developing shared norms and expectations, ways of determining collective responsibilities and individual actions that contribute to shared goals, and decision-making structures for determining and allocating responsibilities (like the Assessment Committee we mention in Section III), then they are best able to respond to external accountability in constructive rather than defensive ways. In the absence of internal accountability, however, educational institutions are either unable to respond to external demands, or do so with compliance-related efforts determined largely by administrators in a top-down fashion.

The best evidence for the centrality of internal accountability comes from K-12 education, where examination of external accountability in a number of states has clarified the importance of internal accountability mechanisms. In addition, in
Grubb’s work on community college teaching, he identified a number of colleges and departments that he thought merited the designation of “teaching institutions” because of their insistent focus on the quality of instruction; all of them were colleges (or departments) where mechanisms of collective decision-making and responsibility had been created, converting an institution of isolated individuals into a real collective able to devise collective goals and move systematically toward them (Grubb and Associates, 1999, Ch. 8). Finally, there’s the well-known case of Miami-Dade Community College under the leadership of Robert McCabe: when a problem arose, McCabe would appoint a committee of the relevant faculty, administrators, and staff; let the committee work through a conventional policy analysis of what causes of the problem were responsible and how they could be best remedied; allocate the responsibility for carrying out the reforms to a new office, or sometimes to a specific section of an existing office, so that implementation would not be forgotten; and maintain an advisory group of faculty and administrators (Roueche and Baker, 1987). In this way the responsibility for diagnosing the problem and creating a solution was shared among a larger working group, and consistently faculty, administrators, and staff were responsible for institutional improvement.

Internal accountability is in turn one component of institutional capacity. In this context, the capacity of an institution to respond to external demands. Another dimension of capacity includes leadership, but particularly leadership that is shared widely among faculty and administrators, distributed leadership (Spillane, 2001) rather than leadership held in the hands of a few top administrators. A third component is the knowledge, skill, and resources about the work that must be done, not just the technical skill but also the energy and motivation of the faculty, administrators, and staff who must devise and participate in mechanisms of internal accountability. As part of this, we should stress that certain kinds of resources are necessary for capacity building: the participants must have the time and the energy to be able to participate in changes, and sometimes additional funding may be necessary — to hire outside individuals with specific skills, to hire additional instructors to reduce teaching responsibilities, to create forms of staff development or blocks of time for collaboration. Where colleges are stretched to the limit and have absolutely no “slack” or spare time, finances and energy, then such changes may be close to impossible.

There are still other resources that we suspect are necessary for the development of internal capacity. One is stability — of external accountability mechanisms, of institutional goals, of personnel (leadership and faculty who participate in decision-making structures), or funding and other external demands. Where these elements change rapidly and abruptly — where the state’s assessments change, for example, or new presidents and chancellors come in with radically new goals, or state funding goes through the kinds of gyrations that have taken place the past few years — then it’s hard to make forward progress on the development of internal accountability. Another is surely trust among members of an institution. The coherence of an institution — the extent to which its components are part of an overall plan to move students forward, rather than fragmented and uncoordinated — may also be crucial. Stability, trust, and coherence are quite abstract resources: it’s difficult
to know how to measure them, and if they are missing it’s difficult to know how to create them. And while their value has been documented for K-12 education, they have not been part of the dominant concerns in community colleges. But if these resources are important for student outcomes, then only a serious system of institutional and student assessment will be able to clarify these linkages.

**Outcomes**

While systems of *external* accountability, and therefore *internal* accountability, do not always focus on outcomes, they often do. The range of student outcomes that could be measured is exceedingly broad, and the incorporation of student learning outcomes into WASC accreditation standards expands the kinds of outcomes that might be assessed. Somewhat arbitrarily, we will classify outcomes into three types: cognitive outcomes; non-cognitive outcomes; and conventional measures of progress.

- **Cognitive outcomes** include the majority of what are usually considered learning outcomes — writing ability, the ability to understand a passage of reading (whether from a literary text or from an automotive manual), the ability to use arithmetic and algebraic procedures correctly, the content knowledge of the various sciences as well as scientific procedures and methods, and the like. The student learning outcomes (SLOs) specified by WASC certainly include all of these cognitive outcomes.

There are various kinds of cognitive outcomes that are worth distinguishing from one another. For example, in the framework presented by Shavelson and Huang (2003), broad abilities might include reasoning (including verbal reasoning, quantitative reasoning, and spatial reasoning); comprehending; problem-solving; and decision-making. These are abilities that are often referred to as “higher-order abilities”: they can be taught in a variety of academic and occupational settings, they often find their way into programs of general education, and in their general forms they are difficult to assess since they encompass so many applications. In more specific academic and occupational areas, knowledge can be divided into declarative knowledge, the command of facts and empirical findings; procedural knowledge, knowing how to carry out various operations; schematic knowledge, or knowing why findings occur, usually requiring mental models or schema of how something being studied (a car, a cell, a mathematical problem) works; and strategic knowledge, or the understanding of when certain knowledge applies and how to apply it.

Very often, student assessment — for example, the initial assessment of students’ reading, writing, and mathematical abilities before they enroll in courses, to see whether they need developmental education or ESL, or conventional multiple-choice tests — focuses only on *declarative* and *procedural* knowledge, and the deeper kinds of understandings involved in schematic and strategic knowledge are untested. We note that while the subject or content knowledge of community college faculty is usually adequate, their understanding of different approaches to pedagogy — what some have called pedagogical content knowledge, or the ways that different pedagogies can be applied to specific disciplines and subjects — is
often lacking (Grubb and Associates, 1999). However, incorporating schematic and strategic knowledge into classrooms usually requires an understanding of a range of pedagogical approaches, a use of teaching methods well beyond conventional lecture and information transfer, and assessments that might include particular kinds of writing, problem-based approaches, portfolios, capstone projects, and other non-traditional assessment. It may, therefore, be necessary for some community college faculty to have substantive professional development in pedagogical approaches before they can articulate all these forms of knowledge, incorporate them into their teaching, and then propose assessments. But at least in theory, all these various forms of knowledge can be assessed, and can be incorporated into the learning outcomes of colleges.

A second large category of outcomes includes non-cognitive outcomes. While educational institutions have, in most times and most cultures, given priority to cognitive outcomes, a range of non-cognitive abilities (or “intelligences”, as Gardner, 1983, calls them) are important for performance on the job and for functioning in daily life, whether as family members, citizens, community members, or consumers. Occupational programs are often particularly rich in non-cognitive abilities, including the kinesthetic or manual abilities of craftsmen and engineering technologists; the spatial abilities of draftsmen, designers, carpenters and metal workers; the aural or musical abilities of musicians, medical personnel, craftsmen, metal workers, and others who need to hear how the transformation of materials is taking place. Often the project-based or “authentic” assessments used in occupational courses — moving between two- and three dimensions in drafting and fabrication classes, diagnosing and treating certain health problems, diagnosing and correcting auto problems, designing and fabricating certain kinds of clothing in fashion classes, different forms of cooking in culinary arts — assess a variety of non-cognitive as well as cognitive abilities simultaneously.

There are several kinds of non-cognitive abilities that are often overlooked but are especially important to community colleges. Interpersonal skills or abilities involve interactions with others, whether co-workers or fellow community members; and intrapersonal abilities reflect the knowledge of self. In some occupations — all those involving contact with the public, for example, and all those involving work in teams — interpersonal skills are crucial, and instructors may create work groups both to enhance these abilities and to assess them. And developing intrapersonal insight or “intelligence” is crucial, particularly for “experimenters” who don’t know what they might want to do, for those students who have trouble making decisions, and for those who are not confident in their abilities because of negative prior experiences in formal schooling. While it may be difficult to decide what progress in these dimensions means and how to measure progress, this too might be an important area of student assessment, particularly in colleges with large numbers of “undecided” students.

Finally, a third category of outcomes, conventionally used in external accountability, includes measures of progress through colleges and into post-college experiences. These include rates of completing courses, programs, credentials, and licenses; transfer rates to subsequent education or employment rates; measures
The ideal result of college-going is a combination of learning and completion, and therefore of student outcomes measured in several different ways.

It is worth distinguishing initial student assessment, the cognitive and non-cognitive abilities that students enter with (and the purpose of most current assessment programs), which is part of SA or student ability to benefit; intermediate student assessment, when a college examines the grades or learning outcomes partway through a sequence of courses, particularly to determine whether a student is making adequate progress; and final student assessments, measuring learning outcomes toward the end of a program or series of courses.

Student Assessment and Institutional Assessment

With the help of Figure 2, we can now see where different kinds of assessments fit. Most forms of external accountability have been measures of student progress, which are only some of the possible outcomes. The student learning outcomes (SLOs) specified by WASC potentially include both cognitive outcomes and non-cognitive outcomes, but assessing these outcomes certainly does not preclude measuring forms of progress through programs. In addition, it is worth distinguishing initial student assessment, the cognitive and non-cognitive abilities that students enter with (and the purpose of most current assessment programs), which is part of SA or student ability to benefit; intermediate student assessment, when a college examines the grades or learning outcomes partway through a sequence of courses, particularly to determine whether a student is making adequate progress; and final student assessments, measuring learning outcomes toward the end of a program or series of courses.

The variety of techniques or instruments appropriate for student assessments is simply enormous, and we cannot possibly do justice to the variety of technical issues in developing them. Conventional quizzes, tests, and student papers are perhaps the most familiar, particularly at the course level. But others include projects of various kinds including capstone projects, some of which involve multiple measures (like a project creating a physical object, a written report, and an oral presentation); the well-know classroom assessment techniques, which provide quick information to instructors specifically for the purpose of improvement (Cross and Steadman, 1996); the creation of case studies; performance in internships and other external placements, assessed either by instructors and/or external supervisors; oral presentations or performances, including artistic performances; portfolios; pass rates on external licensure exams; a creative activity like a dance choreographed
The powerful advantage of an inspections mechanism [as used in Great Britain] is that it provides direct evidence about instructional conditions, and — because inspectors develop expertise about teaching in a wide variety of college — a means of comparing instructional methods across colleges... many institutions in England have developed their own internal process for carrying out such reviews, anticipating the issues that an external review would raise but also contributing to a college’s ability to engage in continuous improvement.

or a print created; a trouble-shooting exercise in an occupational class; or other performance-based assessments. Indeed, we suspect that the enormous variety of ways of assessing student outcomes, and the technical difficulties of doing so, are what make the prospect of student assessment so daunting. However, there are by now many materials to help create student assessments, and the process we outline in Section III is one where a systematic and cumulative process can help any institution begin the process of creating such assessments.

In addition, educational institutions have different ways of examining instructional conditions (IC), non-instructional conditions (NIC), and student ability to benefit (SA) — which are components of institutional rather than student assessments. For example, the Community College Survey of Student Engagement (CCSSE) collects information from students about a variety of dimensions including the extent of active and collaborative learning, academic challenge in coursework, student-faculty interaction — all of which are dimensions of instructional conditions (IC); student effort, a dimension of students’ ability to benefit from instruction (SA); and support for learners, a dimension of non-academic conditions (NIC). Colleges might, for example, use the CCSSE in its entirety, or use selected questions from it, or use it as a guide for developing their own surveys of students about various dimensions of IC, SA, and NIC. (However, using CCSSE “off the shelf” would limit the process of a faculty deciding which dimensions of an institution’s are worth measuring.) In addition, the various examples collected by the New England Association provide illustrations of institutional assessments.

A very different type of institutional assessment, particularly valuable in assessing instructional conditions as well as the quality of student support, would involve an inspections process. Inspections, as they have developed in Great Britain, involve teams of insiders (e.g., faculty and administrators) and outsiders (like the members of a WASC team) observing in classes, examining the operations of student services, and otherwise judging the quality of an institution. While the inspections process can (like any form of assessment) be badly misused, the particular forms developed in Great Britain for Further Education Colleges — similar to community colleges — have been widely cited as helpful and supportive (Grubb, 2000). The powerful advantage of an inspections mechanism is that it provides direct evidence about instructional conditions, and — because inspectors develop expertise about teaching in a wide variety of college — a means of comparing instructional methods across colleges. While an inspections process including outsiders is similar conceptually to a WASC review, many institutions in England have developed their own internal process for carrying out such reviews, anticipating the issues that an external review would raise but also contributing to a college’s ability to engage in continuous improvement. In the U.S., such inspections mechanisms have been incorporated into the SALT procedures used by the New England Association of Schools and Colleges for K-12 education. As part of a larger assessment system, then, a process like the inspections process would provide a college with much more information about the conditions of instruction than any other method.

We note the value of Figure 2 in clarifying the relationship between institutional
a fully-developed approach to improving the effectiveness of community colleges would collect information about both institutional conditions and student outcomes, and in the long run be able to demonstrate the linkages among them.

...
In Part I we started to examine the inner workings of colleges, to develop ways of thinking of how research is used, and what aspects of the process correspond to elements of accountability and assessment. We continue this analysis, now delving into the classroom — the very heart of the teaching (or learning) college — to clarify yet other elements necessary to creating student learning outcomes.

In every classroom, conventional as it may appear, there are at least four elements contributing to its success or failure. Instructors, with their own approaches to the subject matter and to pedagogy, are the most obvious. Students, their levels of preparation, and their attitudes toward college and learning, are the second obvious group of participants. The curriculum or content is the third; sometimes it is generated by the instructor, sometimes it is dictated by the textbook used, and sometimes it is imposed from outside the classroom — for example, when a department decides to use a standardized curriculum for a particular course, when the content of transfer classes is determined by universities, or when occupational instructors follow curricula established by industry associations or employers or licensing requirements. And the institutional setting matters a great deal, since individual college practices, district rules and regulations, state funding and regulatory requirements, sometimes accrediting associations and other external agencies, and some federal agencies have their own influences, ranging from obvious to subtle.

When these four elements are in alignment, or in equilibrium, or consistent with one another, then classrooms are more likely to run smoothly. To be sure, they may not have much content — that is, there may be a low-learning equilibrium if instructors and students establish “treaties” (Powell, Farrar, and Cohen, 1985) where instructors demand little and students work little, or where instructors and students have been socialized to follow the textbook slavishly. What is optimal is to establish a high-learning equilibrium, where all four elements are dedicated to learning at a higher level or in more sophisticated ways. But potential problems emerge whenever any of the four elements are out of alignment with the others. If, for example, instructors disagree with the curriculum, they may undermine or embellish it, sometimes for the good and sometimes for the bad. If institutions fail to support instructors, as they often do, then they undermine the possibility of being a “teaching college”. And if teachers and students disagree about the content or about teaching methods, then classes may become “distressed”, with evident hostility between instructors and students, or may “collapse”, with very little learning going on (Grubb and Associates, 1999, Ch. 6, 8 and 9). The particular problems that emerge will, of course, depend on the specific details of how the four elements conflict with one another, and they will therefore vary from class to class and from
Our general knowledge of students and their attitudes toward learning is sorely lacking, partly because the empirical analyses of teaching in community colleges usually focus on instructors rather than students.

In general, we know the most about the attitudes of instructors toward students, subject matter, and the institutions they work in; there’s a substantial empirical literature and an even larger normative or “how-to” literature. Our knowledge of the institutional settings, and their consistency or inconsistency with instructional improvement, is also substantial, and certainly those working in particular colleges know a great deal about their local institutional conditions. But surprisingly enough, our general knowledge of students and their attitudes toward learning is sorely lacking, partly because the empirical analyses of teaching in community colleges usually focus on instructors rather than students (e.g., Hillocks, 1999; Grubb and Associates, 1999; Seidman, 1985; Richardson, Fisk, and Okun, 1983). Along the same lines, the conventional description of students needing remedial education describes them by demographic characteristics (first generation college students, more likely to be minority or immigrant, etc.) and external demands (employment and family), but aside from evidence of low self-esteem and external locus of control, there has been little effort to understand how they think about their education (Roueche and Roueche, 1999, Ch. 3). Similarly the curriculum — and particularly the trajectory of the developmental curriculum, as individual students move from initial assessments to early developmental courses, to “college-level courses” and then the more advanced courses necessary for completion of credentials or transfer — is also thin, particularly where coursework is organized as a series of independent, instructor-dominated classes. Without confronting both of these, developmental education is unlikely to be particularly effective, and the conventional questions about organization and “promising practices” are likely to lead nowhere.

Student Attitudes: To be sure, community college instructors know a great deal about students and their lives, and they are generally sympathetic to the “busied up” conditions caused by the need to work and (often) maintain family responsibilities. But, based on Cox’s (2004) close observations of composition classes in one community college, they are much less likely to understand how students think about the purpose of college and the nature of learning. Most obviously, many students are highly vocationalist: they are using the community college as a route to employment. As one mentioned,

I want to get my Cisco certification... and then I could get a job and then get paid once I get my CCNA. And then I guess get a degree, like a bachelor’s or something, and then get paid even more.

Vocationalist intentions in turn lead to highly utilitarian conceptions of learning, embodied in the common question “Why do I have to learn this?” Anything that is apparently unrelated to their vocational goals — certainly including developmental education, as well as general education requirements and much that smacks of being “academic” — is systematically avoided. As one student noted, ‘There a lot of stuff
you don’t use, so what’s the point [of learning it]?” So a great deal of the content that matters to instructors — and certainly the broader academic content necessary for “modern” occupations like health, business, and technical occupations — is undermined by student perceptions.

Surprisingly, given these vocationalist notions, students are often uncertain about their careers (see also Grubb, 1996, Ch. 2). They are unclear about their long-run goals, and where their additional schooling will lead them except to higher earnings, as one noted, proposing three wildly different occupations, “I ain’t sure if I want to be an EMT, nurse, lawyer — just whatever pays, so I can get out of debt.” They are often ignorant about the educational requirements of occupations to which they aspire, and unclear about the balance of formal schooling versus experience in getting jobs they want.

Given these long-run uncertainties, they tend to focus on short-run and highly credentialist goals: earning the GPA necessary for passing courses (what Becker et al., 1995, labeled the “GPA perspective”), earning credits for transfer, or completing a credential. For example, one student justified his failure to point out a discrepancy between what the instructor said and the textbook in these terms:

I’ve thought about saying something that would kind of disprove him, but I just keep quiet, do the work, and [pause] yeah, just get the grade.

Another noted the relationship between instructor directives and their vocational goals: “I have to do what he tells us to do, so I can pass the class, so I can get somewhere.” Optimally, they would like to learn as much as possible as they are passing courses and accumulating credits, but if they have to choose, then earning credits counts for more than the learning for which these credits are supposed to stand. And so anything that seems unrelated to earning credits, or that requires extra work to learn something is systematically devalued.

This attitude is exacerbated by an intense utilitarianism on the part of these students, who weigh carefully the costs and benefits of everything they do. If the effort (costs or time or extra courses taken) outweighs the benefits (credits with vocational applicability), then they will avoid that effort, even if it might lead to more powerful learning. They are concerned with “getting their money’s worth” in every class, about the potential for “waste” of money or time. As one student noted, after finding out that she could have enrolled in a shorter certificate program

I would just be taking those classes in that field . . . I would have been able to go into the field, and then come back to school and take the stupid [core] classes later.

The result is that — in contrast to most instructors, who would like their students to be as captivated by their subjects as instructors themselves are — students must depend on extrinsic and vocationalist (and uncertain) motivation to get them through coursework; as one noted the importance of vocational goals in keeping her in school.
I don’t enjoy school. I don’t think it brings anything to me, to be honest. . . I’m just in school to have some kind of career.

But again, the view that learning is not enjoyable means that students resist real learning and instead emphasize GPAs and credits, the tangible evidence of vocational progress. For example, rather than seeing the revision of writing as central to learning how to write, they express frustration since they want to “get over it”. One asked himself “What’s the least amount of change I need to make to get the paper accepted?” Another admitted making changes in response to instructor comments that she didn’t understand:

I just correct them, and I just get it over with, and get it accepted — “accept my paper and let’s go”. That’s it. That’s the class. I don’t care, as long as I pass it.

In addition, most students seem to think that learning means accumulating factual and testable information — “informative information”. One complained about an instructor in these terms:

On tests, she tests us about stuff that’s in the book, but then she comes up with her own stuff. For example, she gives an example of a person and asks, “Would a behavioral psychologist treat this?” So she kind of comes up with brand-new things, and a lot of people are lost in that class too. . . . The only class I feel that I learned something is in math and in criminal justice. Because, for example, my criminal justice teacher, he kind of gives us notes up there and that’s our test. Right. And then I understand it.

Another noted that “Mr. D. does put stuff on the board, and I appreciate that — that’s like my enlightenment”, but she complained about his “giving us random essays that he finds”, thereby negating a form of enrichment that the instructor was trying to provide. They prefer “stuff on the board” and lecture — hopefully engaging lecture, with humor and anecdotes — as the most efficient way of learning, despite the fact that many instructors in community colleges avoid lecture both as an inferior pedagogy and as an inappropriate relationship between teacher and student. Many students express real dislike of discussion (“don’t waste my time for 45 minutes”) and group work as “high school”. One complained that her instructor was not “teaching anything”, and commented sarcastically at the beginning of a class, “Are we all ready for roundtable? Honestly, I feel like I’m back in high school — this is so stupid.” She and other students interpreted the absence of lecture as an absence of instruction, and their understanding of professorial authority meant that any activities that transferred initiative to students — as in all student-centered approaches including whole-class and group discussions at “roundtable” — were rejected as “not teaching anything”.

...most students seem to think that learning means accumulating factual and testable information — “informative information”.

[Students] prefer “stuff on the board” and lecture — hopefully engaging lecture, with humor and anecdotes — as the most efficient way of learning, despite the fact that many instructors in community colleges avoid lecture both as an inferior pedagogy and as an inappropriate relationship between teachers and students.
As part of their conception of learning as accumulating facts, many students are unable to interpret the different pedagogical approaches that their instructors take. They literally can’t figure out why there are differences:

Well, my professors: they have different styles of teaching. I appreciate that there are different styles of teaching, but it just kind of throws your whole — well, me personally, it just kind of throws my whole life around.

And so instructors who seek to have students interpret reading rather than simply regurgitate the facts, or treat the social sciences as ways of understanding the world rather than a series of conclusions or “laws” (as in the “law” of supply and demand), may find students resisting such instruction as “not learning”. In such cases an observer can see students shutting down discussion rather than trying to develop their ideas. In response to one instructor’s analysis of gender differences, two male students fended off discussion by saying:

Men and women are just different. It’s like comparing apples and oranges. Sometimes things can be simple. They don’t always have to be complex.

The most obvious problem with these views of college and learning is that many students systematically undermine their own learning, because they focus on grades rather than content, on efficiency rather than understanding, on useful or “relevant” courses rather than those that might amplify their intellectual sophistication for some future, uncertain and poorly-understood benefit. And because schooling is not “fun” — indeed, because their previous schooling has often been relatively unsuccessful and often demeaning — they are often fearflul, scared of being caught unprepared, feeling responsible in new ways (“I’m all by myself now”), intimidated by professors whom they view as distant authorities (“you’re real hesitant to go to them because of the way they are”). As one older student admitted, “Despite feeling mature enough and committed, it was still really scary. Oh my God, it was a life-altering change.” Their mechanisms of fear management are also counter-productive: keeping quiet in class; avoiding contact with the professor (“they don’t want to look like a fool”); avoiding classes with writing requirements; scaling down ambitions; avoiding assessment, including failing to turn in required work; postponing college; and dropping out — or “stopping out” with the intention of returning, a “decision” that often leads to dropping out permanently.

Finally, these views among students are inconsistent with what many instructors are trying to do, so there is a great deal of misconception and misalignment in these classes. Indeed, the counter-productive behavior of students often generates counter-productive reactions from instructors: shifting from discussion back to lecture and worksheets, as ways of controlling a class; retreating to facts and information transfer, since that seems to be what students want; dismissing students as “not ready” for college, or as unprepared when the real problem is that they need more encouragement; or expressing complete bafflement. As one instructor noted,
I don’t know if it’s because they’re [students] lazy or because they didn’t understand, or because they understood, but for some reason they have some kind of philosophical disagreement on whether that should be done or not.

These views among the students in Cox’s observations are surely not universal. Some students actively dislike lecture and understand that conceptual understanding is more important; some are actively “taught” or socialized by their instructors to be more active. As one student who had been part of participatory learning communities commented, “Talking gets a lot more people involved in discussion instead of, like, teachers standing up there doing lecture. I tend to lose my attention when there’s lecture” (Grubb and Associates, 1999, p. 87). The attitudes of students in composition and remedial classes, which are often not voluntary, may differ from their attitudes in courses they choose, and students in occupational areas to which they are committed may be engaged in deeper ways than students in academic or not-fun courses. Our point is not that all students are like those we have described in a single college, though we suspect that these attitudes are widespread among community college students. Rather, our interpretation is that student attitudes toward the purpose of college and toward learning are often unknown by instructors (including the best-intentioned), and the danger is that they may be inconsistent with what instructors believe and detrimental to learning. Until instructors understand these attitudes, there may be some kind of disequilibrium in the classroom, many kinds of missed opportunities, and in extreme situations a complete collapse of content.

The attitudes of students are complex, and no one knows much about how they develop. Surely family influence matters a great deal; earlier educational experiences “teach” students to think of schooling in different ways, for good and for ill; older siblings or peers may have their own influence. The selection mechanisms of who goes to community colleges, as distinct from who goes to elite colleges and who doesn’t go to college at all, may mean that community colleges have disproportionate numbers of students who think learning is not fun, and many more students who are fearful of formal schooling. Many of the students observed by Cox were influenced by movies and other pop culture about what college is “supposed” to look like. The powerful trend in this country toward vocationalism and economic goals has affected every level of schooling (Grubb and Lazerson, 2004), so it’s not surprising that students think in utilitarian and vocationalist ways. Students’ credentialist notions of college are consistent with a more general belief in credentialism; the limited economic circumstances of many college students force them to consider their budgets, as well as the benefits and costs of going to college, in more careful ways than upper-income students in elite universities. The most thoughtful instructors take care, particularly at the beginning of courses, to “teach” or socialize students about their preferred pedagogy, but many more seem to assume that students know how to learn in seminar formats, in lecture formats, in workshops or labs. And if different instructors have different conceptions of college and learning, then students get inconsistent messages, and — unless they are sophisticated in interpreting and then learning from different pedagogical
In a collaborative model of teaching, part of the instructor’s responsibility involves understanding how students perceive college, the curriculum, and the nature of learning, and when necessary, helping them revise their perceptions in ways that facilitate their learning. Changing the attitudes of students may be difficult, because it may have to fight against larger social trends and pressures; but it is likely to be better accomplished through the collective actions of all instructors than by the efforts of instructors one-by-one.

The implications of these student perceptions of college and learning are profound. If faculty begin to articulate student learning outcomes — the kinds of learning they want to see in particular classes, in sequences of classes, and in majors — and then develop the appropriate assessments, these efforts may make no difference if instructors are teaching in ways that students don’t understand, or if learning conceptual or “academic” material is resisted by students as “irrelevant” or useless (“what’s the point?”). If students think GPAs and credits are the point of college, then the whole exercise of formulating learning outcomes more carefully will be irrelevant. Instead, the process of developing SLOs must simultaneously incorporate students’ — as well as instructors’ — thinking about learning and the goals of education, in order to restore equilibrium to the classroom.
Any attempt to view assessment as a one-shot activity is almost surely doomed to become a form of shallow compliance. Therefore improvement-oriented institutions need to put into place a process for systematically developing a cycle of assessment—not just a set of assessments, but a system that can lead to initial assessments, then improvements in educational practices, and then to the revision and expansion of the assessment system itself.

If assessment is to be continuous, on-going and stable, then it must be overseen by a group that takes responsibility for all aspects of assessments. This might be an existing evaluation committee or institutional effectiveness committee, but in fact many colleges contemplating assessments have found it useful to create an Assessment Committee (or an equivalent committee with another name) with stable membership by individuals in certain administrative and faculty positions. The committee would certainly include the chief instructional officer; faculty with responsibility for the various academic and occupational subjects, including the chairs of departments; those responsible for remedial/developmental education in whatever configuration it takes, as well as those in charge of ESL and other forms of bilingual education; surely the head of institutional research and representation from student services and other non-instructional components. The linkages to other aspects of college governance are issues we take up subsequently, though it is surely desirable for the Assessment Committee to have representation from budgeting and institutional planning groups, so that spending, long-run planning, and instructional developments can be coordinated. On some campuses representation from the Academic Senate or from faculty unions may be appropriate. The representation of students may be advisable, particularly if students are to play important roles in carrying out assessments, and the roles of community members (including employers) needs to be considered—though employers could be included on sub-committees created for specific programs, for example.

By this point the Assessment Committee may seem like an amorphous and unwieldy group. However, it’s crucial to remember, that in a self-reforming institution focused on instruction, the Assessment Committee would be the central committee in a college, so that concern over the nature and effectiveness of
A lack of knowledge among participants as well as emotional rejection of assessment — or rejection on the grounds that assessment takes away faculty control, rather than reinforcing the faculty role, for example — are among the two most common barriers to progress.
The development of methods to determine the views of students is not a trivial task, and might well be addressed by a consortium of colleges, by a state association, or by some other collective group. Research that has been done so far (e.g., Cox, 2004) has relied on ethnographic methods, and they are too time-consuming and cumbersome to use on a wide scale. However, it is possible that student interviews and focus groups — in which students discuss their goals, their best and worst instructional experiences, where they think they have learned the most, and their attitudes toward instructors — would reveal a great deal about student perceptions. These interviews and focus groups need to be carried out by neutral observers, including institutional researchers, counselors, and individuals from outside the institution (like students at nearby universities). The inclusion of counselors might be especially powerful since in some innovative approaches counselors have the responsibility for reshaping the attitudes of students — for example, in introductory courses like College Success or in learning communities incorporating counselors (Grubb, 2004). The development of methods to determine the views of students is not a trivial task, and might well be addressed by a consortium of colleges, by a state association, or by some other collective group.

The final element in ascertaining the alignment of participants around instruction is to determine, as honestly as possible, what the institutional attitudes and priorities for teaching have been. The discussion of Grubb and Associates (1999, Ch. 8 and 9) provides a template for the questions that must be answered, including difficult political questions about administrator support versus lack of interest; the roles (positive and negative) of Academic Senates and unions; the resources (including time as well as money) given to instructors for instructional improvement and experimentation; the nature of existing staff development and whether it serves to improve instruction or merely to provide momentary entertainment; and the role of instruction among the many other priorities of a college. Some of these assessments are painful to make. For example, instructors we interviewed were ambivalent about labeling administrators as “managers”, or “pejorative”, authoritarian and unengaged, or “so busy counting beans and filling out forms that the creative and the innovative is way, way in the back seat”. But if colleges do not confront the institutional as well as the classroom dimensions of instruction, then the development of student learning outcomes can’t make much difference for students.

4. Setting goals

Every discussion of assessment emphasizes the importance of setting goals, for a college as a whole as well as for sub-units like departments, before creating a set of assessments. At this level, goals describe student outcomes including cognitive outcomes, non-cognitive outcomes, and existing measures of student progress. For example, English departments might establish their goals in terms of writing proficiency plus the ability to analyze texts in particular ways; business departments might define learning outcomes in terms of the ability to analyze local economic conditions and develop business plans, or in terms of organizational and procedural efficiency; auto programs might want to emphasize diagnostic abilities; general education programs might want to specify a list of civic competencies. At this stage an Assessment Committee may need to break into sub-committees for some goal setting.
But it’s also crucial to set goals or priorities for a college as a whole. For example, if large numbers of students are enrolled in business programs, then creating assessments for the business department might be a high priority; if the completion rates for transfer programs are already quite high, then further improving transfer might be a low priority. Since general education requirements in California (the IGETC program established in conjunction with UC and CSU) are stated in terms of completion of specified courses with B grades or better, the goals for general education are in the first instance measures of student progress, not cognitive learning outcomes, and the hard work of defining gen ed learning outcomes might need to wait until there is some consensus in IGETC about what these should be. The vision of an assessment system, therefore, is not that every goal is specified and then assessed at the outset, but rather that the process of assessment starts with the highest institutional priorities, followed by a longer process of continual improvement.

5. Developing assessment instruments

Only when goals have been specified is it possible to develop assessment instruments: the tests, writing exercises, oral presentations, projects, occupational demonstrations, portfolios, evaluations by outsiders as well as instructors, and other ways of developing both quantitative and qualitative information about student outcomes. Some of these are easily developed and interpreted; others are much more difficult.

No matter what form they take, assessment instruments should be governed by conventional criteria from classical measurement theory, particularly validity and reliability. Validity refers to an instrument measuring what is intended to measure, rather than (inadvertently) measuring something else; word problems in math, for example, sometimes prove to be tests of reading comprehension rather than procedural knowledge in math. Reliability takes several forms, but it is often thought of as test-retest reliability — the fact that re-testing a person on another day would yield the same result — or inter-rater reliability, the fact that two individuals scoring the same assessment (like a writing passage or an occupational performance) would give the same scores. Norm-referenced tests or assessments measure individual outcomes compared to the sample of people taking the test, while criterion-referenced tests measure performance compared to certain norms or criteria — for example, compared to conceptions of adequate or proficient performance that faculty may identify. For purposes of judging improvement criterion-referenced assessments are clearly preferable, though they require faculty to specify the goal or criterion that they want to meet. Again, as in so many areas of assessment, there are many sources to turn to for the technical details of measurement theory.

The process of developing scoring rubrics — for example, to identify qualitative differences in writing, or the outcomes of a particular project, or the activities required in a capstone project, and then to assign scores (e.g., from 1 to 5) in ways that are reliable — has been well-developed, and there are several guides to developing such rubrics (e.g., Arter 2001). However, such processes have not been
applied to many of the outcomes that community colleges might want to emphasize. For example, some non-cognitive outcomes like the ability to formulate goals, or the ability to take steps toward attaining goals, might be assessed through activities including writing, contacts with counselors, participation in career development classes, and participation in internships intended to provide information about employment opportunities. These sources of information might then be combined into an overall rubric or score, particularly as a way of examining the progress of “experimenters”. But scoring rubrics have conventionally been applied to cognitive outcomes, and colleges that place a priority on measuring non-cognitive outcomes may find themselves in new territory.

6. Implementation of assessments

The obvious next step is to begin using such assessments. While this may be a conceptually trivial step, it is often a politically difficult stage because it requires faculty, counselors, and other student personnel to change their routines, perhaps even to abandon their prior methods of assessment, and to begin using new assessments — and resistance from faculty has been one of the most consistent barriers to assessment, as we discuss in Section V. One task of the Assessment Committee and its sub-committees, then, is to assure the participation of faculty in using new assessments. This will of course be easier to the extent that faculty have participated in developing goals and then assessments, to the extent to which assessment instruments are familiar rather than new and challenging, and to the extent that the process of implementation is supported with staff development rather than introduced on a rushed timetable with little explanation.

Similarly, some assessments require the participation and the cooperation of students — for example, questionnaires to students about their goals and intentions, assessments of learning as well as non-cognitive outcomes, and institutional assessments like the Community College Survey of Student Engagement or a locally-developed alternative. Again, students are more likely to be willing to participate if they have had a hand in setting goals and developing assessments; to the extent that assessment tasks are part of their regular coursework rather than activities they have to undertake on the side; and to the extent that performance on assessments are part of the evaluation of students rather than an exercise that does not affect a student’s learning or progress. This argues for trying to create learning assessments that are embedded into coursework and other normal learning activities (including internships and the like), rather than being independent exercises.

7. The translation of results into useful forms

In some cases the results of assessments are self-evident. Conventional course grades are intended to indicate whether a student has passed or not. Scores on criterion-references tests and assessments (including projects) indicate whether an individual has achieved an acceptable level of performance; the rubrics that translate the results of a complex assessment like a writing exercise or a portfolio submission into a scale also indicate whether an individual has achieved proficiency. In other cases, however, where there are multiple assessments, or assessments in different
...non-use of available information remains a problem in many community colleges, and anything that enhances the use of information will help further the goal of institutional improvement.

forms, then translating results into forms that can be used by instructors, counselors, and other student service personnel may be difficult. The kinds of comparisons that may be desirable depend on the issues that come up in different community colleges. For example, some colleges are particularly concerned about the progress of different racial groups or immigrant students; some colleges may need to distinguish younger students from older students, “experimenters” from those with relatively well-defined goals, or students from particular local high schools. The calculation of some well-known measures — transfer rates are the most obvious examples — have been the subject of endless discussions about what is the “right” measure; but if a student record system is set up appropriately, any conceivable measure of transfer can be calculated — partly in order to see where in a series of linked events the transfer process breaks down. And in many colleges data exist in separate data bases, and these need to be linked or aggregated in order to have comprehensive information.

As in other aspects of assessment, several systems for such translation are available. For example, the National Center for Research on Evaluation, Standards, and Student Testing (NCRESST) has developed the Quality School Portfolio, a web- or desktop-based analysis package aimed at K-12 education that allows multiple kinds of information about student performance to be aggregated in different ways, to compare groups of students at one time, to monitor progress of students, to compare instructors, schools, districts, or states, and to display the results in various ways to facilitate discussions about potential responses. The Berkeley Evaluation and Assessment Research Center has developed GradeMap, a system of transforming student scores into maps of student progress using Item Response Theory (IRT) models.

The translation stage is one that should be governed by several obvious criteria: flexibility and ease of use, comprehensiveness, simplicity and understanding of the underlying statistical transformations. The choice or development of a data management system is perhaps a task of the office of institutional research, in consultation with various potential users of the system. But it’s important to remember that non-use of available information remains a problem in many community colleges, and anything that enhances the use of information will help further the goal of institutional improvement.

8. THE USE OF ASSESSMENTS IN REVISIONING INSTRUCTION AND STUDENT SERVICES

Stages 1 through 7 all lead to the creation of assessment data, but of course the existence of better assessments does not necessarily generate any improvements — though one might anticipate that the very process of defining goals would then encourage a department or program to revise its teaching, and the process of uncovering student attitudes toward learning might cause some instructors to change their instructional methods. The process of examining the results of assessments, and deriving improvements based on these data, should therefore be considered an independent stage in the development of an assessment system.

External accountability mechanisms provide one stimulus for improvement. For
example, transfer rates and employment rates are included in every accountability system. For individual colleges, the question then becomes whether these rates are high or low compared to other colleges; why rates might be high or low, including the potential problems of specific demographic groups; whether rates are increasing or decreasing over time, or whether they vary with local economic conditions and business cycles; and finally what the potential solutions to identified problems might be. Solutions might include improvements in instruction (IC from Figure 2), improvements in student support services (NIC), or changes in the motivation and engagement of students (SA), potentially affected by counseling and other mechanisms. In occupational programs, external licensing sometimes creates targets for improvement, for example if pass rates on licensing exams are too low. And the process of accreditation by WASC will create other forms of external pressure to create student-learning outcomes and then to start using them for instructional improvement. In our interpretation of WASC regulations, and in light of what other accrediting organizations have done, this influence is more likely to involve pressure for improvement and monitoring of self-study reports, rather than the kinds of punitive steps that California has taken through the Academic Performance Index for K-12 education or that the federal government has taken in No Child Left Behind. We note that steady pressure, rather than punitive measures, are more likely to lead to improvement rather than compliance modes, with the kinds of results that we will summarize in Section IV.

In other cases, the stimulus for improvement comes from the institutional goals and the department-level goals set in Stage 3. For example, remedial/developmental education may create goals for movement through remedial coursework into “regular” college courses. If these goals are not met, or if certain kinds of developmental students are not making adequate progress, then this creates opportunities for improvement somewhere in the sequence of initial student assessment, remedial coursework, student support, and “exit” into regular courses (Grubb, 2001). Academic departments may create goals of preparing students to carry out certain analytic tasks; occupational departments may have goals of proficient performance on job-related tasks; and gen ed programs might articulate civic competencies or attitudes. These all create targets for improvement.

While we have stressed the use of assessments in college-level improvement, it’s worth noting that certain kinds of analyses can be performed only at the state level. For example, it is currently unclear in California whether differences among colleges in transfer rates are due entirely to variation in student socio-economic status, or whether some colleges have particularly high transfer even controlling for socio-economic status. Similarly, employment rates from occupational programs surely vary with local economic conditions, but how much they vary is something that only a state-level analysis could determine. The responsibility of individual colleges, therefore, is to carry out the appropriate college-level analyses and to formulate appropriate responses. The responsibility of the Chancellor’s Office or some other state office is to carry out the state-level analyses necessary to explain patterns among colleges.

The exercise of moving from information to improvement is an example of Steady pressure, rather than punitive measures, are more likely to lead to improvement rather than compliance modes...

The responsibility of individual colleges... is to carry out the appropriate college-level analyses and to formulate appropriate responses. The responsibility of the Chancellor’s Office or some other state office is to carry out the state-level analyses necessary to explain patterns among colleges.
The notion in the assessment movement that new forms of information can lead to institutional improvement assumes a capacity for data analysis that is often not present.

Program improvement requires continuous funding for institutional research, staff development, and the development of assessment.

If state policy, or state funding levels, or district-level regulation, or inappropriate forms of external accountability, or accrediting agencies make the conditions for creating systems of assessment worse, then individual colleges will need to blow the whistle on these forms of interference...

Policy analysis, a well-developed process for specifying goals; arraying the options for improvement; assessing the effectiveness, benefits and costs of each of the options; and finally making decisions about which policies or practices to adopt based on these results as well as political feasibility — which might, for example, reflect the desires of the community, the constraints of state policy-makers, or the norms of community colleges. In this process, the specification of goals (in Stage 3) and the collection of data about performance (Stage 4) are obviously necessary; but from Figure 2 it is equally necessary to have certain information about the institutional practices that might cause the observed outcomes. Therefore certain kinds of institutional assessment, as well as student assessment, are integral to the process of improvement.

But while policy analysis has been carefully formulated and extensively described (e.g., in Bardach, 1996), it is not common within community colleges. The capacity of examining data and generating potential improvements is not usually part of the preparation of either faculty (except perhaps instructors in political science or statistics) or administrators. Indeed, the notion in the assessment movement that new forms of information can lead to institutional improvement assumes a capacity for data analysis that is often not present. The implication once again is that the creation of an assessment system may require forms of staff development, in this case aimed at instructors and administrators as well as institutional researchers, to enhance the capacity to examine assessment data with the perspectives of policy analysis.

In addition to expertise, the movement from assessment to improvement requires certain other kinds of resources, broadly defined. One is stability: stability of institutional goals; relatively stability of external assessments, so that changes can be readily incorporated into assessment systems; sufficient stability of personnel (particularly of the Assessment Committee) so that there is adequate continuity and institutional memory from year to year; stability of funding, so that assessment is not sacrificed in the process of scrambling for resources. A second is leadership, which in this case means consistent support from the administrators most responsible for the direction of a college: the president, the chief academic officer, perhaps the district and the board of governors. (We note that it is particularly destructive to have high turnover in the president’s position, particularly if new presidents fail to respect earlier policies and developments.) A third, obviously, includes fiscal resources, since the development of assessments, the provision of certain forms of staff development, and the maintenance of institutional research all cost non-trivial amounts of money. It appears, from other regions, that these are not enormous sums, but program improvement requires continuous funding for institutional research, staff development, and the development of assessments. Individual colleges cannot always control these conditions, but those that are serious about creating assessment systems as mechanisms of improvement can at least watch for external influences that might undermine these conditions. And if state policy, or state funding levels, or district-level regulation, or inappropriate forms of external accountability, or accrediting agencies make the conditions for creating systems of assessment worse, then individual colleges will need to blow the whistle...
on these forms of interference — the subject of Section V on speaking truth to power.


A final stage in the process is intended to come full circle, back to Stages 2, 3, 4, and 5. An assessment system itself may need to be examined and revised — for example, if the implementation is initially uneven, or if the wrong kinds of information are collected, or if information proves incomplete, or if information fails to be used for improvement. Therefore a complete assessment system should have capacity for self-improvement — for periodically stepping back, examining how well the system has worked, and making changes as appropriate. This is the responsibility of an institution-wide Assessment Committee. Such revisions might be stimulated by external accountability, or by accrediting associations, or by licensing requirements, but they should be — like the creation of an assessment system in the first place — a college-level responsibility.

It’s possible that certain forms or procedures can help the process of reformulating an assessment system. For example, the New England Association of Schools and Colleges has developed a simple Program Assessment Inventory as part of the self-study process that asks individual programs — academic and occupational majors or concentrations, general education programs, developmental education and ESL departments, student services — to specify their goals for student learning (analogous to Stage 3); the ways learning is assessed (Stage 4); the changes the program has made as a result of assessment (Stage 7); and the plans for improving assessment in the next two years (Stage 8). The purpose is to provide the Assessment Committee (or the self-study committee for an accreditation review) with information to aggregate to the college level, about which programs have on-going assessment, what types of assessment exercises are commonly used, whether assessment results are used for improvement, and what the general priorities are for improving assessment.

Overall, the development of an assessment system is a lengthy process, one that cannot happen overnight. But the specification of a process should de-mystify assessment, clarifying that a series of incremental steps — each of them familiar, each of them extensively practiced in other areas, each of them with extensive resources available — can over time generate a complete system for the purposes of improvement.
The path of creating assessment systems may seem long and arduous, and starting down the path may seem like an article of faith — untested faith that all the individual work and institutional resources necessary to create an assessment system will in fact lead to improvement in various student outcomes. Fortunately — because faith is often an uncertain basis for action, at least in secular matters — there is at least some evidence about the effects of assessment. The North Central Accrediting Commission (NCAC) instituted its first “Statement on Student Assessment and Student Academic Achievement” in 1989, among the first public bodies to assert the need for student assessment as a way of college becoming more learning-centered. After ten years the NCAC pulled together what it had learned in ten years of steady effort. None of the results are surprising, but that’s part of the good news: assessment works well when it is developed and implemented as planned, though there are several obvious barriers to its implementation.

Most obviously, NCA found that colleges varied substantially in the extent to which they had implemented assessment. They divided colleges into Level One institutions, with assessments in their infancy or that had stalled; Level Two institutions, making progress consistent with the value the institution placed on assessment; and Level Three institutions, with assessment systems that were structured, systematic, on-going, and sustainable. While it is not clear what fraction of institutions fell into each of the three categories — these categories were never used to take a “census” of how institutions progressed — the NCA has used these three categories to develop a matrix, called the Levels of Implementation, that describes each of 8 institutional characteristics (collective values, missions, etc.) according to Level One, Two, or Three. The Levels of Implementation can then be used by institutions to assess their own progress, to strengthen their programs of assessment, and to inform evaluation teams.

In examining the hundreds of evaluation team reports and self-study reports over the initial ten-year period, Lopez (1999) concluded that three broad factors were responsible for hampering the development of assessment systems:

1. basic misunderstandings about the purpose and nature of assessment;
2. emotionally-based resistance to assessment from those responsible for it, including both faculty and administrators;
3. inadequate information and skills needed to conduct assessment.

We note that the first and third of these can be remedied through workshops and staff development that provides basic information about assessment, how to do it, models and experiences from other institutions, and other ways of informing faculty and administrators. In some cases, of course, new expertise may have to be brought into a college, for example by hiring a new institutional researcher or a director of assessment.
Rather than undermining the autonomy of faculty, the development of goals and assessments by faculty and their use in instructional improvement, elevates faculty to stronger roles within community colleges and provides them opportunities for the kind of research into teaching that is surely legitimate for community college faculty.

However, the second problem — emotional resistance to assessment, particularly among faculty who feel that all outside intervention is unwarranted — falls into a different category. In these cases leadership — particularly in the form of judicious sticks and carrots — may be necessary, but a top-down and coercive approach is likely to further alienate faculty. It’s perhaps more appropriate to think of developing a different institutional culture, a system of internal accountability where the responsibilities of instructors and administrators to one another and to students creates a greater acceptance of assessments as mechanisms of institutional improvement and instructor professionalism. Rather than undermining the autonomy of faculty, the development of goals and assessments by faculty and their use in instructional improvement, elevates faculty to stronger roles within community colleges and provides them opportunities for the kind of research into teaching that is surely legitimate for community college faculty (Boyer, 1990).

Whatever the solution to emotional resistance may be, this particular barrier clarifies that the development of assessment requires rethinking the conceptions of what faculty and administrators do, the relationships among faculty and administrators, and the conception of what a “college” is.

The three overarching issues in turn have manifested themselves in 8 more specific problems that colleges face:

- difficulties in involving faculty and students in assessment;
- difficulties in developing program goals and measurable objectives;
- difficulties in developing (or selecting from available sources) direct and indirect measures of outcomes aligned with program goals and measurable objectives;
- difficulties in collecting and interpreting data;
- difficulties in disseminating assessment data because of insufficient or incomplete feedback loops;
- difficulties in obtaining or reallocating the funds needed for assessment activities themselves;
- difficulties in linking the assessment process with institutional planning and budgeting processes, so that the funding and human resources necessary for improvement can be assured;
- difficulties in understanding and providing for the collaborative roles of administrators and faculties.

Almost all these barriers are elements in the planning process we outlined in Section IV. Involving faculty and students is part of Stage 1; developing program goals is the essence of Stage 3; developing measures of outcomes is the purpose of Stage 4; and so on. The point is that, if colleges can learn from the experience of the NCA, and anticipate these kinds of barriers, they develop mechanisms to overcome them in each of these stages. This clarifies once again the importance of considering the development of an assessment system as an on-going and sequential process, resolving a series of potential objections as they come up, rather than a one-shot imposed by the administration or some external group.

The good news is that, for colleges that are successful in developing a coherent assessment system, the rewards are substantial (Lopez, 1999, p. 38):
This is a story of assessment as a driving mechanism, an activity whose initial results lead to other activities designed to improve instruction, support students, generate adequate funding for institutional improvement, improve assessment itself, and — along the way — integrate faculty and administrators in common goals.

...only through serious efforts to improve their quality can colleges have the moral authority to speak truth to power.

When an institution has fully implemented a structured, systematic, and on-going assessment program, it will typically have become a student-centered learning organization that is committed to continuously improving the education its students obtain. Because its culture is focused on student learning, assessment is woven into the fabric of everything its faculty, staff, and students regularly do.

These are institutions that fully deserve to be called learning colleges, or teaching institutions. These colleges are capable of self-improvement and reform; they can meet the challenges of external accountability since they have developed the systems of internal assessment necessary to respond in ways consistent with their own goals and missions.

Furthermore, in such institutions the assessment system may itself lead to improvements in other systems of the college. The NCA case study of GoodPlace Community College — not a real college, but a composite of the successful colleges in the NCA²⁷ — illustrates how changes in assessment can cascade throughout an institution. From an initial Assessment Committee co-chaired by the dean of Teaching and Learning, the college then developed a Core Curriculum Subcommittee to define the goals of the gen ed program; a Program and Curriculum Review Committee to examine curriculum design and effectiveness in the light of assessment results; other committees charged with evaluating progress, testing, advising, portfolios covering core outcomes, and capstone courses. Funding for assessment increased over this period, with new positions designated for an assessment coordinator, and a director of testing and job placement. Support services were enhanced, especially by establishing a Testing, Career Advising and Placement Center, and then a Faculty Development and Innovation Center to provide on-going staff development for faculty — one of our favorite innovations to enhance the quality of teaching and learning.²⁸ Finally, the Office of Institutional Research and Development expanded, partly to provide data for program review and partly to secure grants for faculty development. This is a story of assessment as a driving mechanism, an activity whose initial results lead to other activities designed to improve instruction, support students, generate adequate funding for institutional improvement, improve assessment itself, and — along the way — integrate faculty and administrators in common goals.

To reinforce what should be obvious by now, none of this can happen in compliance mode. Only if colleges are making serious efforts to improve the quality of teaching and learning through assessment will it become clear what the benefits of improving teaching and expanding student support are. And only through serious efforts to improve their quality can colleges have the moral authority to speak truth to power.
As we argued in the introduction, external accountability mechanisms are lurking behind most efforts to improve student and institutional assessment. These accountability mechanisms vary enormously in their structure and quality. Some — like the development of student learning outcomes required by accrediting associations — rely heavily on college-level initiative and development, rather than imposing specific measures or top-down approaches. Others — WIA’s performance measures are the best examples — specify precise methodologies for calculating performance measures, with no allowance for local conditions, existing data collection mechanisms, institutional missions, or the extent of participation. Some, like the performance-based funding initiatives around the country, have real consequences; others, like the Performance for Excellence program in California, do not link college-level performance to any particular consequences, and so create a kind of moral imperative but not concrete incentives. Some accountability mechanisms require outcome measures that almost everyone values (even though their measurement may pose technical difficulties), like transfer rates. Others — the Academic Performance Index in California and the Annual Yearly Progress measures of the federal No Child Left Behind legislation for K-12 are good examples — require assessments that many educators consider invalid measures of student learning; that are unreliable because they are measured with considerable error, unrecognized in calculating year-to-year differences; that distort the missions of local schools by forcing them to concentrate on limited approaches to restricted subjects (English and math); and that often distort classroom pedagogies by forcing instructors to teach the skills measured by simple multiple-choice tests. Poor-quality accountability systems often lead to resistance and compliance mode, and cannot possibly improve schools and colleges in meaningful ways over the long run.

The improvement of external accountability systems is therefore an issue to take seriously, not only by the state and federal agencies and accrediting commissions that create external accountability, but also by the colleges that are its subjects. Only if there are feedback mechanisms from local colleges that government agencies and accrediting associations take seriously can the experiences with accountability be used for improvement.
...another benefit of serious efforts to develop assessment systems is that they could lead to a more collaborative partnership between colleges and the agencies that are trying to hold them accountable.

from the federal government, where the Department of Labor refused to modify its regulations in response to a flood of complaints about the proposed WIA performance standards, and where the Department of Education has been unresponsive to the mounting complaints about No Child Left Behind. No doubt there are similar state-level experiences, but our point should be obvious: just as the development of effective assessment systems requires a collaborative process, with extensive communication within a college, so too the development of effective external accountability requires a collaborative process with adequate communication.

In such a system, colleges should have greater authority to critique top-down accountability if they have made genuine efforts to develop assessment systems and use them for improvement. Then they will have real information about what assessment accomplishes and what it fails to accomplish; about what its costs are in both money and opportunities lost as well as its benefits; about the distortions caused when accountability mechanisms take certain forms, or fail to account for local economic conditions, or student characteristics, or local missions. And so another benefit of serious efforts to develop assessment systems is that they could lead to a more collaborative partnership between colleges and the agencies that are trying to hold them accountable. If movements for accountability continue to expand in the coming decades, if resources continue to be scarce and demand for places overwhelming, if the pressures to move students through colleges more successfully continue to increase, then such a collaboration will benefit colleges, instructors, and students. But if it proves impossible to speak truth to power, assessment will descend into compliance, and then no one will benefit.
Endnotes

We stress throughout this report that assessments, information, and data can be either quantitative or qualitative. It is inappropriate, for example, to think of data as only quantitative. There are also methods of converting essentially qualitative judgements into quantitative scales, for example in developing scoring rubrics with a scale of 1 to 5 for the quality of written essays.


We have borrowed this phrase from Michael Kirst. Some policies — perhaps compensatory education under Title I, at least until No Child Left Behind — have benefited from social learning, as have some states. The chaotic political conditions in California are antithetical to social learning, particularly with term limits.

“Speaking Truth to Power” is the title of Wildavsky’s (1979) analysis of policy analysis.

Our prior efforts to clarify the requirements of these sources of accountability, and to suggest both short- and long-range responses, are contained in Grubb and Badway (1999).

These documents are available from the WASC AACJC website, www.aacjc.org.

These refer to the standards formulated by WASC, available on their website. Four standards encompass eleven different sections.

The API is measured for a school, as an average of student scores, and changes in the API may therefore reflect changes in the school’s composition rather than improvements in the school’s teaching. The tests incorporated into the API have been constantly changing, so improvements in scores are unreliable. There are no student incentives to do well on the API tests, which are used only for holding schools accountable; many schools report that students put little effort into completing the tests once they understand that they do not individually benefit. The scores of individual schools have random components, and these random components (or standard errors) are even larger for changes than for any one year’s scores; but the state treats change scores as precise or non-stochastic.

See especially Adelmann and Elmore (1999); and the various chapters in Carnoy et al. (2003). We know that community colleges don’t like to rely on information from K-12 education, but the fact is that external accountability as been around for much longer in K-12 education, and there has been infinitely more research on its empirical consequences. There has been very little for community colleges. See also the principles outlined by Shulock and Moore (2002), a framework based on reviewing the literature on accountability mechanisms in other states but not on empirical research about the effects of accountability.

For evidence from K-12 see Elmore (2003) and Lemons et al. (2003). In Grubb’s analysis of teaching, community colleges with top-down leadership were almost invariably described as “grim” places where faculty felt isolated — the very opposite of colleges with internal accountability. The notion of distributed
leadership has become powerful in K-12 education, where it has become a widespread conceptual underpinning for empirical work on school leadership.

Again, K-12 research has documented the importance of trust (Bryck and Schneider, 2002) and of coherence (Newman et al, 2001) to the effectiveness of schools.

The issues surrounding credentialing versus other interpretations of the effects of schooling have been endlessly debated; see Grubb and Lazerson (2004), Ch. 7 for a review. There is no strong evidence that signaling or credentialing explains the majority of the earnings effects of schooling, despite the persistence of these views.

Among the most useful materials I have identified so far is the manual of the Middle States Commission on Higher Education (2003). See also the web site on both student and institutional assessment from the New England Association of Schools and Colleges, http://209.113.248.220/assessment, particularly the Self Study Clues and examples.

See the CCSSE website at www.ccsse.org, from which the survey itself is available as well as a description of the 5 institutional dimensions measured by the survey.

See NCPI (1999), Chart 4, clarifying that relatively few colleges and universities (one quarter or less) have studied the relationship between course-taking, teaching methods, or advising on student performance. See also Project DEEP, co-sponsored by AAHE and NSSE, which will draw a sample of colleges with high graduation rates and high scores on the NSSE sub-scales, apparently assuming but not demonstrating causality between the two. See www.iub.edu/~nsse/html/deep/deep_project_announce.htm.

For notions of equilibrium, especially between teacher and student, see Harkin and Davis (1996) and Grubb and Associates (1999). This four-part model is explicitly presented in Lampert (2001), Ch. 3, Cox (2004), and briefly in Cohen, Raudenbusch, and Ball (2003).

For example, see the discussion of Assessment Committees in the materials of the North Central Accrediting Commission, www.ncacihe.

We note that institutional research takes different forms on different campuses. Sometimes it does almost no research, and instead manages data for required reporting; and some colleges do not have any institutional research. In this report, we think of institutional research as a function that could carry out a variety of tasks extending from conventional reporting to state and federal governments, to sophisticated data management, to research about the progress of students and the effectiveness of different programs. But assessment and institutional research should not be considered the same, even though institutional researchers may be important to assessment.

For example, as in footnote 7 above, the tests for the API provide incentives for schools to improve, but they do not affect student learning, grades, or progress, so that some students simply blow them off.
In addition to WASC standards see also Beno (2004). The general WASC expectations are that “an institution of higher education engages in continuous improvement of institutional effectiveness”, with the emphasis on the institution itself formulating the cycle of continuous improvement.

The analogue in K-12 education is the recent demand for “data-driven reforms”, assuming that principals have the capacity to interpret data with their faculty; see Grubb, Furco, and Tredway (2003).


These results are available on the NCA website, www.ncacihe; see especially Lopez (1999) and (2000), other papers by Cecilia Lopez, and the Goodplace Community College Case Study. No other accrediting association appears to have distilled its experience in the way that NCA has. In addition, see the survey of assessment undertaken by the National Center for Postsecondary Improvement (1999). While this survey found large numbers of institutions collecting data on different student assessments (Chart 1), institutions reported that they are not using these assessments for academic decision-making, and they believe the information has little or no impact on institutional performance. In terms of the stages we presented in Section III, some colleges seem to have progressed partly through Stages 4 and 5, but few have made it to stage 7.

See NCA (2002), which presents this matrix, plus Lopez (2000) about its use.

Some of the accrediting associations have gathered the stories of individual instructors who have become experts in the assessments appropriate to their own subjects; see, for example, the North Central Accrediting Commission website, www.ncacihe.

Lopez (1999), p. 8, amplified in pp. 9 – 36, with attention to which of these problems can be corrected through staff development and workshops and which require changes in attitudes and institutional culture.

Available at www.ncacihe.

See Grubb and Associates (1999), pp. 312 – 313. A review of such teaching centers, for both two- and four-year colleges in the U.S. as well as other countries, can be found at the web site of the University of Kansas Center for Teaching Excellence, www.ku.edu/~cte/resources/websites.html.

See also Shulock and Moore (2002, p. 7): “Accountability should be two way: institutions are accountable to state policy makers for performance as set forth in the accountability framework and policy makers are accountable to institutions for maintaining a consistent policy focus on the issues and priorities in the framework.”
References


Middle States Commission on Higher Education (2003). *Student Learning Assessment: Options and Resources*. Philadelphia: Middle States Commission on
Higher Education.


