DOCUMENT RESUME

ED 340 758 TM 018 014

AUTHOR Loacker, Georgine
TITLE Designing a National Assessment System: Alverno's Institutional Perspective.
SPONS AGENCY National Center for Education Statistics (ED), Washington, DC.
PUB DATE Nov 91
NOTE 79p.; Commissioned paper prepared for a workshop on Assessing Higher Order Thinking & Communication Skills in College Graduates (Washington, DC, November 17-19, 1991), in support of National Education Goal V, Objective 5. For other workshop papers, see TM 018 009-024.
PUB TYPE Reports - Descriptive (141) -- Speeches/Conference Papers (150)

EDRS PRICE MF01/PC04 Plus Postage.
DESCRIPTORS Academic Achievement; *Accountability; *College Students; *Educational Assessment; Educational Change; Educational Improvement; Evaluation Methods; Evaluation Utilization; Higher Education; National Programs; *Outcomes of Education; Program Design; Program Evaluation; Psychometrics; *Student Evaluation; *Systems Development
IDENTIFIERS *Alverno College WI; National Education Goals 1990; Performance Based Evaluation

ABSTRACT

Lessons from the 18-year institutional experience of Alverno College in Milwaukee (Wisconsin) with assessment for the improvement and verification of student learning are summarized. Principles are described that should inform a national assessment system with the dual purposes of improvement and accountability by incorporating the following key elements: (1) public abilities/outcomes and developmental performance criteria; (2) multiplicity of performances across varied contexts; (3) feedback and opportunities to interpret information received; (4) relation to instruction; (5) analysis of patterns of change over time; (6) provision for research and evaluation; (7) a context that supports assessment; and (8) a supporting conceptual framework of explicit educational values, assumptions, and principles; an articulated assessment theory; and an articulated psychometric theory. These elements are considered in light of the Alverno College assessment program. Designers of a national system should attempt to include these essentials to assure improvement in education. One three-page figure outlines the Alverno system. A 99-item list of references is included. Five appendices give additional information about assessment at Alverno. Reviews by E. M. Greenberg, M. A. Miller, and M. L. Tenopyr of this paper are provided. The author's response to these reviews is included. (SLD)

DESIGNING A NATIONAL ASSESSMENT SYSTEM: ALVERNO'S INSTITUTIONAL PERSPECTIVE

Georgine Loacker

Alverno College
Milwaukee, Wisconsin

Paper commissioned by the U.S. Office of Education,
National Center for Education Statistics,
in response to the National Education Goals Panel:
America 2000: An Education Strategy.

Georgine Loacker is director of the Council on Student Assessment and coordinator of the Communication Department at Alverno College. As a professor of English at Alverno, she continues to be a major contributor to the development of its ability-based education and student assessment process begun in 1973. She continues to contribute to assessment theory and to design in-course and cut-of-course assessments of student abilities. In August of 1989, she conducted workshops and seminars on assessment in Australia and New Zealand for six weeks as a Visiting Fellow of the Higher Education Research and Development Society of Australasia.

She writes and conducts research on the process of assessment of individual students. Of her publications, the most recent are "The Power of Performance in Developing Problem-Solving and Self-Assessment Abilities," Assessment and Evaluation in Higher Education, University of Beth, Summer, 1988, "Faculty as a Force to Improve Instruction Through Assessment" in J. McMillan, ed. Assessing Students' Learning, Jossey-Bass, 1988, and, with others, "Problem Solving at Alverno College" in D. Boud and G. Feletti, eds. The Challenge of Problem Based Learning, Kogan Page, London, 1991.

DESIGNING A NATIONAL ASSESSMENT SYSTEM: ALVERNO'S INSTITUTIONAL PERSPECTIVE

Georgine Loacker

Alverno College
Milwaukee, Wisconsin

ABSTRACT

The purpose of this paper is to summarize what we have learned from the 18-year institutional experience of Alverno College with assessment for the improvement and verification of student learning and, from that summary, to infer principles that should inform a national assessment system. This paper contributes to the larger purpose of developing a process to assess higher order thinking and communication skills of college graduates in support of the National Goal of Literacy and Adult Learning.

The major argument of this paper is that a national assessment system should aim to achieve the dual purpose of improvement and accountability. It can do so by incorporating the following key elements:

- public abilities/outcomes and developmental performance criteria
- multiplicity of performances across varied contexts
- feedback and opportunities to interpret information received
- relation to instruction
- analysis of patterns of change over time
- provision for research and evaluation
- a context that supports assessment
- a supporting conceptual framework of explicit educational values, assumptions, and principles; an articulated assessment theory; and an articulated psychometric theory

Such an assessment system would involve a set of durable abilities with central definitions to be adapted/elaborated in individual contexts.

The paper addresses its topic by setting forth:

1. A brief summary of the Alverno program
2. What we have learned from nearly 20 years of having the program in operation and how the principles learned can contribute to the design of a national assessment system—with accompanying implications and questions and map that guides the reader throughout is Figure 1, pp.8-10.

The conclusion, proceeding from a cumulative set of recommendations, asserts that the designers of a national assessment system should take an admittedly difficult step. They should attempt to build into the system the essentials that we have discovered assure improvement in education, which is, after all, the reason for accountability.

DESIGNING A NATIONAL ASSESSMENT SYSTEM: ALVERNO'S INSTITUTIONAL PERSPECTIVE[1]

Introduction

What we have learned at Alverno College through 18 years of implementing an ability-based curriculum and what important suggestions that might have for a national assessment system is what this paper is all about.

Because the abilities assessed at our institution include those specified in the national goals, an underlying assumption of this paper is that the general elements[2] that we find have contributed to the student development of those abilities can be operationalized in a broader context. While the practices of any one institution are not generalizable to other contexts, including a national assessment system, the underlying elements and principles are likely to be informative, useful, and potentially shared.

Another assumption of this paper is that its readers accept the idea of a national assessment system rather than a single national test, that they see improvement of learning in terms of individual student development as the ultimate goal of national assessment, and that they are in the process of considering what it means to assess student ability. Such a system would involve a set of durable abilities with central definitions to be adapted/elaborated in individual contexts.

This paper also shares an assumption with that of other papers commissioned for this project: that the authors should widen the lens, and review all the elements we believe should be part of a national assessment system. While we should identify implications, issues, and questions that flow from our recommendations, we should not prematurely impose limitations or feasibility criteria that could limit either a national vision or a set of national opportunities. This paper then sets aside explicit concern for all the difficulties of implementation in favor of broadening the scope and potential of a national assessment system.

I. The Alverno Program

Since 1973, the Alverno curriculum has been a performance-based, outcome-oriented approach to liberal arts education. To earn a degree, a student demonstrates eight broad abilities: communication, analysis, problem solving, valuing in decision-making, effective interaction, global perspectives, effective citizenship, and aesthetic response at increasingly complex levels (See Appendix A), in a wide variety of settings and contexts (Alverno College Faculty, 1985a, 1985b).

The general education courses that students take provide them the opportunity to develop and demonstrate each of the eight abilities at the core of the curriculum. Requirements for different areas of study ensure that students take a breadth of courses and are able to use their abilities in varied disciplinary and interdisciplinary contexts. Throughout their study in major and minor areas, students continue to develop abilities identified as learning outcomes by faculty in the discipline areas. These outcomes, which are distinctive to each major and minor, relate to and extend general education abilities (See Appendix B for examples).

---

[1] This paper was written in relationship to another paper commissioned by NCES in order to enable the authors to focus on two separate areas identified by the Center, and at the same time to maximize the space allotted and to respect our readers' patience. In collaborating in writing the papers, we aimed to establish logical relationships among the principles, recommendations, issues, and questions we set forth for a national assessment system. This paper's companion piece, *Developing a National Assessment System: Assessing Abilities that Connect Education and Work* by Marcia Mentkowski, assumes many of the principles learned and consequent recommendations set forth in this paper and expands them to include relation to the world of work. Because Mentkowski's paper provides detailed data to support the conclusions report here, it is best to read her paper in conjunction with this one.

[2] These elements have been documented, researched, and disseminated in varied reports and articles.

Students are assessed, on the basis of explicit, public criteria for their ability to demonstrate specified levels of each ability for their general education and more advanced levels according to their major.

Generating Abilities and Performance Levels.

The eight abilities and their levels were, and continue to be, identified by the faculty through an extensive dynamic process. This initially involved a thorough review of the literature, which continues as an ongoing means of refining the abilities and levels. Basically, the identifying of outcomes is a careful process of induction out of disciplinary and pedagogical expertise of the faculty. Also inherent in the process is continuing analysis of student performance on assessments.
Faculty also examined the existing curriculum in each discipline. Traditionally, each department had described its curriculum as a structure of **knowledge**, beginning with basic general concepts and progressing toward more complex and specialized studies. This time, the faculty worked from the assumption that there is also a progression of **abilities** implicit in the movement from introductory survey to advanced seminar. The focus, then, was to discern the developmental patterns already embedded in the normal curriculum of the disciplines, rather than to redefine those fields or to create a whole new curricular structure.

The process to articulate outcomes expected in major and minor fields, including professional areas, began with and continues to include a comprehensive review of the literature. In this case also the faculty specified the outcomes out of their disciplinary and pedagogical expertise, supplemented in the professions by direct experience in areas like nursing and business. It was also supplemented by studies of Alverno alumnae, outstanding professionals who are not Alverno graduates, and interviews with off-campus employers in various fields (DeBack & Mentkowski, 1986; Mentkowski, 1988; Mentkowski, O'Brien, McEachern & Fowler, 1982; Mentkowski, Rogers, Deemer, Ben-Ur, Reisetter, Rickards & Talbott, 1991; Schall, Guinn, Qualich, Kramp & Schmitz, 1984).

Assessing Student Abilities

Students demonstrate their abilities through the assessment process, a key component of the curriculum. At Alverno, assessment is both a way to measure student development and an aid to student learning. It represents a broad, individualized view of the learner's progress. It is "a multi-dimensional attempt to observe and judge the individual in action" (Alverno College Faculty, 1985a). Its function is not simply to rate or classify students but rather to assist them to gain insight into their abilities and direction for their further learning.

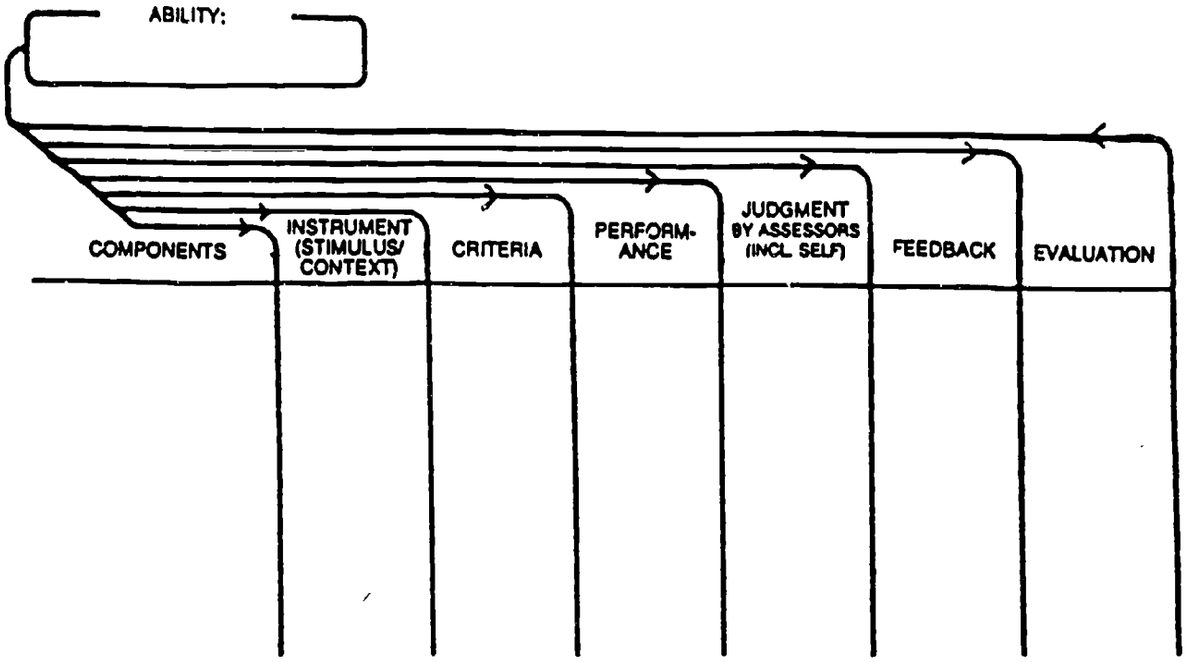
Throughout their academic work, students engage in assessments designed by the faculty; some are parts of specific courses and others are part of the general curriculum but outside their course work, thus incorporating concepts and levels of ability learned in multiple courses. Many are specific to fields of study; others are designed for all students. Often, assessments involve simulation; in all cases, they provide samples of behavior that are measured against explicitly stated criteria and followed by feedback. Although faculty primarily serve as assessors, seasoned professionals from off campus assist as external assessors of student performance. Approximately 400 members of the urban community from business and professional areas serve as volunteer assessors (Alverno College Faculty, 1984). These assessors participate in a training program, designed and implemented by the faculty, that continues for them as long as they assess. Through it, they continually refine their ability to interpret the criteria designed by the faculty, to exercise judgment on student performance, and to provide meaningful feedback.

Introduction to the academic program for entering students begins with a day-long assessment, which helps to identify each student's level of communication abilities. It provides information to be used diagnostically in advising students. After entry, ongoing assessments provide information that is used to diagnose as well as to give credit for a student's progress.

All assessments incorporate the elements identified in the following principles (Alverno College Faculty, 1985a):

1. Assessment is an integral part of learning.
2. Assessment must involve a sample of behavior.
3. Assessment must involve a performance of an ability representing the expected learning outcomes of a course, a program, a department, and/or the institution.
4. Assessment involves expert judgment based on explicit criteria.
5. Assessment must incorporate structured feedback.
6. Assessment must occur in multiple modes and contexts.
7. Assessment must incorporate an external dimension.
8. Assessment is cumulative.
9. Assessment instruments must incorporate open-ended possibilities for demonstrating a given ability.
10. Self-assessment must be an essential part of assessment, as well as a goal of the process. It is an essential ability for the autonomous lifelong learner.

Alverno faculty have designed a generalized model (Alverno Faculty, 1985) describing the flow of the assessment design and implementation process, assuring the inclusion of crucial elements and feeding back into an evaluation of each aspect:



In our publications, we have provided detailed descriptions of how a faculty member and a group of faculty might use this generalized model (Alverno Faculty, 1985; Loacker, Cromwell & O'Brien, 1986). When any of us design an assessment, we clarify what ability we are asking the student to demonstrate. We identify what components of the ability would be included in order to provide more focus for the design of the stimulus.

Once we design the stimulus—whether it is a question or a set of directions, whether it will include something like a videotape—we determine more specific criteria. Then we use the stimulus with students and end up with a set of performances. We ask each of the students to judge her performance on the basis of the identified criteria. Then we judge the performances and give feedback that tells the students which criteria they met; which they showed deficiency in meeting, with evidence to clarify why and how; what they might have demonstrated that went beyond the criteria, and what they need to do further. Finally, our study of the student performances assists us to evaluate the instrument and our own teaching in relation to it. Did the stimulus work? Were the criteria clear and sufficient? Was there some aspect that we did not teach? That we did not give the students sufficient practice in?

For every assessment that faculty design, whether an individual one within a course or a more comprehensive one within the student's total academic program, they include all of these elements even though they might not always work with them in the same order.

Use of Assessment to Evaluate and Improve the Curriculum

This ability-based assessment process generates the evidence that students are learning the abilities. Through a continuous improvement process, assessment results are sampled, studied, and analyzed to provide information for the refinement of abilities, levels and criteria, assessment techniques, and learning strategies. Thus, faculty analyze samples of student performance and synthesize results within and across groups so that they can make practice-based observations of student performance in the curriculum.

These curriculum evaluation activities are the responsibility of individual faculty in regard to their in-class assessment. For example, a faculty member may analyze student performance on a particular ability across the assessments in a particular course, by individual and by group, in order to judge the effectiveness of instruction. This process can give the faculty member a picture of the pattern of performance criteria students are meeting over time, to plan further instruction as well as course revision.

Thus, the assessment process generates continuous performance data on the degree to which students meet performance criteria across and within different levels of abilities, within the context of general education, the disciplines, and professional areas. This data is the basis for a number of analytic strategies carried out at the classroom, department, and cross-college levels for a number of purposes: assessment for individual student development, credentialing, and continuous course and curriculum evaluation and improvement.

In addition, structures are designed in support of the curriculum, for each ability and for assessment in general, to assure the collaborative carrying out of the responsibility of using assessment results to evaluate and improve the curriculum. These structures include the Assessment Council, a group of faculty specialists in performance assessment who meet weekly with staff assessment specialists and undertake various studies. They also include interdisciplinary departments representing each ability, as well as the regular discipline departments. Another one of these structures is the Office of Research and Evaluation (Alverno College Faculty, 1985a).

Demonstrating the Value, Impact, Validity, and Effectiveness of Student Outcomes of College in Relation to the Curriculum and Post-College Performance

A corollary process at the institutional level[3] steps back from the performance assessment system to research and evaluate the value, impact, validity, and effectiveness of the system itself, the curriculum as a whole, and the broad outcomes of college.

This dynamic process examines the validity of the assessment process and the validity of performance assessment techniques designed by the faculty. This process concentrates on demonstrating that changes in the development of student and alumna abilities/outcomes over time are related to the curriculum, while controlling for changes due to background factors, maturation, or other aspects of the college environment. This process also compares student performance in the curriculum over time to external standards drawn from a variety of sources. These include comparison of student performance to disciplinary and professional standards, including the criteria of external credentialing groups; to effective alumna performance in work, personal life, service and citizenship; to the perceptions and performance of outstanding professionals, and to descriptions of what is possible for humans to achieve across the lifespan. Alverno chooses student/alumna outcomes as the criterion, because student learning is at the heart of and central to the mission of the institution and the internal criterion for its effectiveness. The performance of its alumnae as lifelong learners in work, service, and citizenship roles is the primary external criterion of institutional effectiveness.

Results from these comparisons enable faculty to make judgments about the validity of their educational assumptions and principles and their assessment theory and practice. Results enable them to judge the impact of the curriculum, and the effectiveness of the institution, and engage in questioning the values that underlie the institution's mission. Further, these results enable external educators and other groups to judge the credibility, integrity, validity, and impact of the Alverno curriculum.

Thus, this "institutional assessment system" provides not only information faculty can use to improve the educational process, but also information that can assist outsiders to make independent judgments. In fact, by hosting semiannual seminar days and annual workshops, and by organizing and facilitating multi-institution consortia (three externally funded ones since 1983) who work at the College for extended periods, the College opens itself to scrutiny and judgment. Presentation, publication, consultation, and commissioned reviews also enable outsiders to examine college practices and its research and evaluation results.

---

[3] In 1976, Alverno institutionalized its research and evaluation function, a dynamic system that yields information necessary for program improvement, demonstrating quality and effectiveness and researching educational assumptions. The Office of Research and Evaluation is expected to generate evidence that tests, investigates, and examines Alverno's educational philosophy, principles, and practices and to also contribute to higher education research, evaluation, measurement, and institutional assessment. (The Office of Research and Evaluation is budgeted at close to three percent of the College's educational and general budget and reports to the faculty and the Vice President for Academic Affairs. The approach calls for an interdisciplinary team of research staff who collaborate with faculty, an interdisciplinary committee made up of senior faculty and administrators and chaired by the Director of the Office, and an external advisory panel.)

The Office is responsible for (a) demonstrating the value, impact, validity and effectiveness of student abilities/outcomes in relation to the curriculum, and in relation to the expectations and needs of business, industry and community institutions, and professions, so that graduates can fulfill their responsibilities in work, personal life, service, and citizenship. Office goals also include (b) initiating and maintaining the quality of research and evaluation as a concept and function at the College, (c) contributing to program, student, and faculty development, (d) establishing Alverno as an accountable educational institution for its various constituencies, and (e) eliciting constructive critique from colleagues and establishing Alverno as a contributor to higher education research and postsecondary practice.

Thus, the Office provides evidentiary support for Alverno's contributions to the advancement of undergraduate education. A more extensive discussion of the Office's simultaneous contributions to the College's own purposes and goals and to the more general purposes of educational research and postsecondary practice is found in Mentkowski, Rogers, Deemer, Ben-Ur, Reisetter, Rickards & Talbott (1991). The question discussed is, "Can findings from such intra-institutional studies add up to anything across colleges?" How can an institution meet its own purposes and simultaneously contribute more broadly? For example, this paper is an exercise in describing evidence for what we have learned that can contribute to the design and development of a national assessment system.

Engaging in this institution-wide process has meant creating a context where institutional assessment yields educational improvement (Mentkowski, 1991c). This means that assessment is a means to achieve both accountability and improvement agendas. To meet these dual demands, assessment approaches encourage a multiplicity of approaches within a larger system and pay careful attention to how these approaches link into and flow from the purposes and goals served...the context for the enterprise. At the same time, the process encourages coherence by re-examining the explicit and implicit links between educational goals and student outcomes. Finally, feedback is the essence of assessment. It is the catalyst for investment by participants in all of its phases, especially using results to improve. Developing the institutional assessment process meant making a commitment to a dynamic plan and a process that is realized across a number of years of effort. This means relying on educator input every step of the way, creating interactive processes where everyone who has a stake in the enterprise becomes involved, and defining public criteria and standards against which judgments of the "good" are made. It also means translating the results into "live" information that can be easily interpreted. It means creating feedback that relies on more than one data source, that focuses on patterns to encourage the broadest possible implications, that is developmental and encourages productive change. Throughout, we have learned that assessment systems, whether at the level of the individual student or the institution, embody and advance our educational values.

The specific characteristics of our institutional assessment process are described in Appendix C. Approaches and strategies are explained in more detail, because much of the "how to" that is implicit in our recommendations for a national assessment system is drawn from the methods we have created.

Evidence for the Credibility and Benefit of "What We Have Learned" for External Use

Alverno's mission includes a charge to elicit from colleagues constructive criticism of Alverno scholarship and research on teaching, learning, and assessment. In this way, Alverno educators hold themselves responsible and accountable for a continuing contribution to the advancement of undergraduate education.

Thus, the College documents evidence of opportunities to disseminate its findings and to open itself to critique. The number of citations in the literature and collaborations and consultations with other institutions suggest some progress toward this broad institutional goal of contribution and eliciting critique. For example, since 1973, there have been a total of 3,232 individuals from 894 institutions who have visited Alverno for at least a day or up to 10 days for in-house workshops. Since 1978, 20,132 copies of books about Alverno's philosophy and educational frameworks have been disseminated, excluding reprints or Office of Research and Evaluation publications. In 1990 alone, 4,278 publications (including reprints but excluding Office of Research and Evaluation publications) were disseminated.

The Office of Research and Evaluation reports like documentation on the degree to which the Office met similar goals from 1977 to 1987 (See second edition of Mentkowski & Doherty, 1984b). The Office disseminated 19,800 copies of five major articles and chapters developed from the research outcomes that were also distributed externally by outside publishers. Research outcomes were described or cited in 14 news articles and at least 56 outside publications.

From 1977 to the present, the Research and Evaluation staff created 104 publications and made 280 presentations. The Office reached over 2,000 institutions and representative departments in all 50 states and 33 countries through these presentations, together with countless publications distributed during presentations, and 3,528 publications mailed upon request.

This documentation is some evidence of eliciting critique and of contribution related to that part of the College's mission to examine whether and how Alverno frameworks contribute generally to undergraduate education. It is also some evidence that the research and evaluation efforts support this larger contribution.

II. What We Have Learned and How the Principles Learned Could Contribute to a National Assessment System

A number of principles that Alverno faculty have learned through their practice have implications for the design of a national assessment system. These principles continue to be confirmed by the demonstrable success of Alverno students in developing abilities and by the ongoing articulation of the self-renewing system that makes these principles operative. Figure 1, in addition to providing a "map" of this paper, applies these principles to a national assessment system design, through recommendations, implications, issues, and questions.

What have we learned? From a serious and very long look at our practice, we would point out the following as relevant and significant learned principles:

1. An ability-based performance assessment system, with certain key elements,[4] can work both to evaluate student performance and to develop student knowledge and ability.

2. Making expected outcomes explicit and public to all, identifying developmental criteria for performance, and communicating them to students ahead of time, contributes to effective performance by making learning more accessible and enabling performance.

3. Feedback on performance in relation to developmental performance criteria and the opportunity to interpret that information leads to further learning and improvement of student and program performance.

4. Students learn complex abilities, including self-sustained learning, in the curriculum through a variety of contexts.

5. Students can transfer abilities when they are assessed in contexts that are valid for what students learned and for how they will perform abilities later.

6. When an assessment system examines changes in student abilities/outcomes over time, including who changes and why, and relates those changes to the curriculum, the system yields information necessary for meaningful improvement.

7. We can validate an ability-based performance assessment process, and institute an instrument validation process that gradually improves instrument validity. We can establish the educational value, impact, validity, and effectiveness of the abilities/outcomes.

8. A dynamic assessment system incorporating input from and feedback to faculty, as well as administrators, provides for the effective use of information to keep abilities, performance criteria and standards responsive to and in advance of the needs of our society.

9. Creating a context for assessment is as important as creating the assessment method.

---

[4] Public abilities/outcomes and criteria, multiplicity of performances across varied contexts, expert judgment, feedback, and self-assessment

10. The effectiveness of an assessment system concerned with the improvement of learning depends partially on a coherence that comes from the following articulated components:

    - educational values, assumptions and principles that are tied to the mission statement of the institution

    - an assessment theory (what are the components of good assessment?) consistent with those values and assumptions

    - a psychometric theory (how do we best measure and credential performance and give feedback to students on their abilities?) consistent with those values and assumptions

Figure 1. Designing a National Assessment System: Alverno's Institutional Perspective

WHAT HAVE WE LEARNED?	RECOMMENDATIONS	IMPLICATIONS	ISSUES AND QUESTIONS
1. An ability-based performance assessment system, with certain key elements,* can work both to evaluate student performance and to develop student knowledge and ability. • Meeting "exit" standards can be effectively combined with individual student development as criteria for excellence. • Incentive and feedback elements can be effectively combined to ensure that students are invested in performing their best, and can receive feedback that they can use to improve. • Both an accountability and an improvement agenda can, therefore, be met with the same system.	1. A national assessment system should include ability-based performance assessment, with certain key elements,* so that the system provides for individual student development, as well as evaluation of performance; in other words, to assure improvement as well as accountability.	1. Designers and implementors of a national assessment system will need to find a way to make performance assessment work on a national basis. 1. A major challenge will be to provide for accountability without eliminating freedom to take the risks and learn from the failures that are necessary for the development of the learner, whether individual or institution.	1. How do we know we have the right abilities/outcomes? How do we make them integrated, developmental, and transferable? 1. How do we create criteria? 1. How do we sample student performance? when? how often? 1. Can performance assessment be cost-effective? 1. How do we synthesize information from individual assessments to aggregate across students, institutions? 1. How do we create developmental assessment (multiple tracking over time) used both diagnostically for placement and for assessing educational progress on broad outcomes?
2. Making expected outcomes/abilities explicit and public to all, identifying developmental criteria for performance, and communicating them to students ahead of time, contributes to effective performance by making learning more accessible and enabling performance.	2. A national assessment system should make the abilities/outcomes explicit and public and communicate them to students and faculty in advance to enable students to improve performance.	2. Institutions will need to know what their faculty-defined abilities are. 2. A national assessment system will need to link up to institutional efforts.	2. How are outcomes/abilities defined? 2. How are performance criteria defined developmentally?
3. Feedback on performance in relation to developmental criteria and the opportunity to interpret that information leads to further learning and improvement of student and program performance.	3. A national assessment system should provide • feedback at various levels (individual student, faculty, institution, state, federal, public) • structured opportunities to interactively interpret the findings and discuss the implications for improvement	3. A national assessment system that identifies student strengths and weaknesses will incur a national commitment for educational improvement.	3. How do we create a system where all types of institutions can and will use the information to improve? 3. How does feedback work to invest students, faculty, institutions, states and the public in assessment to improve learning?
4. Students learn complex abilities, including self–sustained learning, in the curriculum through a variety of contexts.	4. A national assessment system should sample student performances in relation to instructional opportunities.	4. A national assessment system that identifies strengths and weaknesses in instruction will incur institutional commitment to improve instruction. 4. Strategies will include qualitative measures such as student portfolios and a description of the learning context.	4. How do we sample student performances in relation to instruction? 4. How is assessment linked to instruction? 4. Can students perform to standard? 4. How will institutions best describe the learning context for sampled student performances?

* Public abilities/outcomes and developmental performance criteria, multiplicity of performances across varied contexts, expert judgment, feedback, and self–assessment

Figure 1 (continued). Designing a National Assessment System: Alverno's Institutional Perspective

WHAT HAVE WE LEARNED?	RECOMMENDATIONS	IMPLICATIONS	ISSUES AND QUESTIONS
5. Students can transfer abilities when they are assessed in contexts that are valid for what students learned and for how they will perform abilities later.	5. A national assessment system should define abilities and developmental performance criteria generically but assess them in contexts that are valid for what students learned and for how they will perform later.	5. Not all institutions may teach to national goals. 5. Students will need to learn how to complete performance assessments.	5. How assess in context? • How consider context of the course, program, curriculum, total academic experience? • How generic should abilities and criteria be? 5. How will a national assessment system integrate and synthesize diverse institutional abilities and criteria. 5. How define contextual validity?
6. When an assessment system examines changes in student abilities/outcomes over time, including who changes and why, and relates those changes to the curriculum, the system yields information necessary for meaningful improvement.	6. A national assessment system should link changes in student abilities/outcomes over time, including who changes and why, to student performance in college curricula and feedback the information to institutions.	6. Institutions will need to be able to marshall evidence for the value, impact, validity and effectiveness of curricula by describing what they do and their evidence for student achievement. 6. A national assessment system will need to find ways to link up with institutional efforts.	6. How do we link information from entering student abilities/outcomes to graduating student abilities/outcomes? How do we relate changes in student abilities/outcomes to curriculum? 6. What are best methods for analyzing change? 6. How do we aggregate information from institutional assessment systems?
7. We can validate an ability-based performance assessment process and institute an instrument validation process that gradually improves instrument validity. We can establish the educational value, impact, validity and effectiveness of the abilities/outcomes.	7. In order to examine the educational value, impact, validity and effectiveness of a national assessment system, designers should build in a research and evaluation component.	7. Institutions will be concerned about the educational value of a national system. All design elements will need to be planned from the start.	7. How do we design and validate an assessment process? 7. How establish the validity of instruments? 7. How define construct validity? 7. What is good evidence? 7. How validate expert judgment?

Figure 1 (continued). Designing a National Assessment System: Alverno's Institutional Perspective

WHAT HAVE WE LEARNED?	RECOMMENDATIONS	IMPLICATIONS	ISSUES AND QUESTIONS
8. A dynamic assessment system incorporating input from and feedback to faculty, as well as administrators, provides for the effective use of information to keep abilities, performance criteria, and standards responsive to and in advance of the needs of our society.	8. A national assessment system should be a dynamic system based on faculty-defined abilities, as well as other sources, to make the outcomes, criteria, and standards responsive to and in advance of the needs of our society.	8. A dynamic system will need to identify elements that change and elements that remain stable. 8. A dynamic assessment system raises questions about the meaning of validity and reliability.	8. How create a dynamic system? 8. How do we set performance levels so they reflect changes in what is being taught and what needs to be learned? 8. How define validity in a changing context? 8. How define reliability when change rather than consistency is measured?
9. Creating a context for assessment is as important as creating the assessment method.	9. Creating a context for a national assessment system that yields educational improvement should be planned for and implemented as an essential part of the process.	9. The purpose of a national assessment system will have to shift from testing-for-selection to assessment-for-improvement, in the public eye.	9. How best are students, faculty, institutions, states, federal agencies, and the public invested in a national assessment system? 9. How create a community of judgment?
10. The effectiveness of an assessment system concerned with the improvement of learning depends partially on a coherence that comes from the following articulated components: • educational values, assumptions and principles that are tied to the mission statement of the institution • an assessment theory (what are the components of good assessment?) consistent with those values and assumptions • a psychometric theory (how do we best measure and credential performance, and give feedback to students on their abilities?) consistent with those values and assumptions	10. A national assessment system should have at its root a coherent set of articulated components and principles: • educational values, assumptions, and principles underlying the national goals • an assessment theory that describes the components of "good" assessment • a psychometric theory that describes how we best measure and credential performance, and give feedback to students, faculty, institutions, states, federal agencies and the public on student achievement	10. In order to establish the integrity and credibility of a national assessment system, we will need to continuously re-examine and re-articulate its components and principles.	10. Can institutions articulate and identify shared educational assumptions and principles? 10. Do assessment assumptions and principles hold up? Are values shared? 10. How will a national assessment system with multiple purposes, functions, uses, and users contribute to coherence across educational contexts?

A discussion of each of the above statements follows, including a summary of supporting evidence, a statement of the related recommendation for a national assessment system, and important implications and questions.

Principle Learned #1. An ability-based performance assessment system, with certain key elements, can work both to evaluate student performance and to develop student knowledge and ability.

The above description of the Alverno program summarizes how cumulative series of assessments enable students both to develop and to demonstrate the levels of the abilities required to advance through the program and ultimately to graduate. Each level is marked by academic credit accompanying the successful completion of courses. When students show they have achieved a given level of ability, they still receive feedback to assist them to develop further. Throughout, this paper explains how key elements like explicit criteria and feedback are necessary to and actually operative in the program.

Evidence that faculty have evaluated student performance and confirmed student abilities—and continue to do so—exists in numerous academic records. Further evidence that students have developed knowledge and ability, and continue to do so, exists in our research reports and in records of institutional growth and influence. As an institution that has increased in enrollment 123 percent since the current program began in 1973, we daily experience the fact that an ability-based performance assessment system can work both to evaluate student performance and to develop student knowledge and ability.

On another level, we take samples of student performances from the instruments used within the curriculum to evaluate and improve institutional performance. We find that linking multiple purposes and levels of analysis within the same instrumentation preserves sought-for connections between assessment and instruction, teaching and learning, accountability and improvement. Such linking assures that the same abilities are assessed, no matter the immediate purpose. It protects students from having to take multiple assessments that may or may not affect their learning. As the remainder of this paper will continue to develop and reinforce, it is, above all, possible.

Recommendation for National Assessment System

Our over-arching experience of the program as a whole leads to the following recommendation:

Recommendation #1. A national assessment system should include ability-based performance assessment, with certain key elements, so that the system provides for individual student development, as well as evaluation of performance; in other words, to assure improvement as well as accountability.

Implications, Issues, and Questions

The recommendation incorporates all the more specific recommendations that follow in regard to individual elements of assessment. The difficulties in making performance assessment work nationally are expressed in the questions that immediately come to mind.

- **How do we know we have the appropriate abilities/outcomes?**

Our initial step 20 years ago at Alverno was to incorporate all faculty in a careful process of identifying and articulating abilities, with ongoing review of the literature and current practice. Since then, we have been creating a system that provides for continuous review and revision of the abilities on the basis of what we learn from our practice. The remainder of this paper provides examples, as well as further explanation, of how that system works.

The challenge of creating a system nationally with built-in provision for continuous improvement, of course, includes involving faculty across institutions and finding ways to assess performance contextually, to synthesize the data nationally, and to provide feedback loops to benefit students, faculty, and institutions as a whole.

- **How do we create criteria?**

Clearly, a dual agenda for a national assessment system that incorporates concerns for improvement and accountability will have at its center, criteria against which student performance in college will be judged. These criteria serve not only as a way to profile student strengths and areas to be developed; they will also need to represent the standards that society expects of its graduates.

How we create these national criteria will communicate a good deal to the judges who use them, the students who experience them, and the public who expect them to serve as a guiding light to improved performance. The institutions responsible for the degree to which students meet them, have a similar responsibility for creating the context for developing abilities in students and for articulating sets of outcomes that can serve as a basis for defining criteria.

- **How do we sample student performance? When? How often?**

The issue of how to sample student performance is key to developing a national assessment system that combines individual development with meeting "exit" standards. Nationally, faculty have made it clear on a number of occasions that expecting students to perform on complex assessments that are unrelated to their learning context is not an acceptable goal (Forrest & A Study Group on Portfolio Assessment, 1990).

Our experience suggests that it is possible to sample complex performances for individual development that can also serve as "exit" measures. We believe that the current efforts in elementary and secondary education to assess student portfolios and other kinds of performance are an indication that meeting such a goal is on the more immediate horizon. How we determine the frequency of sampling to observe patterns of development rather than merely discrete performances is difficult to achieve. No doubt this issue will remain before us in developing a national system.

- **Can performance assessment be cost-effective?**

Clearly, current efforts to assess student performance are judged <u>not</u> to be cost-effective because the information they yield cannot be used to improve curricula. With performance assessment, costs are often misleading, because much of the effort has to be expended during the design phase and in the judgment phase. This is in contrast to the way most costs for testing are currently parcelled out. In testing, design costs are also at issue. But once the measure is "created," routine administration and automatic scoring is labeled "cost-effective." Often costs associated with using the information are not included.

In performance assessment, using the information becomes part of faculty and institutional responsibility and is motivated by the improvement that users experience. Individual students also become engaged in the improvement process, no longer complaining that assessment takes too much time—their nonexpendable commodity. Benefits outweigh costs, because the assessment process becomes part of a continuous improvement agenda.

The issue of "cost and benefits" is not an easy one to address (Read, 1985). It takes effort to translate "anxiety about testing" to "confidence in assessment." Values around time and money and where it should be spent are often undisclosed and in conflict. Those who opt for the benefits may be seen as naive and unrealistic, because the start-up time is daunting, and the methods have to be worked out as one goes along. We have no magic answer to the costs issue. Suffice it to say that we have institutionalized performance assessment, continuing to improve it despite a 123 percent increase in enrollment, and the second lowest private school tuition in the state ($6,390 annually). Our students are generally first generation college students; 21 percent are minority students. Our colleagues in large and small institutions express similar questions about the cost/benefit concerns that we hear about in the national media. With us, they are making the investment because they have already experienced benefits. With us, they make no claims for broad use without extensive field tests.

Our judgment at this time is that the elementary/secondary experience in designing and field-testing performance assessment will be an important "cost-saver" for higher education. This experience will provide some more specific answers, but it is likely that the issue will continue to surface. Another helpful source will be advances in computer technology for handling complex responding by students, and complex content coding.

- **How do we synthesize information from individual assessments to aggregate across students? Across institutions?**

This question will be one of the most difficult to answer. Certainly, standard criteria should be considered as part of the answer. Still, collecting performances from local contexts presents a difficulty that is hard to surmount. To use standard criteria to make reliable judgments about performances from multiple contexts becomes more difficult when one identifies the many aspects of context that affect the performance.

Our own institutional experience with aggregating information from assessor judgments in order to transform the information in ways that can be scaled and compared has taught us that researchers could scale, in a reliable way, qualitative narrative comments that faculty have given as feedback.

Further, even when one is judging complex abilities, and different students meet criteria in different ways, this information yields differential patterns that provide differential profiles of how students met criteria, particularly when faculty judgment has included an indication of exactly which criteria each student met. The challenge is to specify abilities and—even more difficult—to determine the level of specificity for the criteria that will make criteria analysis possible, without destroying the picture of the ability represented by the criteria.

Two questions arise: (1) what kind of confidence do we have in faculty judgment and (2) what kind of confidence do we have that the criteria to be specified will reflect the ability? Dealing with these questions is a first priority. Following that are questions of finding strategies that will assist in synthesizing the information so that discussion about the degree to which students are meeting the criteria can occur in a national context.

An important related occurrence is that as the cache of student performance builds to illustrate the criteria across multiple contexts, a clearer understanding of both the meaning of the ability and the validity of assessor judgment will accrue. As the pattern of judgment from such assessment becomes clearer, it has been our experience that finding a way to synthesize also becomes easier to achieve.

Yet all of that just begins to deal with the question of aggregating information from performance assessments across students. To find a way to aggregate such information across institutions will take a great deal of inventiveness and courage to risk moving ahead on some important convictions despite the barriers of feasibility issues.

Related to this is one of the final recommendations of this paper, which involves the establishing of a national center that would train and validate expert judges who could sample performances, work to make their judgment process valid and reliable, and investigate these issues for the benefit of those assessing and promoting learning on the local level.

- **How do we create developmental assessment for diagnostic use and for assessing educational progress?**

The question about how we create developmental assessment (multiple tracking over time) that can be used diagnostically for placement and for assessing educational progress on broad outcomes is a challenging one. Clearly, we will not necessarily be using the same measures for assessing an ability at each stage in a student's college career. It is often not helpful to ask students to complete assessments where they cannot perform, just to get "proof" that they cannot perform the abilities. We can, however, collect a range of performance samples that will enable us to apply developmental performance criteria, and to get a picture of student progress over time.

An important issue in that discussion will be another major challenge presented by combining the two purposes of assessment in one set of instrumentation: to provide for accountability without eliminating freedom to take the risks and to learn from the failures that are necessary for development and improvement. The possibility of making the profound conceptual shift that lies at the heart of that challenge is confirmed by our experience with our students, who eventually learn that their success is not dependent upon a single performance, that the quality of their achievement is not dependent on its comparison to the achievement of others, and that a string of apparent successes does not necessarily constitute growth or improvement.

Principle Learned #2. Making expected outcomes/abilities explicit and public to all, identifying developmental criteria for performance, and communicating them to students ahead of time, contributes to effective performance by making learning more accessible and enabling performance.

Making expected outcomes and developmental criteria public renders Alverno faculty accountable to students and puts them into a dialogue with each other and with colleagues in their field throughout the academic community. That dialogue leads to ongoing development of understanding on the part of all involved, of what should be learned and how it should be learned. Explicit outcomes and criteria enable students to try out performances and strategies to improve them. They enable faculty to relate learning experiences in history or science to others within those areas, as well as in areas like nursing or business or philosophy (Loacker, 1988; Loacker & Palola, 1981).

Evidence that the performance of students is affected by their knowledge of expected outcomes and criteria comes from varied sources: day-to-day student self-evaluations within individual courses as well as the overall program, instruments like an inventory of learning strategies used by first-semester students, departmental reviews, and longitudinal research.

Student Self-Evaluations

Student self-evaluations over time, as well as self-assessments on specific performances, are a commonly used faculty strategy to sample student reflections on their learning. These are gathered through journals, essays, or interviews that elicit student reflections on what they have learned and what learning opportunities they have had (Deahl, 1990; Kramp & Humphreys, 1990). Faculty also ask students to review a series of sequential performances from their assessments, to reflect on how they have developed their abilities, and to link these changes to specific curricular elements.

Analysis of these self-reflections reveal that students use abilities and criteria as a means of understanding and learning to do what is expected of them. They explicitly relate what they have accomplished to their growing understanding of what they are aiming for.

Student as Learner Inventory

Students use the opportunity to complete the Student as Learner Inventory (Alverno College Office of Research and Evaluation/Assessment Committee, 1986; Rogers, 1988) to reflect on their own approaches to learning. Through the inventory, students examine the fit between their own approaches and those identified through research on the learning patterns of successful students (Much & Mentkowski, 1982; 1984). Students also compare their learning strategies to faculty expectations gathered through research on faculty perspectives on what makes for successful learning in the curriculum.

Recent analyses of patterns of student responses gathered during their first semester have enabled faculty to identify those students whose self-descriptions of their approaches to learning may place them at risk. For example, the instrument discriminates students who do and do not understand or reflect successful learning patterns—for example, students who acknowledge or deny inconsistency in their work, accept or reject the use of criteria and feedback, and work behind or beyond specific course expectations. Therefore, the instrument provides evidence that making outcomes and criteria explicit contributes to the student's ability to construct successful learning pathways, thus making learning more accessible and enabling more effective performance. The instrument also informs faculty about the "who changes and why" question because it allows for analysis of individual differences.

Departmental Reviews

Scheduled departmental reviews use questionnaires, interviews, or panels to collect data from students, alumnae, or external groups, to study the degree to which abilities/outcomes, as identified by faculty, contribute to effective performance. While there is considerable variability in the strategies used across departments, some departments report that students describe departmental outcomes as those they have achieved (Albro, Devitt, Salem, Sharkey & Wojno, 1990) and alumnae report, through Behavioral Event Interviews, that they use these abilities in their professional positions after college (Kennedy, 1988).

Longitudinal Research[5]

As part of a longitudinal study that sampled two complete classes, the Office of Research and Evaluation conducted open ended, in-depth, confidential interviews at the end of each year in college and afterward Mentkowski & Doherty, 1984b). The research staff analyzed how students construct their abilities, learning, and development. Interview analysis identified patterns that describe student and alumna use of developmental performance criteria. Effective use of criteria was clearly demonstrated across the four years of college and linked to effective performance (Much & Mentkowski, 1984).

Student constructions of the learning process reveal a developing understanding of the role of criteria (See Appendix D). Beginning students were apt to construct criteria as vague directions for what to learn and arbitrary standards beyond their control. As they progressed, students saw criteria as pictures of the abilities to be performed. Advanced students saw criteria as flexible guides to independent learning, providing a framework for self-assessment (See Appendix E). The ability to use criteria to evaluate their own performance, i.e., to self-assess, plays a central role in the student's ability to engage in independent learning (Loacker & Jensen, 1988) and development after college (Deemer, in press; Deemer & Mentkowski, in press; 1990; Mentkowski, Much & Giencke-Holl, 1983).

Recommendation for National Assessment System

As a result of everything we have learned about the importance of outcomes being made explicit and public, we would recommend the following:

---

[5] Data sources. Results are reported from (a) curriculum-embedded performance assessments, (b) college designed instruments and interviews, and (c) a battery of 12 external measures of generic abilities, learning styles, and moral, intellectual and ego development (human potential measures). These were completed longitudinally on three occasions (1976/1977; 1978/1979; 1980/1981) by the entire entering classes of 1976 and 1977 during college (N=706), and most measures were completed again on a fourth occasion five years later. Measures of abilities, learning styles, motivation, cognitive, moral and ego development were employed along with in-depth, confidential interviews, surveys of student perceptions and background characteristics, and Behavioral Event Interviews (McClelland, 1978) of alumna performance across settings in work, personal life, and service. Student participation rates ranged from 84 to 99 percent; alumna (n=358) rates ranged form 59 to 88 percent. Data from curriculum-embedded performance assessments in the curriculum, academic reports, and a faculty rating of performance characteristics with background factors controlled, were related to changes on the battery of external measures using multiple linear regression, ANOVA for repeated measures, and path analysis. Interviews were coded via ethnographic and thematic analysis.

More specifically, the battery of 12 human potential measures and college-designed instruments were administered to two complete entering classes and one graduating class (altogether about 750 students). A subsample (n=80) completed in-depth interviews as well. The entering classes completed the same battery two years after entrance and again two years later, near graduation, and again five years later (1986/1987). Thus, we have a set of longitudinal results that can be double-checked against results from a cross-sectional study of 60 graduating seniors who participated in 1978 as seniors and 1980 as alumnae, who were compared with entering students who later graduated (controlling for retesting and attrition, with initial selection factors, such as disposition to change, probably uncontrolled). The data on students who completed the 12 external instruments on the three occasions during college provide a parallel stream of longitudinal information alongside these same students' progressive performance on five college-designed measures. The design includes two age cohorts (age 17 to 19 and age 20 to 55 at entrance) to examine the effects of maturation, and two achievement cohorts (high and low, based on number of consecutive assessments completed in the curriculum) to examine the effects of performance in the curriculum. Two class cohorts, with the second cohort analyzed for weekday versus weekend time frames, further enhance representativeness, although only further longitudinal cohorts could truly control for effects of curriculum and societal changes. The time series design holds time constant and allows performance in the curriculum to vary, so we can attribute change to performance in the curriculum in the absence of a control group of students who did not attend Alverno. As mentioned, we also control for several age, background, and program variables as well as pretest scores when we study the effects of performance in the curriculum. In studies of current students, their portfolios and other curriculum performance assessments are judged on dimensions of performance by expert judges and related to abilities that define their major.

> Recommendation #2. A national assessment system should make the abilities/outcomes explicit and public and communicate them to students and faculty in advance to enable students to improve performance.

Implications, Issues, and Questions

Inherent in the above recommendation are significant implications, one of which is: Institutions will need to know what their faculty-defined abilities are.

- **How are abilities/outcomes defined?**

We have noted that defining outcomes/abilities that make sense to students and faculty, as well as to state, federal, and public constituencies will be a central activity of a national assessment system. We believe that this issue is at the heart of developing a system that can benefit students, faculty, and other constituencies.

While we do not argue that only one definition is credible, we do argue that abilities have multiple components and that abilities are defined as integrated, developmental, and transferable. We have developed this point elsewhere (See Mentkowski, 1991d, a paper commissioned for this project), and have cited research evidence from a range of sources to support this definition.

- **How are performance criteria defined developmentally?**

Analysis of student performance quickly identifies those samples that meet criteria and those that do not. Gradually, judges begin to set midpoints, and so developmental criteria begin to emerge. Whether this process for generating criteria will work for a national assessment system is open to question. Our current experience with 11 institutions ranging from high school to medical school suggest that faculty and administrators find it fascinating to discern effective from ineffective performance, but more important, to distinguish the elements that define sequential, pedagogical criteria that enable them to teach and assess for the abilities involved. Further elaboration is found in another author's paper (Mentkowski, 1991d).

> Principle Learned #3. Feedback on performance in relation to developmental criteria and the opportunity to interpret that information leads to further learning and improvement of student and program performance.

Student Performance

When Alverno students write papers, participate in projects, or make presentations, faculty give them feedback intended to clarify how well they met the given criteria and, when applicable, if they showed some aspect of the ability that the criteria had not included. In that feedback, faculty also aim to suggest needed direction so that each student can find strategies to improve.

Our experience has consistently been that students do learn to make meaningful use of feedback in their own development. As with their use of criteria, we find frequent reference to feedback in their ongoing self-reflections as well as in our longitudinal research.

Student Self-Reflections

In addition to the systematic student self-reflection exercises described in relation to Principle #2, a good example of how students use feedback is in their portfolios. For all students, writing and video speaking portfolios collect performances across the entire curriculum. In several departments these are incorporated into portfolios that represent a student's work in the major. The portfolios are designed to show development rather than merely discrete samples. Therefore, the entries include feedback and revisions whenever possible so that the portfolios reveal whether and how students use feedback and when they are able to do independent revising of aspects that have not been pointed out in the feedback.

Longitudinal Research

In their analysis of the longitudinal, in-depth interviews, the Office of Research and Evaluation staff analyzed patterns of students' use of feedback as well as of criteria. They found that student perceptions of feedback developed from *experiencing feedback as general affirmation or rejection of themselves* to *seeing it as the provision of explicit information on their progress* to finally *expecting feedback that helps them see patterns and relationships to their performance in other areas* (Much & Mentkowski, 1984). The same analysis revealed patterns of commitment to improvement. Beginning students tended to want to improve and to know they should improve. Intermediate students showed that they think about how to improve, become aware of their weaknesses and build on their strengths. Advanced students took initiative and used resources to improve (Much & Mentkowski, 1984).

Program Performance

Ongoing feedback to faculty and academic administrators is at the heart of what makes our assessment system dynamic. The sources of such feedback are faculty analyses of student performances, ongoing faculty review of related current research in given abilities, student reflections on their own learning, collected observations of external assessors, and the longitudinal research and other studies by the Office of Research and Evaluation. Opportunities to interpret the information are built into the regular agendas of the Assessment Council, departments, and the half-day and week-long sessions that are ongoing structures for faculty development (Alverno College Faculty, 1985; O'Brien, Matlock, Loacker & Wutzdorff, 1991).

One example of how feedback works to improve the program can be seen in the ongoing process of design, implementation, and review of the entry level assessments. Since 1973, these have included faculty-designed instruments to assess student performances in reading, writing, speaking, listening, media literacy, quantitative literacy, and computer literacy; and a standardized multiple-choice reading test. As of September 1991, the assessment of reading has incorporated several changes on the basis of feedback from the process. We have eliminated the standardized test and have thoroughly revised the reading performance assessment, both of which are assessed by staff assessors trained by the faculty. Examination of ongoing results had shown very little relation between student performance on the two instruments. Also, through a regular process by which staff assessors report problems each semester, difficulties with items like main idea and fact vs. opinion continually surfaced.

At the same time, the Communication Competence Department, an interdisciplinary group that provides direction and assures quality in the learning and assessment opportunities for communication abilities for students, was doing a concentrated study of contemporary development in reading instruction and assessment. The members synthesized current research and compared the standardized and faculty-designed instruments in relation to important elements. Once they found that the performance assessment incorporated most of those elements, they revised a few of the criteria and assessment items, and they eliminated the standardized instrument altogether.

Another example of how feedback contributes to improvement of the program shows the use of data from individual studies by the research staff. Studies of faculty-designed generic instruments for communication and valuing abilities (Friedman, Mentkowski, Earley, Loacker & Diez, 1980) produced data that faculty used to rethink generic criteria at various levels of both abilities.

Recommendation for National Assessment System

Given the significant role that our experience tells us feedback plays when it is related to present and further development, the following recommendation seems inevitable:

> Recommendation #3. A national assessment system should provide
> - feedback at various levels (individual student, faculty, institution, state, federal, public)
> - structured opportunities to interactively interpret the findings and discuss the implications for improvement

Implications, Issues, and Questions

- How do we create a system where all types of institutions can and will use the information to improve?

This is our biggest challenge, but it is the one which we have observed in the emerging commitment of the assessment community (Hutchings & Marchese, 1990). More and more institutions are making the effort, and although statistics suggest that a smaller proportion have institutionalized assessment processes than are starting up, improvement is clearly on the higher education agenda.

Any efforts to create a national system can call on this motivation, but will also incur all the problems that have already surfaced nationally. In our view, how a national assessment system will fare is open to debate. Some are clearly for; some are clearly against. That is why we have made recommendations to consider the context for assessment. We refer the reader to Principle Learned #9.

- How does feedback work to invest students, faculty, institutions, faculty and the public in assessment to improve learning?

Feedback is the essence of assessment. But we have yet to demonstrate the full range of feedback strategies that will continue to invest the multiple audiences who will need to benefit from assessment information.

How different kinds of feedback work, who should deliver it, and how it links to improvements in learning are important issues for startup, and are likely to continue to be important.

> Principle Learned #4. Students learn complex abilities, including self sustained learning, in the curriculum through a variety of contexts.

Because opportunities to develop the required abilities are infused throughout the curriculum, students consciously work, for instance, to develop problem solving in art classes as well as in mathe natics, education, or nursing. They demonstrate their progress in terms of each specified level of the ability through assessments in individual courses. They demonstrate their progress in a more integrated way with increasingly complex subject matter--especially through more comprehensive assessments, based on a semester or more of learning, that range from a half-day simulation of a school board committee on the censorship of books to a week-long

art exhibit, planned, designed, advertised, and implemented by students to exhibit their own work. In these cases, problem solving is demonstrated in integrated situations that involve other complex abilities like critical thinking and aesthetic response.

In addition to regular academic records, faculty verify students' improvement by their feedback to them and in the evaluative narrative statements they write for graduating students. Students also analyze their own experience of that improvement. As explained above, in a range of departments, faculty ask students to describe changes in performance over time by analyzing consecutive performances and the specific causes to which they attribute change. Therefore both faculty and students continuously witness the students' development of complex abilities in contexts all across the curriculum.

Student Performance

Students have consistently shown change on the College's own assessments designed by the faculty. Each graduate has, along the way, engaged in more than 100 active performance assessments in and outside of her various courses. Faculty design each assessment to elicit a particular level of one of eight required abilities, using the course's discipline content as a context. Each graduate's performances have been variously assessed by faculty, peers, and community professionals (and always by herself) according to criteria that remain stable across all disciplines.

We think it is important that so many students have shown consistent change through this complex network of performance measures. It suggests that the complex outcomes identified by the faculty are indeed developable and visible in performance to faculty, students, and professionals from outside the college; that a complex ability is recognizable across settings, despite the varied forms it may take in different disciplines and professional environments; and that such abilities can be developed sequentially to increasingly complex levels (Mentkowski & Doherty, 1984b).

Longitudinal Research

Office of Research and Evaluation studies found that students perform abilities as the result of instruction in the curriculum (Alverno College Assessment Committee/Office of Research and Evaluation, 1980; Friedman, Mentkowski, Deutsch, Shovar & Allen, 1982; Friedman, et al., 1980). For example, our study of the Communication generic instrument indicates that it validly discriminates instructed from uninstructed performance as does the Valuing generic instrument. Weekday students performed better after two years in the learning process in speaking, writing, listening, and reading criteria than weekend entering students who are usually older and more experienced. On level 4 of the Valuing process, weekday students performed better after two years of instruction than did weekend entering students (Friedman, et al., 1980). More important, patterns of student performance validate the sequential levels of the Communication ability. The cumulative sequence of levels 1, 2, 3 and 4 of Communication was confirmed for instructed students; weekend entering students used a different sequence. In a study of the developmental nature of the criteria for the Valuing ability, levels 2 and 3 were found to be similar in complexity for students.

For the Social Interaction generic instrument, we have had more difficulty demonstrating that instructed students perform at higher levels than uninstructed students. We did find that instructed students interpret social interaction skills differently from uninstructed students, and maturity and motivation affect performance in a group discussion (Friedman, et al., 1982).

Office of Research and Evaluation longitudinal studies of student perspectives found that students attribute learning outcomes to curricular elements and develop self-sustained learning or learning to learn (Mentkowski & Doherty, 1984b; Mentkowski, 1988). One of the most prominent curricular elements gleaned as causal

from the interview examples is *experiential validation:* applying abilities within and across courses, demonstrating them on assessments and during internships, and using abilities in multiple settings. One student said, "You can see you've really been learning in school because you can *use* it out there...it's not just something memorized...it's something you can actually work with...it's the experiences they give you and that have shown me that I've learned." And another said, "They've challenged me to use all my skills on the spot."

Among other elements identified, feedback and self-assessment are attributed as causal to developing outcomes by students.

The research staff also studied student performance on Human Potential Measures, a battery of 12 instruments drawn from outside the College. These studies demonstrate that growth and changes in students' human potential result from the College's curriculum (Mentkowski & Strait, 1983). Almost all colleges promise personal growth outcomes and expect that college will make a difference in broad abilities, lifelong learning, and life-span development. Studies of college outcomes have shown that college as a whole causes change (Astin, 1977; Feldman and Newcomb, 1969; Heath, 1977; Jacob, 1957; Pace, 1979; Pascarella & Terenzini, 1991). Our longitudinal research added a dimension that few, if any, studies have demonstrated—namely, change over time linked to student performance in a particular curriculum. The research questions were: (1) Do students change on instruments drawn from outside the college that measure human potential for learning, abilities, and life-span development? and (2) Can we attribute change on these measures to student performance in the curriculum?

Students clearly showed significant developmental changes on 12 measures across all three occasions (Mentkowski and Rogers, 1985; Mentkowski and Strait, 1983). Generally, the change that occurred can be related to student performance in the curriculum. This is the case even when we account for change due to the pretest scores, age, religion, parents' education and occupation, high school grade point average, prior college experience, marital status, year of entrance, residence at home or on campus, full- or part-time attendance, and type of major. (The time series design holds time constant and allows performance in the curriculum to vary, so we can attribute change to performance in the curriculum in the absence of a control group of students who did not attend Alverno.)

These results of all the external instruments together show that students appear to change more on these external measures during the first two years than during the second two years, but the changes in the second interval are more directly attributable to students' successful participation in the College's curriculum. This finding suggests that there may indeed be a college atmosphere effect, as studies of college outcomes have shown, but the curriculum does have a decided added value as well, particularly as students experience studies in their major or professional fields.

Recommendation for National Assessment System

Given the important role we have learned that multiple contexts play in developing and assessing complex abilities, we would recommend the following:

> Recommendation #4. A national assessment system should sample student performances in relation to instructional opportunities.

Implications, Issues, and Questions

- **How do we sample student performance in relation to instruction?**

Here we meet an earlier issue. In this section it takes on a new cast. We consider assessing in context. Which contexts? General education? The major fields? At graduation?

Elementary and secondary efforts will provide some advance information in regard to performance assessment at the local level; clearly, this question will be a focal point in field tests.

- **How is assessment linked to instruction?**

Faculty perspectives often include this important issue of carrying out effective assessment connected to instruction. Faculty are so accustomed to assessing in the context of instruction, that they believe that assessment cannot occur unless the judge understands the context in which the performance was created. To what degree will this approach meet the needs of a national system?

Sampling student performance in relation to instruction will be a key concept to investigate. Clearly, how this issue shakes out will determine to a large part the nature of faculty investment.

At the same time, it is important for faculty to take a firm role in rebuilding the public trust in higher education, and to expend the kind of effort necessary for assessment that is both linked to instruction and capable of meeting accountability demands.

- **Can students perform to standard?**

Once we set standards, educators will worry about whether students can reach them. Assessment system designers need to be prepared for some institutions not wanting to get involved because their students may not meet standards. Evidence that students can learn the complex abilities being assessed will not sway those who look at a national assessment system as just another high-stakes test. In fact, it has been our experience that students often look "worse" at the beginning because performance assessments measures not only knowledge recognition but also the internalizing of abilities like problem solving or critical thinking. But the temptation will be to fall back on recognition measures in order to give wary users some confidence in the system. One antidote will be to feed back students' actual performances with clear profiles of strengths.

- **How will institutions best describe the learning context for sampled student performances?**

Alverno Faculty make it a regular practice, at the beginning of a course, to describe to students the context for learning. While this practice is probably less frequent at the department level or institutional level, such descriptions are clearly essential. Finding ways to do this will be an essential part of creating a national system.

Principle Learned #5. Students can transfer abilities when they are assessed in contexts that are valid for what students learned and for how they will perform abilities later.

On-Campus Student Performances

Alverno Faculty experience each student's transfer of abilities in performances on multiple assessments. Cumulative academic records enable faculty to assert that students have made critical thinking, for example, a usable part of their personal repertoire. These records indicate that the students have shown their ability to think critically in situations initially including perhaps an analytic literary paper and an introductory management case study and eventually broadening to comprehensive assessments that may require a teaching demonstration for a peer group or a financial plan for an off-campus business person to assess externally.

Faculty give students credit for the performance of each progressive level of ability because the students show in their assessment performances that they can apply given abilities to a new context. We find that students can make that application when the context calls for the knowledge and level of ability they have been required to develop in their learning and other assessment experiences.

Off-Campus Student Performances

When students participate in off-campus internships, their performance is also evaluated by their mentors, whether from business, the arts, health sciences, education, or scientific research centers. In this aspect of the assessment process, the expert judgment of professionals from the public and private sector supplements that of the faculty. Their judgment assists us to confirm that students are able to transfer their developing abilities to the workplace (Hutchings & Wutzdorff, 1988).

Longitudinal Research

Office of Research and Evaluation longitudinal studies provide data to support student transfer of abilities. Results from indepth, confidential interviews during college show the student's experience of what is involved in such transfer. The studies of student perspectives cite evidence that students make relationships among abilities and their use (Mentkowski & Doherty, 1984a, 1984b; Mentkowski, 1988). For example, a student described *making relationships among abilities and their use* in the following terms: "Things are pulled together more for you through the abilities... a math class and a music class may have nothing to do with each other. But if you think about it, you are doing problem solving in both...it's really the same process. You don't experience that unless you can go to your abilities and see that it's interrelated, and you can pull it together more for yourself." Still another said, "You have to take these abilities like valuing in different classes...I looked at valuing from the philosophical and psychological standpoints in a death and dying course...it has caused me to see things from many different points of view...to try to get values out of a biochem experiment, looking for relationships in a lot of things, and looking for universality where there seems to be none, is really hard on your head..."

The longitudinal studies of alumna perspectives show that students continue after college to use abilities they have developed (Much & Mentkowski, 1982; Mentkowski, 1991d; Mentkowski et al., 1991; Giencke-Holl, Mentkowski, Much, Mertens & Rogers, 1985). In the analysis of the alumna perspectives interviews, two major categories of complex abilities emerged. Both younger and older women, across all professional groups, cited reasoning abilities—using such terms as "analysis," "problem solving," "decision making," "planning," and "organizational abilities"—as important to their career performance. Alumnae also consistently emphasized interpersonal abilities learned in college as critical to effective work.

Recommendation for National Assessment System

From everything we have learned about the importance of relating assessment to instruction and to future use if we expect the transfer of abilities, a clear recommendation follows:

> Recommendation #5. A national assessment system should define abilities and developmental performance criteria generically but assess them in contexts that are valid for what students learned and for how they will perform later.

Implications, Issues and Questions

- **How assess in context?**

Context includes the course, program, curriculum, indeed, the total academic experience. We have found the need to develop multiple measurements to adequately tap abilities across these multiple settings.

The key issue here is how broad or specific criteria need to be in order to cross settings appropriately. We can say quite directly that this issue raises different perspectives across the disciplines: what is appropriately broad to a behavioral scientist is too specific for a humanities faculty member. The discussions that result, however, are likely to generate criteria that can cross contexts. Our recent experience of building a codebook of abilities to measure alumna performance across a range of settings and professions outside college makes this goal seem within reach in college (Rogers & Talbott, 1990). But much will depend on how one deals with the next issue.

- **How will a national assessment system integrate and synthesize diverse institutional abilities and criteria?**

Who will contribute abilities and criteria? What kinds of institutions are likely able to make such a contribution, and are these representative of the "users" of national assessment system information?

The past practice of calling together experts in a field to identify items for the SAT or GRE is a worthy model: expert judgment in the identifying of abilities and criteria is an essential component.

For a national assessment system, however, a few experts will not do. Participation of practitioners at every level is necessary. New Jersey managed such an activity, and more and more sets of abilities are appearing as syntheses already made (e.g. U.S. Department of Labor, "What Work Requires of Schools," 1991). The Association of American Colleges (1991) has just completed an effort to define several majors (Fong, 1988). These examples are a start in this activity.

- **How define contextual validity?**

Questions of validity are sure to surface once the design of a national assessment system gets underway. Demonstrating the validity of performance assessed in context in ways that meet a national agenda will be on the minds of the supporters--but mostly the critics.

Our efforts to define contextual validity are reported in this paper. We are aware from this experience that contextual validity criteria do not enable us to generalize beyond one context to another, unless we can create broad criteria that cross contexts. How successful will we be at doing this?

Principle Learned #6. When an assessment system examines changes in student abilities/outcomes over time, including who changes and why, and relates those changes to the curriculum, the system yields information necessary for meaningful improvement.

As suggested above in Principle #4, we have been able to conclude that student performance of complex abilities changes over time in relation to performance in the curriculum. In order to produce information that could contribute to improvement, we found that a further level of analysis is necessary.

Longitudinal and Other Research Linking Change in Student Abilities to the Curriculum

As the results of longitudinal studies are broken open into intra- and inter-individual change patterns, a picture of who changes and why emerges (Mentkowski, 1990b; Rogers, 1991). The picture forms from six different sources that yielded data to link outcomes specifically to college instruction (Mentkowski & Doherty, 1984b; Mentkowski, 1988; Mentkowski, 1991d, Mentkowski et al., 1991). The sources analyzed included: (1) student performance on faculty-designed assessments that showed change as a result of instruction; (2) confidential interviews in which students and alumnae attributed changes in learning to curricular elements; (3) student performance on 12 external instruments that showed change linked to instruction; (4) alumna ratings and confidential interviews that showed graduates' use, in post-college settings, of abilities developed in college; (5) Behavioral Event Interviews of alumnae that showed them, in various settings, performing abilities developed in college; and (6) Job Competence Assessment (McClelland, 1976), (including Behavioral Event Interviews) of professionals who are not Alverno alumnae that showed the impact of education on their demonstrated abilities.

All of these sources validate the testimony of faculty who judge that students are learning, of external assessors who judge as successful the performance of some of these abilities, and of other students and alumnae who say they are developing these abilities and whose reports become more complex in describing their abilities in college, at work, and in their personal lives (Mentkowski & Doherty, 1984b).

In the case of student performance on faculty-designed instruments, studies specifically linked the abilities of Communication and Valuing to instruction (Friedman, et al., 1982; Friedman, et al., 1980). Student perspectives studies showed that the Communication and Social Interaction abilities learned in college are useful for functioning in personal and professional roles. On the other hand, there are other complex outcomes and abilities where the link to performance in the learning process is less clear. For example, changes on Rest's measure of moral judgment, the Defining Issues Test (1979), show significant, incremental gains during college and plateauing after college, with results decidedly linked to the curriculum. Changes on Watson and Glaser's (1984) Critical Thinking Appraisal show significant, incremental change across three data points during college, and during the five years after college, but these changes are not related to performance in the curriculum. Winter's (1976) Test of Thematic Analysis, a production measure of critical thinking, showed less overall change during college, but some change could be attributed to the curriculum. There were no changes on Loevinger's measure of ego development (Loevinger & Wessler, 1970) during college, while there was a significant change on the measure in the five years after college (Mentkowski et al., 1991; Mentkowski & Strait, 1983).

This pattern of results, showing where changes in abilities do and do not occur, becomes essential both for faculty investment in the system and for faculty ability to use the information to improve the curriculum. For example, what is faculty response to the finding that students develop critical thinking as measured by the Watson-Glaser Critical Thinking Appraisal? Do they say, "Great, we saw change during college and our graduates continued to improve after college." No, faculty questioned whether they could trust results from a multiple-choice measure. "Were the abilities really internalized; did they appear in performance at work?" Further, "If changes were not related to the curriculum should we continue to use the measure as an external criterion for performance assessments?"

In contrast, results from the Test of Thematic Analysis, a critical thinking measure eliciting constructed responses, showed less change, but faculty appeared to have more confidence in the results. They engaged in a discussion of how the curriculum was constructed to elicit role-taking, for example, and how this might have affected the results. "What did the findings have to say," they asked, "about our students in relation to the student groups on which the criteria for judgment were developed?"

Using the Information for Improvement

Helpful in the process is the ongoing analysis of longitudinal data and faculty judgment about usefulness of the information. It allows us to pinpoint those external measures that meet our expectations as external criterion measures of faculty-defined outcomes of college (Rogers, 1990).

Further examples of the range of data and its use can be found in the analyses of student performance on faculty-designed instruments. When such an analysis in relation to the developmental criteria of the Valuing ability surprised faculty, they quickly incorporated the results into their understanding of the ability: "We thought that each level of the valuing ability was sequentially related, from simple to complex. Actually, now we understand that performance criteria at levels 2 and 3 are different abilities but similar in difficulty. Since then, faculty have expanded the meaning of the ability and extended the criteria. Another analysis, examining pre-post instruction results from the half-day performance assessment all students complete at the end of the general education sequence, showed clear directions for improving the instrument (Alverno College Assessment Committee/Office of Research and Evaluation, 1982; Rogers 1988). A revised instrument is now in place.

A final example shows how facilitating structures can assure the use of information for improvement. When analysis of longitudinal interviews showed the importance of self-assessment as an element of the assessment process that was critical to self-sustained learning, the data became part of a regular report to the Assessment Committee. This faculty committee of performance assessment specialists brought this information to their review of sample instruments that they had collected across the entire faculty. They examined the instruments for how each one elicited self-assessment from students. Some of the instruments did so in a cursory manner: students were asked to merely rate the strength of an ability. The committee then provided feedback to faculty designers on how to elicit increasingly complex self-assessment from students and sponsored day-long workshops to improve this component in instruments across the college.

Clearly, some kinds of assessment information was of value for immediate revision of abilities, performance criteria, and instruments. Other kinds of information, about the development of critical thinking, for example, deepened faculty understanding of patterns of student development during college and afterward. It also seemed to strengthen their resolve and commitment to performance assessment. The overall effect encouraged the research staff in their decision to rely more heavily on alumna performance data from Behavioral Event Interviews in the measurement of critical thinking when they made their reports to faculty.

Recommendation for National Assessment System

As a result of what we have learned about the information for improvement that an assessment system can yield when it examines changes in student abilities/outcomes over time and relates them to the curriculum, we would recommend the following:

> Recommendation #6. A national assessment system should link changes in student abilities/outcomes over time, including who changes and why, to student performance in college curricula and feedback the information to institutions.

Implications, Issues, and Questions

Clearly, institutions will need to be able to marshall evidence for the value, impact, validity, and effectiveness of curricula by describing what they do and what evidence they have for student achievement. Describing what they do, that is, describing the learning context, would be a first step for an institution to participate in a national assessment system. Comparing the learning context with sample student performances as evidence of learning would then help determine the important link between how students learned and what they learned.

This process would enable institutions to more effectively participate in collaborative efforts across the country to jointly examine and review student achievement. Such efforts provide institutions with opportunities for critique and comparison. There are already existing opportunities for this kind of activity, e.g., the AAHE Assessment Forum, which holds annual meetings where institutions can share results and invite criticism.

Questions remain. It is important to address at least a few:

- **How do we link information from entering student abilities/outcomes and graduating student abilities/outcomes? How do we relate changes in student abilities/outcomes to curriculum?**

We have been citing extensive evidence using this approach. Alexander Astin's (1991) assessment methods also flow from a developmental model. He defines outcomes as "those aspects of the student's development that the institution either does influence or attempts to influence through its educational programs and practices" (p. 38). Astin recommends a longitudinal research method that studies causal connections between inputs (students' entering abilities), environment, and outcomes. "Assessment results are of most value when they shed light on the causal connections between educational practice and educational outcomes" (p. xii).

Astin gives specific technical advice for analyzing assessment data and building the kind of quantitative, longitudinal data base an assessment professional will need to realize the model's benefits. He provides clear steps for consequent statistical analysis that most anyone can follow.

Consistent with the improvement agenda for assessment, Astin argues for a heavy emphasis on using assessment results. He contrasts incentive and feedback models for their value in improving student and program performance, and lays out the advantages of direct feedback. Based on cooperative rather than competitive alternatives, he draws public policy implications for state assessment activities.

Astin's model offers a valid approach to designing assessment systems. Beneath its undeniable advances in thinking, it also raises several questions for the assessment practitioner. The model highlights the importance of student growth as an outcome. Unquestionably, the input-environment-outcome model is a considerable advance on higher education's preoccupation with resource and reputation indicators such as number of books in the library and faculty scholarship records. Our own experience, however, shows that the pre-test/post-test design Astin recommends for analyzing change and linking it to educational programs can fall short of the ongoing, multiple collections of longitudinal data needed for creating intra- and inter-individual change patterns that model the interactive dynamic of student growth and curriculum effects (Mentkowski, 1990b) that are important for intra-institutional studies. Astin's model, tested primarily in large scale, quantitative, cross-institution studies, includes the essential component of "environment," that is, educational practices that must be linked to changes in student outcomes. Our own experience shows that linking student changes over time to the curriculum can be accomplished with large-scale qualitative data bases as well (Deemer, in press; Deemer & Mentkowski, in press; Much & Mentkowski, 1984).

Clearly, we have a good deal of effort ahead of us if we are to design and develop methods for analyzing change in student outcomes over time, and linking that change back to the curriculum (Astin, 1991; Collins & Horn (in press); Mentkowski, 1990b; Rogers, 1991; Willett, 1988; 1989; 1990). This is an area for a good deal of research, but it is probably one of the most promising approaches. This approach deals directly with many of the problems institutions raise when they contemplate a national assessment system. How will a national system attribute change in student outcomes? How will they measure change?

- **What are best methods for analyzing change?**

If a national assessment system relies on changes in student outcomes, and not just on exit criteria, then institutions will be encouraged to look at change as well, and together could work to make meaning out of change data. For example, Alverno and Millsaps College have each collected change data and have collaborated to find best methods for analyzing change (Mentkowski, 1991b). Issues of inter-institutional comparison, which can be disheartening when one is comparing institutions on exit criteria alone, disappear when institutions are discussing how to measure change in relation to curriculum. Institutions work collaboratively when they are discovering who changes and why, and what the patterns of change are. Each institution is able to identify students who are not learning and those who are. Therefore, institutions can unite in a common question: How do we improve learning for each student?

- **How do we aggregate information from institutional assessment systems?**

Further, the question of how to aggregate information from institutional assessment studies becomes more open to discussion when one is describing results from change studies rather than comparing scores. Questions such as "What level did your students reach as a group?" drops away. Rather, "What proportion of your students showed change on the complex abilities we are trying to understand, and can you at this point in time relate any change you see to your curriculum? What do the patterns of change tell you about the complex abilities we are all trying to measure?" These questions yield exciting discussions among faculty who are then focused on improving curricula.

Principle Learned #7. We can validate an ability-based performance assessment process and institute an instrument validation process that gradually improves instrument validity. We can establish the educational value, impact, validity, and effectiveness of the abilities/outcomes.

Validating the Process: Longitudinal Research

One question that soon followed upon the inauguration of Alverno's ability-based academic program in 1973 was that of demonstrating validity. At the time, traditional validation concepts and strategies were not congruent with the underlying assumptions and principles of our assessment system. Thus, it became necessary to re-think the meaning of validity (Mentkowski, 1989; Rogers, 1988). That rethinking entailed an examination of the process by which faculty design and continually refine and revise the abilities, levels, performance criteria, assessment instruments, and learning strategies. It also meant articulating a framework for validity that would preserve the integrity of our system.

Defining validity continues to be a challenging and ongoing exercise. For Alverno, validating the assessment process now includes: (1) the processes by which faculty define abilities and criteria and design and revise assessment instruments; (2) the work of faculty and staff, through the collaboration of the Assessment Council and the Office of Research and Evaluation, to build a community of judgment about student performance; (3) articulating modes of inquiry about our criteria, evidence, judgment, and assessment processes; and (4) studies by the Office of Research and Evaluation that generate evidence and comparisons to norms and criteria from internal and external sources.

Perhaps some of the most valuable information for establishing the validity of the assessment process came from longitudinal studies of student and alumna perspectives, from alumna studies of performance, and from studies of outstanding professionals who are not Alverno graduates. Data showed that the assessment process, in particular, was essential both in the mental constructions and the performances of participants. Specifically, feedback and self-assessment were cited as critical for students' taking responsibility for learning and for using different ways of learning. Self-directed, or self-sustained learning, the ability to learn within a range of situations and settings to become a better learner over time and to adapt and integrate one's abilities, emerged as an essential element in transferring abilities after college (Mentkowski, 1988, 1991d).

Thus, our in-depth studies confirmed key elements of the assessment process and clearly singled out the assessment process as critical to student learning during college. *Self-assessment* emerged again in abilities demonstrated in Behavioral Event Interviews of alumnae performance, and was the most frequently coded ability in a pilot study (Mentkowski, et al., 1991; Rogers & Talbott, in press). This ability, *accurate self-assessment*, was a key ability learned early in the career of the outstanding managers and executives we studied who were not our graduates (Mentkowski et al., 1982). All of this information has heightened our resolve to improve performance assessment; it has reinforced the importance of the dynamic process by which systematic feedback comes to the faculty.

Validating Instruments

It is up to the Office of Research and Evaluation to articulate the meaning and underlying principles of validity and to conduct studies that establish validity. However, the actual process of improving assessment and student learning, which is at the heart of the validity process, operates within the faculty working individually and through their academic departments and the Assessment Council as well as with the Office of Research and Evaluation (Alverno College Office of Research & Evaluation/ Assessment Committee, 1989; Loacker, Loveland, McElroy & Mentkowski, 1991; Mentkowski & Rogers, 1988; Rogers, 1988).

Goals of the overall process suggest the dual function they address:

- improving an instrument's design so that it assesses what it aims to assess and is representative of a valid assessment process and theory (This also enhances our understanding of what is a valid assessment process and theory);

- improving instrument criteria so they can adequately represent the ability to be assessed (This also enhances our understanding of those abilities.);

- improving expert assessor judgment of student performance in relation to criteria and improving feedback for learning (This also enhances our understanding of how expert judgment works);

- improving student learning as the result of the assessment process (This also enhances our understanding of how students learn from the assessment process and their own self-assessment).

These goals are realized through a series of strategies faculty apply to instruments. Which strategy they use depends on where an instrument is in its development, whether the instrument is used in or outside of class, and whether it is used as a milestone measure to judge outcomes across the college.

Strategies for design-based validity include evaluating instrument components in relation to guidelines for instrument design formulated by the Assessment Committee (Alverno College Faculty, 1985a). Performance-based validity strategies include criteria evaluation based on student performance, inter-judge agreement between assessors and reviewers, evaluations of assessor training and assessor use of criteria, judgment,

feedback, and finally, establishing that student involvement in the assessment process leads to learning. Comparison of the instruments with assessment principles and with educational assumptions about teaching, learning, and assessment completes the process (Alverno College Office of Research and Evaluation/ Assessment Committee, 1989).

Therefore, Alverno's definition of contextual validity (Mentkowski & Rogers, 1988; Rogers, 1988) refers both to the multiplicity of perspectives reflected in the instrument's design and use, and to the match of the performance modes that represent current/future performance situations, assuring that students can transfer their performance to a range of settings during and after college. Contextual validity means that an instrument's design and use:

- integrates the educational assumptions, expectations, and purposes for an ability or outcome of a particular institution,
- is consistent with an institution's curricular principles and practices,
- integrates a multiplicity of individual faculty and departmental perspectives in ability definition, instrument design, and judgment of performance so that students will transfer abilities to a range of settings,
- is designed to elicit sustained, open, interactive performance that enhances transfer of abilities to other settings,
- builds in faculty expectations for the multiple demands of work, family, and personal life after college,
- calls for performance in a mode that has a fidelity and depth that matches situations the instrument needs to represent so students can better generalize abilities to situations outside class and after college,
- anticipates that instruments will be designed, evaluated, revised, and validated by departmental or cross-college, interdisciplinary faculty based on their assessment principles and their analysis of student performance, as well as their disciplinary and pedagogical expertise.

Recommendation for National Assessment System

Our experience with designing and implementing a comprehensive validation system, including longitudinal evidence, prompts the following recommendation:

> Recommendation #7. In order to examine the educational value, impact, validity, and effectiveness of a national assessment system, designers should build in a research and evaluation component.

Implications, Issues and Questions

Our experience has taught us that establishing the educational value of the assessment process was necessary to assure commitment to continually improving it. Establishing the validity of instruments was not nearly so critical as demonstrating that a system with certain essential elements was worth working toward because over time, benefits would accrue for students. We have found that if faculty disagree with the basic educational assumptions underlying an assessment system, "tinkering" with particular aspects of it, or with certain of its instruments, will not solve problems of continuing investment.

From a national perspective, one can expect that institutions will be concerned about the educational value of an assessment system. Persons across the country will raise questions about the system itself, but also about why they should invest in it. The implications for validating an assessment process are that the elements of the design need to be educationally sound in the eyes of the persons using the system. This is why it is critical early on to include all the elements of the design at least in the "grand plan" rather than working at it piecemeal, often citing feasibility criteria to justify the piecemeal approach (Mentkowski, 1991c). While this seems difficult to do, the face validity of the design rests on the users' faith in the system's ability ultimately to meet the promises it makes. Our experience is that having "half a system" designed can mean failure for good ideas that are not realized because persons will not "buy in" to a long-term commitment.

- **How do we design and validate an assessment process?**

It has been our experience that validating the assessment process has been more important than validating the instruments that contribute to the process. If one places the "educational value" criterion first, then users will "live with" almost any snafu in the design's elements. They will also trust a research and evaluation system to provide information for improving the system along the way. An important implication, then, is that all design elements will need to be planned from the start and need to be clearly tied to the educational assumptions that underlie the system.

- **How define construct validity?**

Another continual concern is with construct validity. Our experiences have taught us to rethink, re-examine, and extend our conceptions of validity as we worked to measure complex abilities (Mentkowski, 1989). For example, how should we define construct validity? We are just beginning to define such complex abilities as critical thinking. Our definitions shift over time as we learn from the experience of trying to measure them. This adds a whole new dimension to measurement issues. What is construct validity when the abilities are not fully defined? When definitions of the abilities emerge in part during the assessment process, while the assessor is assessing (e.g. "I haven't seen that response before. It is unique"), and when abilities are not unitary, but multidimensional, what are the implications for validation strategies? Clearly, measuring higher order abilities and determining the implications of construct validity issues will be a task for the next decade.

- **What is good evidence?**

Further, what is good evidence? When the unit of analysis is expanding from student selection of predetermined test item alternatives—or even short answers—to include proactive, open, interactive, dynamic, sustained student performance, how does one determine what kind of evidence is critical and necessary? Working out answers to such questions is an ongoing process. We have learned not to wait for final answers but to keep developing answers by trying out new methods. We would recommend the same operational principle for a national assessment system.

- **How validate expert judgment?**

And how does one validate expert judgment? While our studies indicate that establishing inter-judge agreement is clearly an important strategy, there are occasions when multiple assessors are engaged in judgment, not to come to consensus, but to bring a range of perspectives to bear on the performance. For example, a faculty member, a hospital administrator, and an ethicist may all be judging a student nurse's ethical decision-making in a situation where costs and individual needs are in conflict. Here, inter-judge agreement may not be at a premium. Rather, effectiveness for feedback to the student may be the important validity criterion to meet.

The challenges and rewards of pursuing expert judgment as an element of a national assessment system may be previewed in the work of a FIPSE-sponsored Critical Thinking Network (Cromwell, 1986). Here, a consortium of 36 faculty of four disciplinary areas from almost as many institutions from around the country met at Alverno over three summers to consider how to define critical thinking, how students learn critical thinking, and how to assess for critical thinking.

In the process, the four groups—psychology, humanities, natural science, and management—found that expert judgment was a reasonable starting point for designing assessment (Cromwell, 1986; Halonen 1986). In their report, the arts and humanities group describe how they recognized that assessing critical thinking is a natural outgrowth of a process in which liberal arts faculty have been engaging for years. They describe how they analyzed their own judgment throughout the defining of critical thinking, designing of assessment, and analyzing of student performances, and in the process, refined their ability to do each of these. Consistently, they kept several validity issues at the forefront (Mentkowski & Rogers, 1986): What do I mean by expert judgment? Why do I make the judgments I do? How explicit should my rationale for judgment be? How can I make expectations for students explicit?

In this case, establishing construct validity, where the construct was critical thinking, became an interactive process of generating criteria that described elements of critical thinking, using these criteria to judge student performance samples, and gradually refining definitions of critical thinking and the criteria used to assess them (Mentkowski, 1989). Clearly, establishing the validity of expert judgment in the assessment of complex abilities will need a great deal of attention in the development of a national assessment system that includes ability-based performance assessment.

> Principle Learned #8. A dynamic assessment system incorporating input from and feedback to faculty, as well as administrators, provides for the effective use of information to keep abilities, performance criteria, and standards responsive to and in advance of the needs of our society.

An ability-based performance assessment system is dynamic. Perhaps one of the most cogent findings from a review of our 18 years of practice working with an ability-based performance assessment system is that the definitions of the abilities and related disciplinary and professional outcomes change as we try to measure them. The instruments also change rapidly as we improve them after analysis of student performance. We refine the performance criteria as we become more adept at sorting out what aspects of an ability are visible in performance and what aspects of an ability form the basis for expert judgment. As we assess, the process itself is a source of continual information that leads to refinement and therefore, change.

Continuous Improvement in Practice

At the classroom, department, or institutional level, one can step back and observe this continuous improvement atmosphere and see the results of this changing panorama. There are examples throughout this paper. Others, in the last year, include four faculty groups responsible for respective abilities that presented official revisions of the definitions of ability levels: Valuing in Decision-making, Global Perspectives, Effective Citizenship, and Aesthetic Response. Disciplinary and professional departments also published a revised set of advanced outcomes in the major and support areas (Alverno College Faculty, 1990). All of these changes were based on study of contemporary theory and projected responses to societal needs as well as analysis of student performance and evaluation of instruments and criteria.

Research Studies

Over the last 15 years, the Office of Research and Evaluation has expanded their methods as the questions became focused on more specific issues. Once studies of outstanding professionals who are not our graduates were in hand, the staff began a more in-depth look at how advanced outcomes develop in the major. Within the longitudinal studies, they analyzed pathways that lead to abilities demonstrated by effective alumnae. Measurement approaches that look at the development of broad outcomes over time (moral, intellectual, ego development) have become more focused on observing, for example, how the developmental level of an education major interacts with her performance in student teaching.

All of these studies were begun with the assurance that, through the ongoing dynamic structures of the system, the results will be used to improve assessment and learning opportunities in the major.

Recommendation for National Assessment System

Everything we have learned about the necessity of an assessment system being dynamic if learning is to improve makes the following recommendation crucial:

> Recommendation #8. A national assessment system should be a dynamic system based on faculty-defined abilities, as well as other sources, to make the outcomes, criteria, and standards responsive to and in advance of the needs of our society.

Implications, Issues and Questions

Commitment to a dynamic system has consequences for measurement. Rather than building a measurement system that is built on consistency, we need to build one that is based on change as the rule. Here, the assumption is that assessment contexts will vary and they are expected to vary. The expectation is that purposes, definitions, curricula, and faculty-designed instruments undergo revision over time. It will require important decisions about what to keep stable.

- **How do we create a dynamic system?**

One of the more difficult issues to face in creating a national assessment system is the pinpointing of those elements of the system that will contribute most to its dynamic qualities. As we have recommended, the identification of broad, durable abilities should be a point of stability in such a system, while the developmental performance criteria should change with insights from student performance of the abilities. One of the issues that comes immediately to mind is the question of performance levels.

- **How do we set performance levels so they reflect changes in what is being taught and what needs to be learned?**

This is a question at both the local and national level. It is further complicated by the aim to stay in advance of what society needs. How do we keep responsive to what is needed now by graduates who are preparing not only for future positions but also entry level positions at work? As NCES has pointed out, "Low standards may reduce the value of the program, while high standards can be troublesome and perhaps unrealistic for both students and institutions."

Thus, it becomes critical to define criteria developmentally, at various levels of proficiency, so that progress rather than end points alone can be measured. This enables feedback to be developed in terms of strengths and areas to be developed Such developmental feedback, with clear indications of what is beginning performance and what is more advanced performance, is motivating. Students can see where to go to improve. This helps to deal with the problem that not all students enter any learning or work environment with the same sets of abilities, nor do they graduate that way.

We have experienced the raising of standards at our own institution. At the national level, we expect that the standards that define effective performance will be expanded, and that as institutions become more effective at instruction, students will more likely meet the standards. As assessment information gets used, better teaching and student learning results.

Further, many educators realize that while it makes educational sense for them to show leadership in defining abilities and performance criteria, society is not satisfied with current performance levels. Other groups responsible for education—state and federal policy makers, corporate groups—are also expecting to contribute to standard setting (Albert, 1991). At the heart of designing and implementing a national assessment system there remains the complex challenge of finding a way to bring about collaborative synthesis.

The key is to develop a process that ensures that the "reliability" and "stability" expected of assessment instruments serve at the behest of a larger, dynamic assessment system, so that performance criteria and standards can change. Such a process would assure that the specifications of what is measured and to what level are consistently open to question; performance criteria and measurement serve to generate information that will cause regular changes in the criteria and instruments.

- **How define validity in a changing context? How define reliability when change rather than consistency is measured?**

Can performance assessment measures maintain both validity and reliability over time? Traditionally, we have had what seemed to be a steady ruler against which to measure progress. How will measures be designed now? Change will occur in the very techniques formerly counted on to "hold still" across time. The most important issue is whether the assessment system itself is valid rather than whether the instruments themselves are technically "sound" in terms of reliability defined as measuring consistently over time. Once one focuses on the validity of the assessment system as a whole, the issues of reliability of measurement are judged within this framework.

None of this removes the difficult issues confronted if one considers how to identify points of stability when it is impossible to hold a performance criterion "still." We do not claim to have solved this problem. The challenge is to adjust or develop a psychometric approach that is based on change rather than consistency (Messick, 1980; Mentkowski, 1989).

Clearly, one approach we have used institutionally and now recommend is to maintain broad ability definitions that may serve as more stable place holders over time, and to spend effort in training expert judges to use these broad definitions to ground their judgment. Explicit performance criteria elicit evidence from judges and enable more explicit links from a particular performance to the judgment, so the basis for judgment is more open to external critique. A second strategy is to clarify elements of the ability one can more likely observe directly, and those that may form the basis for judgment but are less likely to be directly observed within each situation. For example, in judging critical thinking, one may be able to observe "making relationships" more directly, but one may not as easily make explicit and judge the confidence it takes to "state relationships in the midst of a value-laden argument." Over time, we may develop a clearer understanding of any criterion like the latter, especially since it seems to discriminate effective from ineffective performance.

The issue remains that if performance is measured, it must be measured in context. The context will be variable and will change. Traditionally, we have put careful effort into standardizing the context and conditions of an instrument. For performance assessment, it would be important to shift that effort into identifying abilities that most institutions could buy into, and defining performance criteria that are developmental so that individuals and institutions could see profiles of strengths and areas to be developed. Thus, the performance criteria could be applied by expert judges to quite different performances elicited from quite different settings.

Because the ability has to be measured in the context of a discipline or professional area in order to assess the full range of the ability measured, clearly, understanding the context of the performance is critical to applying performance criteria. Careful specification of context is necessary because of its effect on the criteria.

Because an important source of stability is the generic outcomes themselves, it is important to ask about the reasonableness of assuming that they can be generated since even within the institutional level, contexts are so variable. The meaning of "history," "philosophy" and "management" determines the meaning of "critical thinking in the discipline." Such definitions are themselves in flux: multiply that across a range of institutions.

Then can disciplines and professions become more clear about their outcomes? Some existing efforts suggest that they can. The American Association of Colleges conducted a project where the learned societies were actively involved in defining the major (American Association of Colleges, 1991). While there are clear difficulties in such efforts, and there are inconsistencies in both the approach and the results across the disciplines, as would be expected, some progress indicates that this is a worthwhile effort. Institutions are also becoming more clear about their outcomes (Clayton State College; James Madison University; Kean College of New Jersey; King's College; University of Tennessee-Knoxville). As the assessment movement continues, a range of institutions are involved in clarifying outcomes, and in defining what they mean at the department level.

In addition, these and other institutions are actively involved not only in assessment of these major outcomes but in activities involved in evaluation of the major (Woodward, 1984). Their activities suggest that there are public reports and other sources that a national assessment system could draw on to involve faculty in defining abilities and performance criteria that would, because of these prior efforts, have some acceptability across institutions, the learned societies, and the professions.

Clearly, creating a dynamic system means dealing with a host of problems that we have not yet solved. But it is clear that a key element in creating a successful assessment system is acceptance of the underlying assumption that change rather than consistency would permeate expectations. Dynamic quality would be an expected requirement of the system rather than a stumbling block that confounds. Such a system, and any performance criteria used to judge performance or to set standards, would change as instruction improved and students became more expert at the abilities. That these changes could be incorporated would provide proof to both faculty and other groups that a national assessment system would not deal with either minimum or idealistic indicators, but, rather, a picture of abilities that continues to serve both as beacon and support to student learning.

Principle Learned #9. Creating a context for assessment is as important as creating the assessment method.

We have learned that an ability-based performance assessment system both demands and contributes to an atmosphere supported by structures that ensure strategies for continuous improvement. Initially, we saw the importance of creating a context for the assessment of our students to enable them to demonstrate levels of

ability leading to graduation. We saw that such assessment would gradually assist them to improve their learning and assist us to improve our teaching if we developed it into a system with a strong supportive context (Alverno College Faculty, 1985a; Mentkowski, 1991c). Critical to such a goal was developing a community of learning (Read & Sharkey, 1985) where a gradual commitment to improvement, to questioning our basic assumptions, and to building institutional structures all infused our new ways of thinking. Like the assessment of our students itself, this involved a systematic design effort (Read, 1980). We had to discover—often through failures—the kinds of shifts in attitudes and perceptions that were needed to accompany our move toward an assessment-for-improvement system that was criterion-referenced but simultaneously set standards for graduation. Most of us were grounded in testing knowledge, rather than assessing abilities that linked knowledge to action.

We would not have been able to develop a comprehensive assessment system if we had not concurrently built processes and structures that we could institutionalize, to make sure that the design and development of assessment continued with faculty and student investment. When problems arose, we held faculty and student forums and departmental meetings; we created task forces. Sometimes these were ad hoc; at other times, they became permanent institutional structures. Throughout, the purposes of assessment were discussed, made public, documented and critiqued both inside and outside the institution. We continue to consistently deal with some aspect of assessment at our triennial week-long faculty institutes in order to maintain and develop our purposes, motivation, and educational assumptions and principles.

Three years after we began, we gradually extended the system to include institutional assessment. The same approaches applied. The new Office of Research and Evaluation had to develop strategies for the involvement of students and alumnae in completing instruments from outside the institution, with no credit involved. Strategies for involving and investing faculty in an institutional assessment process were essential.

Recommendation for National Assessment System

From what we have learned about the necessity of a carefully designed and developed educational context to support assessment for improved learning, we would recommend the following:

> Recommendation #9. Creating a context for a national assessment system that yields educational improvement should be planned for implemented as an essential part of the process.

Implications, Issues, and Questions

- **How best are students, faculty, institutions, states, federal agencies, and the public invested in a national assessment system?**

Given our experience, some elements emerge as contributing to the investment of institutions, specifically their students and faculty, in assessment activities for the purposes of generating cross-institution feedback. (Mentkowski, 1988, October):

For students and alumnae:

- Treatment as partners in improving college learning,
- Understanding of the rationale for participation,

- Knowledge that professionals in their discipline are also involved in identifying abilities, performance criteria and standards,
- Individual feedback that contributes to their developing picture of their own abilities and growth, and
- Sense that they contribute to changes in curriculum that will benefit other students.

For faculty:

- Focus on cross-disciplinary questions that inform their understanding of student learning and development,
- Communication of patterns and complexities that suggests reasons and direction for improvement,
- Aggregate information that carries the individual student's voice, that connects individual to group findings, and
- Regular feedback from multiple perspectives, data sources, measures, and criteria.

Each of these elements needs ongoing nurturing and development but the context they continue to create proves worth it.

At this point in time, some 18 years since the initial implementation of the program, the responses of participants in our week-long, on-campus workshops in teaching and assessment encourage us. They report that they discern a context of attitudinal, motivational, institutional, and external support for an assessment system. They tell us that the system we describe in our publications is indeed operational (Alverno College Educators, 1986; Alverno College Faculty, 1985a, 1985b; Earley, Mentkowski & Shafer, 1980; Loacker, Cromwell, Fey & Rutherford, 1984; Loacker, et al., 1986; Mentkowski & Doherty, 1984b; Mentkowski & Loacker, 1985).

Some essential distinctive qualities of institutional assessment have emerged not only from our own practice but that of others (Mentkowski, 1991c):

- Assessment should be a means, not an end;
- Assessment should be a means not only to establish accountability but also to achieve educational benefits;
- Assessment purposes, goals, and methods should emerge from the setting;
- Assessment should incorporate multiplicity;
- Assessment should be structured to encourage coherence; and
- Feedback should be an essential part of assessment.

A review of current practice then suggests six guidelines for constructing an assessment context.

- Make a long-term commitment to a dynamic plan based on an articulated educational mission;

- Create interactive processes and structures that flow from articulated educational principles;
- Define outcomes, criteria, and comparisons publicly;
- Rely on faculty questions and input for direction and definition of outcomes and criteria;
- Translate results into relevant, "live," feedback-usable information about performance that stimulates improvement; and
- Create opportunities for user involvement throughout the process, including external critique/review.

- **How create a community of judgment?**

A review of these guidelines argues that a national assessment system should make clear a dynamic plan that outlines not only a long-term commitment, but the involvement of various constituencies in setting direction for the system itself. Key to the plan will be creating strategies that enable participants to interact around the general goals as well as the specifics of the system and any attendant measurement. Such active involvement has been carried out on a large scale in Washington State (Council of Presidents and State Board for Community College Education 1989) and New Jersey (College Outcomes Evaluation Program 1987). Vermont, Connecticut, and California are also pursuing state-wide efforts at performance assessment (DeWitt, 1991).

Including performance assessment as part of a national assessment system will clearly mean creating processes for specifying criteria, training assessors, making expert judgments, interpreting results, and discussing applications. The key element here is the involvement of the users of the information in all elements of design, implementation, and use of the results, in order to create a community of judgment.

How can performance assessment with the supporting context described here be affordable? Cost effectiveness requires reorganization both of time and money (Read, 1980). Generally, it means spending less effort in using experts to create "items" and more on use of professionals who are already working to improve programs, and involving them in identifying and judging performance samples.

Clearly, there are major difficulties in meeting the requirements for user-involvement in design, implementation, and use of the results. But as a nation committed to world-class quality, why not set it forth as something to work toward? It is tempting to imagine a national assessment system that works to create a community of judgment:

- Institutions and their faculties would involve themselves because they share common interests in assessing abilities identified in the national goals;
- Institutions would train their own expert judges both to sample their students' performance in relation to instruction and to judge it in relation to performance criteria that have been defined and validated nationally;
- A national center would train and validate expert judges;
- Institutions would exchange judges to validate judgments, improve criteria, and garner external insights (Fong, 1988);

- Institutional judges would meet nationally to discuss results and their implications; they would lead such discussions on their own campuses; and

- Institutional judges would discuss the kinds of comparisons that might provide more insight into standard setting and making standards public.

Several purposes seem to characterize persons committed to assessment (Mentkowski, 1991). They want assessment to make a difference. They share some common educational values that center on expanding human knowledge for human use and educating diverse students for a changing and challenging global environment. They realize these commitments by building meaning into the broader processes of program and institutional assessment. A national assessment system then could count on such persons, who have demonstrated that results would matter to those educators, administrators, and policy makers in the public and private sector who are in a position to use the information to improve teaching and learning.

Principle Learned #10. The effectiveness of an assessment system concerned with the improvement of learning depends partially on a coherence that comes from the following articulated components:
- educational values, assumptions and principles that are tied to the mission statement of the institution;
- an assessment theory (what are the components of good assessment?) consistent with those values and assumptions; and
- a psychometric theory (how do we best measure and credential performance and give feedback to students on their abilities?) consistent with those values and assumption.

Over the years, Alverno's faculty and staff have worked to articulate the educational values, assumptions, principles, and practices that underlie its ability-based curriculum with its performance assessment process[6].

Because these elements are embedded throughout this paper and are thoroughly developed in our publications, it does not seem necessary to illuminate them here, but rather to articulate them in relation to a national assessment system and to discuss implications for creating such a system.

Clearly, Alverno educators have been generating theory along with their practice. Articulation of our values, assumptions, and principles has been essential to our continuing development. Further, other institutions have been able to learn from us because we have shared not only descriptions of our practice, but our developing conceptualizations.

---

[6] Alverno educational values, frameworks, principles and practices include a liberal arts/professional focus, a student-centered, outcome-centered emphasis, and a coherent, developmental curriculum (Read & Sharkey, 1985) with these elements:

(a) ability-based, via the disciplines (Alverno College Faculty, 1985b; Alverno College Nursing Faculty, 1979; Earley, Mentkowski & Shafer, 1980; Loacker & Palola, 1981; Loacker, et al., 1984; Read, 1980);

(b) experiential learning (Doherty, Mentkowski & Conrad, 1978; Hutchings & Wutzdorff, 1988);

(c) assessment-as-learning for individual student development, credentialing, and program evaluation (Alverno College Faculty, 1985a; Loacker, 1988; Loacker, et al., 1986; Mentkowski & Loacker, 1985); and

(d) educational research, program evaluation, and institutional assessment for demonstrating the value, impact, validity, and effectiveness of ability-based performance assessment, the curriculum, and the broad outcomes of college, via college outcomes studies (Mentkowski, 1988; Mentkowski & Doherty, 1983; 1984; Mentkowski & Loacker, 1985; Mentkowski, 1991c; Mentkowski et al., 1991).

Recommendation for National Assessment System

Our experience with the profound effect that articulating a conceptual framework (including underlying principles, assessment theory, and psychometric theory) can have on learning, teaching, and assessment prompts the following recommendation:

> Recommendation #10. A national assessment system should have at its root a coherent set of articulated components and principles:
> - educational values, assumptions, and principles underlying the national goals;
> - an assessment theory that describes the components of "good" assessment; and
> - a psychometric theory that describes how we best measure and credential performance, and give feedback to students, faculty, institutions, states, federal agencies, and the public on student achievement.

Implications Issues, and Questions

- **Can institutions articulate and identify shared educational assumptions and principles?**

This most important underlying question remains. While the practices of any one institution are not generalizable to other contexts, the underlying principles are likely to be informative, useful, and potentially shared. For example, three institutions agreed to principles of ability-based performance assessment (Loacker, Wutzdorff, Barnett, Brown, Farmer, and O'Brien, 1988). Another collaborative effort with the Faculty Consortium for Assessment Design coordinated by Alverno faculty (Alverno College/FIPSE Assessment Project, 1987) tested ability-based performance assessment design principles across 24 institutions involving 54 faculty from 1987 to 1990. The W. K. Kellogg Consortium for the Improvement of Teaching and Assessment (1989-1992) coordinated by Alverno College involves 11 institutions across several levels of education: high school, community college, college, university and schools of pharmacy and medicine. The consortium is currently synthesizing educational assumptions that are common across their institutions. They are elaborating these with (a) questions that are prompting constructive change at their institutions, (b) strategies that are working to implement ability-based education, (c) barriers and constraints they are experiencing, and (d) indicators of change toward ability-based, outcome-centered education. At present, this Kellogg consortium is discussing shared outcomes to see if they can be described and sequenced across levels of education.

These inter-institutional consortia experiences have taught us that it is not only possible but likely that some institutions can come together to examine their assumptions, principles, and outcomes. The search for commonalities and differences illuminates a more general theory, and discussion of the range of institutional practices enables a group to key in on those assumptions about learning and assessment that can be articulated. Educational values, assumptions and principles underlying the national goals should be clarified across time. If components and principles are based on the most advanced thinking the nation has to offer, and the "thinking" is consistently re-examined, this will be a powerful incentive to institutions to join such a national effort.

The most advanced educational values and principles will no doubt raise questions for assessment theory and the components of "good" assessment. And if we incorporate ability-based performance assessment as one component, it will require the development of a corresponding psychometric theory that describes how we best

measure and credential performance, give meaningful feedback to groups at every level, including students, on student achievement, and involve these groups in discussions about implications for improvement.

Our recommendation embodies at least four characteristics for a national assessment system that can be expected to contribute to the investment of institutions in the process, and to ensure that the system continuously re-examines and re-articulates its components and principles (Mentkowski, 1991a). A first characteristic is the design of assessment processes that rest on emerging educational assumptions and that are characteristic of education's best practices. A second is a description of conceptual elements or principles that must be present and that define "assessment." Third, one would expect the assessment system to be implemented, and one might generate some guidelines for implementation that would lead to "how to design and do assessment" or "principles of good practice." Such descriptions would require specification of evidence and definitions of validity that would, in turn, imply elaboration of a psychometric theory. Immediately important questions ensue:

- **Do assessment assumptions and principles hold up? Are values shared?**

A fourth characteristic addresses these questions: that research and evaluation efforts symbiotic with the first three characteristics work at validating these educational assumptions and conceptual elements. One might also evaluate the system's practices and strategies. Our expectation is that the integrity and credibility of a national system will rest on continuous re-examination and re-articulation of its components and principles.

There are some emerging expectations and assumptions underlying assessment that we believe should distinguish its current form.

- **How will a national assessment system with multiple purposes, functions, uses, and users contribute to coherence across educational contexts?**

Assessment should contribute to coherence in a particular context: course, curriculum, department, institution, state, national. At the individual student level, assessment is expected to "pull together" student abilities in a set of performances so that student learning outcomes can be judged and improved through feedback. At the institutional level, assessment is expected to become a feedback system that generates ongoing information for various but related uses. An individual student who experiences an assessment can use it to integrate his or her various abilities and knowledge into a demonstration of "outcomes." An institution can tap its institutional assessment process for a synthesis of student outcomes for improving curricula, accountability, and accrediting purposes. Then, institutions might also contribute to a larger and richer picture of college student performance.

Can a national assessment system serve the kind of coherent, integrative function that it can at the college level? It could, if it is made integral to teaching and learning. It could, if the elements and principles that undergird assessment are linked to those teaching and learning assumptions that lead the educational reform movement.

In sum, we argue here that ability-based performance assessment can contribute to educational reform when it relies on specified educational assumptions about learning and development. Exercising explicit educational assumptions at Alverno has meant the development of ability-based performance assessment, within the interdisciplinary context of a liberal arts college with an emphasis on professional preparation, that enables graduates to transfer abilities to work, citizenship, and service.

This experience has led us to identify a final expectation for assessment: assessment has multiple purposes, functions, uses, and audiences within an interdisciplinary institution. It is likely that a national assessment system will find that it too will have multiple purposes, functions, uses, and audiences within a context that is not only interdisciplinary, but that brings to bear inter-institutional concerns and perspectives from the larger pluralistic society. Because of this diversity, it becomes even more useful and necessary to focus on identifying the components of a conceptual framework that would lend coherence to a national assessment system. To do so is necessary if that system is to keep viable the essentials for improving learning—for college students, for institutions for the sake of the student, and thus ultimately for the civic, as well as the growth potential of our country. To do so is to take understanding from the past, to reinvest energy from the present, and to build growth into the future. Clearly, many questions remain unanswered. At least there are signs, throughout the country, of willingness to make a seemingly insurmountable task less so, by beginning to surmount it. The recommendations of this paper have a focus that can be simply stated: make acceptable national connections with the day-to-day learning of every student in order to assure its continuing improvement.

REFERENCES

Albert, L. S. (1991, September). Interview with Kay McClenney and Frank Newman of the Education Commission of the States: Reclaiming the public trust. AAHE Monitor, pp. 3-8.

Albro, D., Devitt, M., Salem, G., Sharkey, S., & Wojno, M. (1990). Social science department review. Milwaukee, WI: Alverno Productions.

Alverno College Assessment Committee/Office of Research and Evaluation. (1980). Validating assessment techniques in an outcome-centered liberal arts curriculum: Insights from the evaluation and revision process. ERIC Document Reproduction Service No. ED 239 559)

Alverno College Educators. (1986). Faculty handbook on learning and assessment (rev. ed.). Milwaukee, WI: Alverno Productions.

Alverno College Faculty. (1984). The volunteer assessor at Alverno College (rev. ed.). Milwaukee, WI: Alverno Productions.

Alverno College Faculty. (1985a). Assessment at Alverno College (rev. ed.). Milwaukee, WI: Alverno Productions.

Alverno College Faculty. (1985b). Liberal learning at Alverno College (rev. ed.). Milwaukee, WI: Alverno Productions.

Alverno College Faculty. (1990). Advanced outcomes in the major and support areas. Milwaukee, WI: Alverno Productions.

Alverno College/FIPSE Assessment Project. (1987). Faculty consortium for assessment design. Judeen Schulte, Project Director.

Alverno College Nursing Faculty. (1979). Nursing education at Alverno College: A liberal arts model. Milwaukee, WI: Alverno Productions.

Alverno College Office of Research and Evaluation/Assessment Committee. (1982). Validating assessment techniques in an outcome-centered liberal arts curriculum: Integrated competence seminar. (ERIC Document Reproduction Service No. ED 239-560)

Alverno College Office of Research and Evaluation/Assessment Committee. (1986). Student as Learner Inventory. Milwaukee, WI: Alverno Productions.

Alverno College Office of Research and Evaluation/Assessment Committee. (1989). Putting the validation process to work: A series of strategies for establishing the validity of faculty-designed performance assessment instruments. Milwaukee, WI: Alverno Productions.

Alverno College Research and Evaluation Committee. (1986). Teacher as researcher: Some elements that describe teacher inquiry that improves teaching. Milwaukee, WI: Alverno Productions.

Association of American Colleges. (1991). Liberal learning and the arts and sciences major: Reports from the fields. Washington, DC: Author.

Astin, A. W. (1977). Four critical years. San Francisco: Jossey-Bass.

Astin, A. W. (1991). Assessment for excellence: The philosophy and practice of assessment and evaluation in higher education. New York: Macmillan.

Banta, T. W. (Ed.). (1988). Implementing outcomes assessment: Promise and perils. New Directions for Institutional Research, 70, 95-98. San Francisco: Jossey-Bass.

Collins, L. M., & Horn, J. L. (in press). Best methods for the analysis of change? Washington, DC: American Psychological Association Press.

Cromwell, L. (Ed.). (1986). Teaching critical thinking in the arts and humanities. Milwaukee, WI: Alverno Productions.

Deahl, R. J. (1990). Developing reflective pedagogical practices: Linking student and teacher reflections on classroom learning. Proposal submitted to the National Center on Adult Learning.

DeBack, V., & Mentkowski, M. (1986). Does the baccalaureate make a difference: Differentiating nurse performance by education and experience. Journal of Nursing Education, 25(7), 275-285.

Deemer, D. (in press). Illustrations of student voices: Studying the outcomes of college from the perspectives of students. Milwaukee, WI: Alverno Productions.

Deemer, D., & Mentkowski, M. (in press). Mapping a learning process: Observations from Alumnae Perspectives Interviews. Milwaukee, WI: Alverno Productions.

Deemer, D., & Mentkowski, M. (1990). Studying the outcomes of college from the perspective of the student: A rationale and method for the student/alumna perspectives interview approach. Milwaukee, WI: Alverno Productions.

DeWitt, K. (1991, August). In Vermont schools, test on how well students think draws interest. New York Times, Education Life.

Diez, M. (1990). A thrust from within: Reconceptualizing teacher education at Alverno College. Peabody Journal of Education, 65(2), 4-18.

Doherty, A., Mentkowski, M., & Conrad, K. (1978). Toward a theory of undergraduate experiential learning. New Directions for Experiential Learning (Vol. 1). San Francisco: Jossey-Bass.

Earley, M., Mentkowski, M., & Shafer, J. (1980). Valuing at Alverno: The valuing process in liberal education. Milwaukee, WI: Alverno Productions.

Feldman, K. A., & Newcomb, T. M. (1969). The impact of college on students. San Francisco: Jossey-Bass.

Fong, B. (1988). Old wineskins: The ACC external examiner project. Liberal Education, 74(3), 12-16.

Forrest, A., & A study group on protfolio assessment. (1990). Time will tell: Portfolio-assessed assessment of general education. Washington, DC: American Association for Higher Education.

Friedman, M., Mentkowski, M., Deutsch, B., Shovar, M., & Allen, Z. (1982). Validating assessment techniques in an outcome-centered liberal arts curriculum: Social Interaction Generic Instrument. Milwaukee, WI: Alverno Productions. (ERIC Document Reproduction Service No. ED 239 558)

Friedman, M., Mentkowski, M., Earley, M., Loacker, G., & Diez, M. (1980). Validating assessment techniques in an outcome-centered liberal arts curriculum: Valuing and Communications Generic Instruments. Milwaukee, WI: Alverno Productions. (ERIC Document Reproduction Service No. ED 239-557)

Giencke-Holl, L., Mentkowski, M., Much, N., Mertens, S., & Rogers, G. (1985, April). Evaluating college outcomes through alumnae studies: Measuring post-college learning and abilities. Paper presented at the annual meeting of the American Educational Research Association, Chicago.

Halonen, J. S. (Ed.). (1986). Teaching critical thinking in psychology. Milwaukee, WI: Alverno Productions.

Heath, D. (1977). Maturity and competence. New York: Gardner Press.

Hutchings, P., & Wutzdorff, A. (Eds.). (1988). Knowing and doing: Learning through experience. New Directions for Teaching and Learning, 35. San Francisco: Jossey-Bass.

Jacob. P. (1957). Changing values in college. New York: Harper & Row.

Kennedy, M. (1988). Abilities that define computer studies. Presentation at the May 1988 Alverno College Faculty Institute, Teachers as researchers: Learning from students and colleagues (videotape). Milwaukee, WI: Alverno Productions.

Knight, M. E., Lumsden, D., & Avioli, P. S. (1991). Assessing outcomes in the major: The Kean College model. Panel presentation at the American Association of Higher Education Assessment Forum, San Francisco, CA.

Kramp, M. K., & Humphreys, W. L. (1990). Narrative as a mode of self-assessment for adult learners. Proposal submitted to the National Center on Adult Learning.

Light, R. J., Singer, J. D., & Willett, J. B. (1990). By Design. Cambridge, MA: Harvard University Press.

Loacker, G. (1988). Faculty as a force to improve instruction through assessment. In J. McMillan (Ed.), Assessing students' learning. New Directions for Teaching and Learning, 34, 19-32. San Francisco: Jossey-Bass.

Loacker, G., Cromwell, L., Fey, T., & Rutherford, D. (1984). Analysis and communication at Alverno: An approach to critical thinking. Milwaukee, WI: Alverno Productions.

Loacker, G., Cromwell, L., & O'Brien, K. (1986). Assessment in higher education: To serve the learner. In C. Adelman (Ed.), Assessment in higher education: Issues and contexts (Report No. OR 86-301, pp. 47-62). Washington, DC: U.S. Department of Education.

Loacker, G., & Jensen, P. (1988, summer). The power of performance in developing problem solving and self assessment abilities. Assessment and Evaluation in Higher Education, 13(2), 128-150.

Loacker, G., Loveland, E., McElroy, E. W., & Mentkowski, M. (1991, June). Developing valid faculty-designed assessments: Many paths to a common goal. Presentation at the American Association for Higher Education Assessment Forum, San Francisco.

Loacker, G., & Palola, E. (Eds.). (1981). Revitalizing academic disciplines by clarifying outcomes. New Directions for Experiential Learning (Vol. 12). San Francisco: Jossey-Bass.

Loacker, G., Wutzdorff, A., Farmer, D. W., O'Brien, J., Barnett, T. V., & Brown, J. C. (1988, June). Common principles from different practices. In Assessment to serve learners: Principles and methods. Presentation at the Third National Conference on Assessment in Higher Education ponsored by the American Association for Higher Education, Chicago.

Loevinger, J., & Wessler, R. (1970). Measuring ego development I: Construction and use of a sentence completion test (Vol. 1). San Francisco: Jossey-Bass.

McClelland, D. C. (1976). A guide to job competency assessment. Boston: McBer and Company.

McClelland, D. C. (1978). Behavioral event interview. Boston: McBer and Company.

Mentkowski, M. (1986, April). Student involvement in learning and assessment. Paper presented at the annual meeting of the American Educational Research Association, San Francisco. Milwaukee, WI: Alverno Productions.

Mentkowski, M. (1988). Paths to integrity: Educating for personal growth and professional performance. In S. Srivastva & Associates (Eds.), Executive integrity: The search for high human values in organizational life (pp. 89-121). San Francisco: Jossey-Bass.

Mentkowski, M. (1988, October). Faculty and student involvement in institutional assessment. In T. Banta (Chair), Politics and Assessment in Higher Education. Paper presented at the annual meeting of the American Evaluation Association, New Orleans.

Mentkowski, M. (1989, June). Catching theory up with practice: Establishing the validity and integrity of higher education assessment (Part I). Paper presented at the Fourth National Conference on Assessment in Higher Education sponsored by the American Association for Higher Education, Atlanta.

Mentkowski, M. (1990, August). Higher education assessment: Connecting to its conceptual base. In Facing the challenge of student, program, and institutional assessment. Symposium conducted at the meeting of the 98th Annual Convention of the American Psychological Association, Boston.

Mentkowski, M. (1990b). Issues in the analysis of change in higher education assessment. In M. Mentkowski (chair), Individuals and patterns: Analyzing change over time. Symposium at the Fifth National Conference on Assessment in Higher Education sponsored by the American Association for Higher Education, Washington, DC.

Mentkowski, M. (1991a, April). Assessment: How is it different? How can we know? Why do we care? Essay prepared for the meeting of the American Association of Higher Education Council on Assessment, Washington, DC.

Mentkowski, M. (1991b, June). Analyzing and understanding patterns of change: Where are we headed? Symposium conducted at the Sixth National Conference on Assessment in Higher Education, San Francisco.

Mentkowski, M. (1991c). Creating a context where institutional assessment yields educational improvement. Journal of General Education, 40, 255-283.

Mentkowski, M. (1991d). Designing a national assessment system: Assessing abilities that connect education and work. Paper commissioned by the U.S. Office of Education, National Center for Education Statistics, in response to the National Education Goals Panel: America 2000: An Education Strategy. Milwaukee, WI: Alverno Productions.

Mentkowski, M., & Chickering, A. W. (1987). Linking educators and researchers in setting a research agenda for undergraduate educators. The Review of Higher Education, 11(2), 137-160.

Mentkowski, M., & Doherty, A. (1984a). Abilities that last a lifetime: Outcomes of the Alverno experience. AAHE Bulletin, 36(6), 5-6, 11-14.

Mentkowski, M., & Doherty, A. (1984b). Careering after college: Establishing the validity of abilities learned in college for later careering and professional performance (Final Report to the National Institute of Education) (rev. ed.). Milwaukee, WI: Alverno Productions. (ERIC Document Reproduction Service No. ED 239 556)

Mentkowski, M., & Loacker, G. (1985). Assessing and validating the outcomes of college. In P. Ewell (Ed.), Assessing educational outcomes. New Directions for Institutional Research, 47, 47-64. San Francisco: Jossey-Bass.

Mentkowski, M., Much, N., & Giencke-Holl, L. (1983). Careering after college (Fifth Report to Participants in a Longitudinal Study of College Outcomes). Milwaukee, WI: Alverno Productions.

Mentkowski, M., O'Brien, K., McEachern, W., & Fowler, D. (1982). Developing a professional competence model for management education (Final Report to the National Institute of Education: Research Report No. 10). Milwaukee, WI: Alverno Productions. (ERIC Document Reproduction Service No. ED 239 566)

Mentkowski, M., & Rogers, G. (1985). Longitudinal assessment of critical thinking in college: What measures assess curricular impact? Milwaukee, WI: Alverno Productions.

Mentkowski, M., & Rogers, G. (1986). Assessing critical thinking. In L. S. Cromwell (Ed.), Teaching critical thinking in the arts and humanities (pp. 117-128). Milwaukee, WI: Alverno Productions.

Mentkowski, M., & Rogers, G. (1988). Establishing the validity of measures of college student outcomes. Milwaukee, WI: Alverno Productions.

Mentkowski, M., Rogers, G., Deemer, D., Ben-Ur, T., Reisetter, J., Rickards, W., & Talbott, M. (1991, April). Understanding abilities, learning and development through college outcomes studies: What can we expect from higher education assessment? Symposium presented at the annual meeting of the American Educational Research Association, Chicago, IL.

Mentkowski, M., & Strait, M. (1983). A longitudinal study of student change in cognitive development, learning styles, and generic abilities in an outcome-centered liberal arts curriculum (Final Report to the National Institute of Education: Research Report No. 6). Milwaukee, WI: Alverno Productions. (ERIC Document Reproduction Service No. ED 239 562)

Messick, S. (1980). Test validity and the ethics of assessment. American Psychologist, 35(11), 1012-1027.

Much, N., & Mentkowski, M. (1982). Student perspectives on liberal learning at Alverno College: Justifying learning as relevant to performance in personal and professional roles (Final Report to the National Institute of Education: Research Report No. 7). Milwaukee, WI: Alverno Productions. (ERIC Document Reproduction Service No. ED 239 563)

Much, N., & Mentkowski, M. (1984). Student perspectives on learning and development in college. Milwaukee, WI: Alverno Productions.

O'Brien, K., Matlock, M. G., Loacker, G., & Wutzdorff, A. (1991). Problem solving at Alverno College. In D. Boud & G. Feletti (Eds.), The challenge of problem-based learning. United Kingdom: Kogan Page.

Olds, B. M., Griffin, J. B., & Gaffney, G. (1991). Portfolio assessment in professional programs. Panel presentation at the American Association of Higher Education Assessment Forum, San Francisco, CA.

Pace, C. R. (1979). Measuring outcomes of college: Fifty years o. findings and recommendations for the future. San Francisco: Jossey-Bass.

Pascarella, E. T., & Terenzini, P. T. (1991). How college affects students: Findings and insights from twenty years of research. San Francisco: Jossey-Bass.

Read, J. (1980). Alverno's college-wide approach to the development of valuing. New Directions for Higher Education (Vol. 31). San Francisco: Jossey-Bass.

Read, J. (1985). Assessment/evaluation and costs. Milwaukee, WI: Alverno Productions.

Read, J., & Sharkey, S. R. (1985). Alverno College: Toward a community of learning. In A. Levine and J. S. Green (Eds.)., Opportunity in adversity (pp. 195-214). San Francisco: Jossey-Bass.

Reisetter, J., & Sandoval, P. (1987, October). Flexible procedures for efficiently maximizing participation in a longitudinal study. Paper presented at the annual meeting of the Midwest Educational Research Association. Milwaukee, WI: Alverno Productions.

Rest, J. (1979). Development in judging moral issues. Minneapolis: University of Minnesota Press.

Rickards, W., Cromwell, L., Diez, M., Rogers, G., & Mentkowski, M. (1991, November). Classroom, curriculum, college and beyond: Intra-institutional assessment strategies that meet accreditation and evaluation purposes. Presentation at the annual meeting of the American Evaluation Association, Chicago, IL.

Rogers, G. (1988, April). Validating college outcomes with institutionally developed instruments: Issues in maximizing contextual validity. Paper presented at the annual meeting of the American Educational Research Association, New Orleans. Milwaukee, WI: Alverno Productions.

Rogers, G. (1991, June). Typical strategies for statistically analyzing inter- and intra-individual change through time in relation to the curriculum: How well are they working? In M. Mentkowski (Chair), Analyzing and understanding patterns of change: Where are we headed? Symposium conducted at the Sixth National Conference on Assessment in higher Education, San Francisco.

Rogers, G., & Talbott, M. (1991). Generic ability codes for scoring the Behavioral Event Interview of Alverno alumnae. Milwaukee, WI: Alverno Productions.

Rogers, G., & Talbott, M. (in progress). Generic ability scoring manual for Behavioral Event Interview of Alverno alumnae. Milwaukee, WI: Alverno Productions.

Rogers, G., & Talbott, M. (in press). Alverno faculty validation of scoring of alumnae generic abilities through the Behavioral Event Interview. Milwaukee, WI: Alverno Productions.

Schall, C., Guinn, K., Qualich, R., Kramp, M. K., Schmitz, J., & Stewart, K. (1984). Competence and careers: A study relating competences acquired in college to career options for the liberal arts graduate. Milwaukee, WI: Alverno Productions.

Watson, G., & Glaser, E. (1964). Critical thinking appraisal. New York: Harcourt, Brace, & Jovanovich.

Willett, J. B. (1988). Questions and answers in the measurement of change. In Ernst Rothkopf (Ed.), Review of Research in Education, 15, 345-422.

Willett, J. B. (1989). Some results on reliability for the longitudinal measurement of change: Implications for the design of studies of individual growth. Educational and Psychological Measurement, 49(3), 587-602.

Willett, J. B. (1990). Measuring change: The difference score and beyond. In H. J. Walberg and G. D. Haertel (Eds.), International Encyclopedia of Educational Evaluation. Oxford: Pergamon Press.

Winter, D. (1976). The Test of Thematic Analysis. Boston: McBer and Company.

Woodward, C. A. (1984). Assessing the relationship between medical eduction and subsequent performance. Assessment and Evaluation in Higher Education, 9(1), 19-29.

Appendix A:

Abilities that
- **involve the whole person**
- **are teachable**
- **can be assessed**
- **transfer across settings**
- **are continually re-evaluated and re-defined**

ALVERNO COLLEGE
Milwaukee, Wisconsin

ABILITY-BASED LEARNING PROGRAM

The curriculum is an ability-based, outcome-oriented approach to liberal arts/professional education. To earn a degree at Alverno College a student demonstrates the eight broad abilities listed below, at increasingly complex levels, in general education and in her areas of specialty.

These abilities constitute liberal education at the college and undergird and infuse advanced study in the disciplines and professions. Within the curriculum of a given major, the student develops the abilities according to the distinctive requirements of the disciplines and professions.

Throughout her course of studies, the student participates in performance-based assessments and learns to assess herself. Her progression toward a degree is based upon these assessments, both internal and external.

With demonstrated achievement at each level the student receives one level unit. For a Bachelor's degree, in addition to 32 units awarded when she has demonstrated the first four levels of each of the eight abilities, the student must achieve another 8 units, at least one of them at level 6. Advanced levels of any given ability require more time and effort to achieve than lower ones. For an Associate of Arts degree in General Studies, a student demonstrates her ability at the first four levels in each of the eight areas.

Abilities and Developmental Levels

1 Develop communication ability (effectively send and respond to communications for varied audiences and purposes)

Level 1 — Identify own strengths and weaknesses as communicator
Level 2 — Show analytic approach to effective communicating
Level 3 — Communicate effectively
Level 4 — Communicate effectively making relationships out of explicit frameworks from at least three major areas of knowledge

In majors and areas of specialization:
Level 5 — Communicate effectively, with application of communications theory
Level 6 — Communicate with habitual effectiveness and application of theory, through coordinated use of different media that represent contemporary technological advancement in the communications field

IN WRITING, READING, SPEAKING, LISTENING, USING MEDIA, QUANTIFIED DATA, AND THE COMPUTER

2 Develop analytical capabilities

Level 1 — Show observational skills
Level 2 — Draw reasonable inferences from observations
Level 3 — Perceive and make relationships
Level 4 — Analyze structure and organization

In majors and areas of specialization:
Level 5 — Establish ability to employ frameworks from area of concentration or support area discipline in order to analyze
Level 6 — Master ability to employ independently the frameworks from area of concentration or support area discipline in order to analyze

3 Develop workable problem-solving skill

Level 1 — Articulate and evaluate own problem-solving process
Level 2 — Define problems or design strategies to solve problems using discipline-related frameworks
Level 3 — Select or design appropriate frameworks and strategies to solve problems
Level 4 — Implement a solution and evaluate the problem-solving process used

In majors and areas of specialization:
Level 5 — Design and implement a process for resolving a problem which requires collaboration with others
Level 6 — Demonstrate facility in solving problems in a variety of situations

Appendix A (continued):

4 Develop facility in making value judgments and independent decisions

Level 1 — Identify own values

Level 2 — Infer and analyze values in artistic and humanistic works

Level 3 — Relate values to scientific and technological developments

Level 4 — Engage in valuing in decision-making in multiple contexts

In majors and areas of specialization:

Level 5 — Analyze and formulate the value foundation/framework of a specific area of knowledge, in its theory and practice

Level 6 — Apply own theory of value and the value foundation of an area of knowledge in a professional context

5 Develop facility for social interaction

Level 1 — Identify own interaction behaviors utilized in a group problem-solving situation

Level 2 — Analyze behavior of others within two theoretical frameworks

Level 3 — Evaluate behavior of self within two theoretical frameworks

Level 4 — Demonstrate effective social interaction behavior in a variety of situations and circumstances

In majors and areas of specialization:

Level 5 — Demonstrate effective interpersonal and intergroup behaviors in cross-cultural interactions

Level 6 — Facilitate effective interpersonal and intergroup relationships in one's professional situation

6 Develop global perspectives

Level 1 — Assess own knowledge and skills to think about and act on global concerns

Level 2 — Analyze global issues from multiple perspectives

Level 3 — Articulate understanding of interconnected local and global issues

Level 4 — Apply frameworks in formulating a response to global concerns and local issues

In majors and areas of specialization:

Level 5 — Generate theoretical and pragmatic approaches to global problems, within a disciplinary or professional context

Level 6 — Develop responsibility toward the global environment in others

7 Develop effective citizenship

Level 1 — Assess own knowledge and skills in thinking about and acting on local issues

Level 2 — Analyze community issues and develop strategies for informed response

Level 3 — Evaluate personal and organizational characteristics, skills and strategies that facilitate accomplishment of mutual goals

Level 4 — Apply her developing citizenship skills in a community setting

In majors and areas of specialization:

Level 5 — Show ability to plan for effective change in social or professional areas

Level 6 — Exercise leadership in addressing social or professional issues

8 Develop aesthetic responsiveness: Involvement with the arts

Level 1 — Express response to selected arts in terms of their formal elements and personal background

Level 2 — Distinguish among artistic forms in terms of their elements and personal response to selected art works

Level 3 — Relate artistic works to the contexts from which they emerge

Level 4 — Make and defend judgments about the quality of selected artistic expressions

In majors and areas of specialization:

Level 5 — Choose and discuss artistic works which reflect personal vision of what it means to be human

Level 6 — Demonstrate the impact of the arts on your own life to this point and project their role in personal future

Appendix B: Advanced Outcomes in Selected Major Areas at Alverno College[1]

BIOLOGY

1. Shows the basis and limitations of scientific analyses.
2. Demonstrates proficient library and computer skills in data gathering and analysis.
3. Designs, conducts, and communicates biological experiments that meet standards for publication.
4. Solves complex biological problems drawing on concepts from several different areas and working independently and collaboratively.
5. Develops value judgments based on ethical frameworks in the conduct of biology and the application of biology in society.
6. Applies concepts from biology to the analysis of environmental problems and issues.
7. Performs appropriate mathematical and statistical analysis.
8. Articulates judgments between competing scientific theories.
9. Applies one's learning in an off-campus, professional setting.

ENGLISH

1. Uses frameworks to analyze, evaluate and place in context literary works from various cultures and genres.
2. Communicates an understanding of literary criticism and questions its assumptions.
3. Participates in the academic discourse of the discipline of English.
4. Demonstrates personal and intellectual engagement in responding to literary works.
5. Articulates understanding of the impact her literary study has on her life.
6. Demonstrates her understanding of the structure and history of the language, linguistic development in England and America, and major grammatical systems.

BUSINESS & MANAGEMENT

1. Critical thinking/communicating: Accurately uses theoretical frameworks from functional business areas to interpret and analyze business situations and effectively communicate the analysis in a variety of business contexts.
2. Enterprising/problem solving: Takes initiative in identifying and solving problems or pursuing opportunities for organizational growth or improvement.
3. Interacting/leading: Uses organizational and management theory to interact effectively in organizational contexts that require leadership of groups or other types of inter-personal interactions.

Alverno College Faculty, (1990)

---

[1] Advanced outcomes have been identified for all other Major areas at Alverno College:
Art Education, Art Therapy, Studio Art, Chemistry, Elementary Education, History, Mathematics, BA in Music Culture, Music Education, Music Performance and Pedagogy, Music Therapy, Nursing, Philosophy, Professional Communication, Psychology, Religious Studies, Social Science

Appendix C: Characteristics of Educational-Framework Driven Institutional Assessment

Since its inception in 1976, the Office of Research and Evaluation has investigated a series of questions at the behest of the faculty, with special attention to linking the outcomes of college to the curriculum, establishing the validity of assessment techniques and the assessment process, and demonstrating the link between college learned abilities and alumnae performance in the world of work, personal life, service and citizenship. For example:

- How are abilities best defined, learned and taught?
- Does involvement in the ability-based performance assessment process lead to learning? Are assessment instruments valid?
- Are changes in abilities/outcomes over time linked to the curriculum? Who benefits and why?
- Do graduates transfer abilities and learning beyond college to work, personal life, service and citizenship?
- Are student and alumna abilities/outcomes "good" compared to the "internal" standards and expectations of faculty, students, and alumnae? Are outcomes "good" compared to "external" disciplinary and professional outcomes and expectations, the performance of effective alumnae and other outstanding professionals, and compared to what is possible for humans to achieve across the lifespan?

Approaches

The Office concentrates on a number of approaches that have been gradually developed to carry out its mission and to respond to these questions (Mentkowski & Doherty, 1984b; Mentkowski et al., 1991). Research, evaluation and measurement strategies are expected to be consistent with the educational values, assumptions and principles that inform the curriculum, including its ability-based performance assessment theory, and the latter's psychometric theory.

Because Alverno's mission is the personal and professional development of its students, research and evaluation questions reflect a student-centered institution and a concern with whether and how each individual student demonstrates this development. Alverno educational frameworks include a coherent, developmental, ability-based curriculum with special attention to experiential, self-sustained learning and assessment-as-learning. Student and alumna outcomes of the curriculum (development, learning and abilities) are the focus of the institutional assessment enterprise.

This student-centered, outcome-centered focus of the institution means that information from institutional assessment has a central purpose: to enhance student development, learning and abilities. Information must be both useful and general. At the program or institutional level, information indirectly benefits individual students. But clearly, information is expected to be used for student benefits. At the same time, the broader picture of student achievement that emerges is multifaceted and collective, a backdrop against which faculty can interpret an individual student's growth. Pictures that accrue from aggregated sets of information over time are expected to inform curriculum development, but also to question the philosophy and principles upon which it is based, and to demonstrate institutional effectiveness through descriptions of student achievement to various external constituencies. Still, these collective pictures are expected to be easily transformed into intra- and inter-individual patterns that do not lose sight of the individual student's development (Mentkowski, 1990b; Rogers, 1991). The pictures are impressionistic in that, as one steps away, a holistic scene appears. As one looks more closely, each dab of paint, each individual color, each brush stroke is evident.

Appendix C (continued): Characteristics of Educational-Framework Driven Institutional Assessment

To carry out educational framework-driven institutional assessment, the Office and its faculty committee developed interactive, interdisciplinary processes that are effective in developing a collaborative interplay that engages faculty questions and contributions (Mentkowski, 1988, October). By engaging the whole faculty in question-asking, and by tapping existing faculty groups related to particular issues, the Office formulates research questions. Nor does faculty leadership and investment stop there. Over the years, faculty have served in various capacities as adjunct members of the research team, as advisors, as interpreters of results, and so on. The Research and Evaluation Committee, comprised of senior-level faculty and administrators, is a springboard and interpreter at the institutional level of question-asking and interpretation. Findings and their interpretations are an outcome of this interplay all the way through the process: formulating questions, data collection and analysis, interpretation of results, and making meaning out of the results for curriculum development. Thus, our institutional assessment system involves the "users" of the information at every stage of design, implementation, implication, interpretation and use of results.

Our question-asking considers external sources as well. Since 1985, we have been active in creating strategies for question-asking that work to integrate research, evaluation and practice at the national level, in that we actively co-lead and support the AAHE Research Forum (Mentkowski & Chickering, 1987), which has generated a research agenda each year since 1986. This involvement ensures that Alverno's research, evaluation and assessment activities are in tune with national questions and issues that educators feel should be the subject of inquiry.

The Office has also created methods that result in sustained participation of samples of students and alumnae in research and evaluation activities that support curriculum and institutional assessment. Key elements are providing extensive educational and disciplinary rationales, and immediate benefits to these participants (such as feedback on instruments they complete that are external to the performance assessment system) (Mentkowski, 1988, October; Mentkowski & Strait, 1983; Reisetter & Sandoval, 1987).

The Office is expected to meet internal and external tests of its value, impact, validity and effectiveness by demonstrating that findings are actually used by the faculty to challenge and inform the educational frameworks of the college, to refine the curriculum, and to promote an atmosphere of continuous improvement. The Office also measures its own value, impact and effectiveness through external tests including external peer review via advisory panel, presentation, publication, consulting, commissioned reviews, and conducting a workshop as part of the College's annual workshop on ability-based performance assessment.

Strategies

Recall the goal to demonstrate the value, impact, validity, and effectiveness of Alverno's educational enterprise. This means using multi-level, triangulated designs to enable multiple, internal and external comparisons (Mentkowski & Doherty, 1984b; Mentkowski & Loacker, 1985). The Office has selected instruments and methods from outside the College that represent a number of external theoretical frameworks in abilities, learning and human development. The Office has also developed multiple instruments and methods on its own.

The "What Have We Learned" section provides evidence for the learned principles that form the basis for recommendations for a national assessment system. Results are drawn from strategies that describe/ascribe/evaluate/validate student development, abilities, and learning through:

Appendix C (continued): Characteristics of Educational-Framework Driven Institutional Assessment

(a) longitudinal analysis of change as a result of curriculum (qualitative/quantitative)

(b) analysis of professional/alumna abilities in relation to work, personal life, citizenship and service

(c) evaluation of general education and the major field

(d) educator-as-researcher/inquirer studies; and

(e) evaluating/validating ability-based performance assessment (for example, defining contextual validity, developing strategies for validating faculty-designed performance assessment measures, defining criteria for "good" assessment).

A first strategy is the longitudinal analysis of change as the result of curriculum (Mentkowski, 1990b; Mentkowski & Strait, 1983), which is a strategy for both research and evaluation of the broad outcomes of college. This strategy provides for more short-term curriculum evaluation benefits at its earlier phases when generating information on current students. It provides for more long-term research benefits at its later phases when describing longitudinal antecedents of alumna abilities, learning and development. Longitudinal strategies employ both quantitative and qualitative methods. Further, these strategies have used instruments and methods that are drawn from a variety of theoretical frameworks in cognitive development, learning styles, and broad abilities or competences. Because the approach draws from a range of theoretical frameworks that relate to faculty educational frameworks in student development, learning and abilities for external comparisons, there is potential for contributions to discipline-based theory and method in adult learning and development.

A second strategy is analysis of professional/alumna abilities, to describe ability models of outstanding professionals who are and are not Alverno graduates, in order to enable faculty to define and refine ability definitions, instruction and assessment, and to evaluate their professional or major fields. Because graduates include examples of activities in other areas than paid employment, faculty have a picture of abilities that are used in personal (e.g., child-rearing; graduate learning), service, and professional domains.

A third strategy builds on the first and second, and extends it for more immediate benefits. This is called evaluation of general education and the major field. Here studies generate ability models for outstanding professionals in each of the three largest major field areas: nursing, management and teaching (DeBack & Mentkowski, 1986; Diez, 1990; Mentkowski, 1988; Mentkowski et al., 1982). Currently, these strategies include inter- and intra-individual pattern analyses of student performance throughout the major, using data generated from faculty-designed external assessment measures, including portfolio assessments (Mentkowski et al., 1991; Rickards, Cromwell, Diez, Rogers & Mentkowski, 1991).

A fourth strategy is "Educator as Researcher/Inquirer" studies (Alverno College Research and Evaluation Committee, 1986). Here individual faculty members or groups of faculty conduct research projects within or across classes for the purposes of direct intervention in teaching and learning activities, so as to improve the immediate relationships between instruction and student learning (Deahl, 1990; Kramp and Humphreys, 1990).

A fifth strategy is evaluating and validating the ability-based performance assessment for individual student development, which includes faculty-designed performance assessment measures. We have developed a workable definition of contextual validity (Mentkowski, 1989; Mentkowski & Rogers, 1985; Rogers, 1988) and strategies for validating faculty-designed performance assessment measures (Alverno College Office of Research and Evaluation/Assessment Committee, 1989). The latter have been field-tested with a range of colleges and universities in a FIPSE-funded project (Alverno College/FIPSE Assessment Project, 1987). Finally, as we mentioned earlier, the Office works to define criteria for "good" assessment and apply them to its work (Mentkowski, 1989).

Appendix D: Developing Perspectives on the Role of Criteria for Student Understanding of Independent Learning and Self-Assessment What Value and Benefit do Assessment Criteria Have for Students?

CRITERIA MAKE INDEPENDENT LEARNING POSSIBLE	CRITERIA MAKE SELF-ASSESSMENT POSSIBLE
.... from content to abilities .... from vague to explicit to flexible interpretation .... from external to internal self-assessment	.... from grades to criteria .... from quantity to quality .... from opinion to evidence
BEGINNING STUDENT	
• Sees learning objectives as vague directions for what to learn • Finds explicit directions too picky	• Sees assessor judgments as arbitrary and vague and dependent on factors beyond own and assessor's control • Finds explicit assessment criteria too picky
• Sees learning objectives as directions for how much content to learn	• Sees assessor judgments as based on standards for how much to learn • Sees number or letter grades as the standards for how close you are to learning enough of the right answers
• Sees competences or abilities as directions for what to do • Asks for explicit directions for what to do to perform, to get validated, or to "pass"	• Sees criteria as feedback on strengths and weaknesses but as vague with little meaning for "passing" • Sees that assessor judgments are based on criteria, but finds interpretation of criteria arbitrary and vague and dependent on personal opinion of the assessor and self • Often doesn't understand why validated or not • Sees criteria expressed as percent of correct response • Worries about motivation to achieve where can pass by just getting by
DEVELOPING STUDENT	
• Sees that criteria given ahead of time tell you what to learn and what to do • Asks for explicit learning objectives and criteria • Sees abilities as steps in a process that you use in school and personal life • Sees learning as a process (you learn how to learn and it doesn't disappear afterwards) • Sees criteria as providing a picture of the ability to perform	• Sees that feedback on strengths and weaknesses provides explicit information on progress and success • Sees criteria as a framework for feedback and self-assessment • Asks for explicit criteria • Motivated to achieve by explicit criteria • Rejects grades as a source of information on progress and success • Sees criteria for assessment as more flexible and ambiguous, as more open to interpretation
ADVANCED STUDENT	
• Sees criteria as one part of a process for learning and assessment • Sees abilities as frameworks for performing and criteria as a picture of the ability for performing and for self-assessment • Sees criteria as a cognitive framework for learning, that enable transfer of learning • Sees criteria as being met in more ways than one, and uses in a flexible way to guide independent learning • Sees criteria as internalized and uses for self-assessment • Creates own criteria	

This handout accompanies a slide-tape of student examples illustrating this framework of student perspectives. The Assessment Committee drew the framework from research on Alverno College students completed by the College's Office of Research and Evaluation (Much & Mentkowski, 1984).

Appendix E: Alverno Students' Developing Perspectives on "Self-Assessment," "Using Feedback," and "Commitment to Improvement" That Lead to Taking Responsibility for Learning and Using Different Ways of Learning.

SELF-ASSESSMENT	USING FEEDBACK	COMMITMENT TO IMPROVEMENT
BEGINNING STUDENT		
• Makes judgments on her own behavior when someone else points out concrete evidence to her • Recognizes that her attitudes affect her work • Recognizes contradictory evaluations of her work • Expects the teacher to take the initiative in recognizing her problems and approaching her about them • Responds to divergent values with self-assessment insights	• At this point, experiences evaluation of her performance as general affirmation or rejection of herself • Her emotional response to evaluation, as of yet, interferes with insight into her performance • Can connect feedback received to subsequent classroom experience	• Knows she should improve, wants to improve, tries to improve in quality ways • Recognizes negative attitudes; expresses willingness to change
DEVELOPING STUDENT		
• Senses when her own performance in a given situation is essentially competent or incompetent • Aware that the learning process requires a change in approach to learning • Knows her strengths • Reflects on a given performance as representative of a pattern in her own behavior • Sees criteria as a framework for feedback and self-assessment • Sees criteria as providing a picture of the ability to perform • Compares self to self, rather than just self to others • Achieves sufficient awareness of self to assess her own abilities and how they contribute to a situation (rather than an undifferentiated sense of how "she" contributed)	• Sees the value in separating emotional response to feedback from more objective stance • Sees that feedback on strengths and weaknesses provides explicit information on progress and success • Accepts criticism and suggestions and follows through	• Thinks about how to improve • Builds on her strengths • Sees that criteria given ahead of time tell you what to learn and what to do • Motivated to achieve by explicit criteria • Performs well in structured situations; follows through if there are external demands • Completes assignments in weak areas; is becoming aware of her weaknesses
ADVANCED STUDENT		
• Sees own abilities apart from a given situation • Sees abilities as frameworks for performing and criteria as a picture of the ability for performing and self-assessment • Emphasizes reliance on self-evaluation and self-assessment • Consistently applies self-awareness of self (therefore, has more knowledge of her abilities—acts accordingly) • Shapes her aspirations realistically, commensurate with her abilities • Gives evidence of internalizing standards of self-assessment • Sets personal standards out of her expectations of her professional needs • Shows interest in her ability relative to other professionals	• Seeks out formative evaluation of her work (doesn't just wait for someone else's summative evaluation) • Self-applies formative evaluations of her work • Acts on feedback • Expects feedback that helps her "take charge" • Expects feedback that helps her see patterns and relationships to her performance in other ability areas	• Knows what she needs to do to improve • Consistently makes an effort to improve processes • Uses resources to help her improve processes • Takes initiative to improve her work, finds help when she needs it

The Alverno College Assessment Committee drew this framework from research on Alverno College students completed by the College's Office of Research and Evaluation (Much & Mentkowski, 1984) and the Department of Business and Management.

U.S. Department of Education - NCES
STUDY DESIGN WORKSHOP
on
Higher Order Thinking and Communication Skills of College Graduates
November 17-19, 1991

Position Paper

"Designing a National Assessment System: Alverno's Institutional Perspective"

Author
Georgine Loaker

Reviewer
Elinor M. Greenberg

Overview

This extremely detailed, thorough and well documented paper and the elegant assessment system it describes, developed over 20 years at Alverno College, reveals what can be done when vision, commitment, perseverance and knowledge are integrated and implemented in a coherent and competent fashion. This document provides the foundation for a number of common elements that could be used in developing a "Coordinated Multi-Option National Assessment and Partnership System" (See review of Capelli) . This paper also gives specific guidance to those who choose the option which I call, the "Development-Based Assessment Option." The latter option has the advantage of building on current efforts already underway (Alverno Network, AAHE Assessment Forum, FIPSE, Perry Network and others). It also offers us the opportunity to utilize significant research that is based on developmental perspectives and philosophies, which are highly regarded and extensively researched, but still relatively unknown and under-utilized.

Useful Measures

Purpose: The suggested dual purpose of improvement and accountability should be adopted for the entire national assessment system, whatever form it takes (Abstract).

Elements: The eight suggested elements should be considered for adoption as criteria by which to develop the entire national assessment system: 1) outcomes,* 2) varied contexts,* 3) feedback* and self-assessment* 4) instruction,* 5) patterns over time, 6) research, 7) supportive context, and, 8) explicit values and theories (Abstract). These eight elements are subsequently collapsed into five key elements.*

Recommendations: The 10 recommendations (Figure 1, p. 8-10) can be further reduced to an outline for use in our

discussion of development-based assessment and can also be used in the overall discussion of a national assessment system.

Justifications: All elements and recommendations proposed in this paper are backed up with detailed descriptions of implementation strategies, lessons learned, research efforts and findings, and issues yet to be resolved. The case for a development-based assessment process is thoroughly and thoughtfully presented in both research and common sense terms.

Primary Advantages: The Alverno developmental model integrates theory and practice; is both general and specific; involves the learners, institution and external entities; is longitudinal and benchmarked; and has potential to provide the foundation for a learner/worker-centered, lifelong learning, training and education system which does not end. If linked to precollegiate and post-collegiate efforts, a truly "seamless" learning system could be envisioned. This might appeal to some K-12 systems as well as to some forward-thinking employers. Its qualitative nature matches the need to view higher order thinking and communication skills along a continuum of development which recognizes the inter-connectedness and inter-dependence of various "skills."

At the sub-skill level, Alverno offers an excellent set of tools, by discipline and at developmental and proficiency levels (Appendix A, B, D and E), which can be used in a variety of assessment approaches. If made available, these tools could be more broadly utilized; for example, in a sub-skill database and directory, in faculty development activities, etc. These tools are applicable not only to a "Development-Based Assessment Option," but to Institution-based, Industry-based, State-based options, and a national data collection effort (proposed by this reviewer), as well.

Primary Disadvantages: The developmental perspective and highly individualized nature of this approach may not be accepted and/or perceived as feasible on a large scale. Issues of cost, faculty development, judgement, and measurement are likely to be seen as barriers to implementation (although this is not necessarily the case). The assessment-for-improvement philosophy may be seen as too long-term in the face of needs and political/public pressures for short-term payoffs and results. Means and ends debates may also take a time.

Since "network organization" thinking is necessary to create a national cluster of such development-based assessment efforts, large and small, and "hierarchical organization" thinking is still the rule in most institutions and businesses, developing new development-based approaches would not be easy. However, a solid base for such a national cluster already exists and it is worth our time and effort to consider how best to incorporate these efforts into our national assessment strategy as a major, not peripheral, component.

Comments by Reviewer

This reviewer has been familiar with the Alverno effort since the early 1970's and has also implemented similar, although

not as comprehensive, development-based research and competency-based systems, both in college and industry settings. Because I know the potential power and effectiveness of such approaches, both from theoretical perspectives and practical experience, I especially appreciate Loaker's efforts to describe and advocate for such an approach. This paper makes a major contribution to our workshop and to our subsequent activities.

Conclusions

Development-based assessment approaches are some of the most sophisticated and promising areas for a national assessment system to pursue. When introduced to academic, community, industry, government and political audiences, these kinds of assessment efforts and the research findings that they net cause significant interest to be generated. They are, I believe, at the "cutting edge" of our efforts to define and make real what is being called "lifelong learning," a concept still at the rhetoric stage.

Therefore, in order to bring these kinds of efforts to broader academic, industry, political and public attention; to reward their success and promise; and to make use of their tools, sub-skill sets, lessons learned and potential power in improving and reconceptualizing a "seamless" system of lifelong learning in America, I strongly recommend that we consider a <u>multiple option approach to a national assessment system</u>. One very important option should be the "Development-Based Assessment Option." My guess is that if training and support were made available, a multitude of similar new efforts would be spawned (especially in small liberal arts institutions and innovative businesses) and that the growing number of institutions and partnerships currently involved would be strengthened in their resolve and effectiveness. My own experience with academic, business and government audiences convinces me of strong interest in this approach.

This is a very important paper that makes a significant contribution to our thinking and to the national assessment effort.

Review of Georgine Loacker's "Designing a National Assessment System: Alverno's Institutional Perspective"

Alverno College's assessment program has been an inspiration to those of us who have struggled to institute state-wide assessment programs. It has not, however, apparently been a practicable model for the institutions, at least in Virginia. This is perhaps understandable at the large, complex universities in which undergraduate education is only one of a number of sometimes conflicting priorities. It is more puzzling in the small liberal-arts or two-year colleges, whose missions focus primarily or even exclusively on educating the undergraduate student. It seems that for a faculty to organize its pedagogy and curriculum around assessment as Alverno has done involves a change in faculty culture that is quite profound, and in Virginia at any rate, faculty have by and large resisted the transformation.

Dr. Loacker proposes to use Alverno's program as a model for a national assessment system. The skills and abilities for which students will be assessed would be developed by faculty in an institution-specific context, although she suggests that there may be some broad ability definitions on which institutions might agree. The measures used to assess performance would be various, again dependent on context, although mention of a national center to "train and validate expert judges" suggests that she believes that the results generated by these various measures might be evaluated according to "performance criteria that have been defined and validated nationally." And the results would be returned to the institution and the individual student in order to improve instruction and learning, although the institutional judges would "discuss results and their implications" in national fora and make "the kinds of comparisons that might provide more insight into standard setting and making standards public."

This approach to assessment raises several major issues. First, it is not clear is how many institutions could have assessment programs that, even while falling short of Alverno's standard, demonstrate a real willingness to link assessment with serious curricular and pedagogical change. Peter Ewell estimates that perhaps 15% of institutions engaging in assessment now (and they are only a fraction of the whole of American higher education) are getting some good out of it. What do we do about the other 85%? How well do we assume that their students are doing? In some ways, this model of assessment is predicated on the very transformation of American higher education to which the national assessment movement is supposed to lead.

The second issue is the problem, even if we had them, of linking thousands of good campus-based programs into a single, coherent national assessment system. Alverno has had some success in getting three institutions to agree to "principles of ability-based performance assessment," and another group of 11 is

"currently synthesizing educational assumptions that are common across their institutions." While encouraging, do these developments suggest that it would be feasible to get even a good proportion of American campuses to agree to a set of outcomes and ways to measure them that could be translated into comparable information? The diversity of approaches to assessment and recent debates over transfer articulation in Virginia ("my general-education program does something quite different from your general-education program") suggest not. Certainly Dr. Loacker does not say, in this paper, what those commonly agreed-upon outcomes and measures might be. She does not seem to assume, for instance, that they would necessarily be the communications, critical-thinking, and problem-solving skills of Goal 5, never mind further specify sets of sub-skills.

Finally, while I share Dr. Loacker's commitment to assessment as a means to improve teaching and learning, early hopes that this approach would be compatible with accountability demands are beginning to seem overly sanguine, As Peter Ewell, again, has pointed out, campus-based assessment reports do not lend themselves to the telling of a story either of student performance or of curricular transformation. The national assessment movement, much as we may wish it were otherwise, seems to be primarily driven by a desire for accountability and hence a desire for relatively simple and comparable data about what American students know and are able to do at any point in time as compared to the previous year or decade. In this context, questions of institutional mission and practices are secondary, and an assessment system characterized by complexity, multiplicity, and a lack of stability is not going to fill the bill.

This is unfortunate, because campus-based assessment is the kind most likely to support rather than damage the teaching-learning relationship, acknowledge what has been a fruitful diversity of institutional purpose in American higher education, honor faculty control over educational matters, and do some good for the individual student. But it is also a system that is costly, redundant, slow-moving, and unlikely to produce easily understood results. It seems to me that any national assessment system should preferably build on the good assessment being done on some campuses now or at least not damage or supplant those assessment programs. But it cannot rely on them, since they are far from ubiquitous and not particularly well suited to the job the national assessment system is supposed to do, give a coherent and probably simplistic picture of higher education's results and progress.

Margaret A. Miller

Review of Designing A National Assessment System: Alverno's Institutional Perspective

MARY L. TENCPYR

This rather lengthy, but well organized paper raises probably all of the conceptual issues that could be raised about the philosophy of nationalizing a program like Alverno's. The extraordinary insights in the paper have no doubt been contributed to by many persons associated with Alverno College over the many years that institution's assessment system has been in effect.

It is well th[illegible] the paper was summarized in Figure 1, since a casual reader could be confused by the lengthy passages that follow.

The ten well formulated principles of assessment given on pages 2 and 3 are the heart of the paper and deserve the consideration of any individual or group interested in educational reform. The learned principles appearing on page 7 also deserve consideration. Although I feel it necessary to criticize some of them, I still believe that the principles can form a basis for further thinking about national assessment.

First, the author has been negligent in defining the terms "knowledge" and "ability." There is inordinate difficulty separating the two, and although the author stresses the importance of assessment within a particular learning context, the ways in which knowledge and ability interact need to be explored. Some investigators do not always differentiate between knowledge and ability within a discipline. For example, it is clear that one cannot do critical thinking in mathematics unless one has the requisite knowledge of mathematics.

It would be appropriate to offer evidence in any assessment system that development actions prescribed actually improve criterion performance; this type of content is noticeably lacking in this whole paper.

Some proof that students transfer abilities should be offered. Again, we have an assertion not supported by research evidence.

When it is said that one can validate an ability-based performance assessment process, some research evidence should be presented.

A dynamic assessment system, one that is constantly changing, cannot serve as a model for a national assessment system. There must be some stable basis for evaluating gains, not only on an individual, but also on an institutional basis. The author should be definitive about how she would move from a dynamic system to a more stable one.

There are some severe questions about whether an assessment system tied to the mission of a particular institution can be expanded into a national system. Alverno's program might be extended and amplified so that it can serve as a basic model for liberal arts institutions, but how will it serve full universities or more technically oriented institutions like California Institute of Technology?

I believe that the term "psychometric theory" to too loosely used in this paper. According to the traditional definition of psychometrics, it is difficult to see how feedback fits in.

I question how national values are to be judged. Many educational reform initiatives have been informed by obtaining the opinions of leaders in many fields. Often these opinions are supported only by anecdotal evidence. However well meaning these persons have been, they have not been able to present the type of data needed for a sound national assessment system. There are research bases from which to draw, and these present some results that may not support the Alverno model. It is true that the typical manager assessment model used in business usually deals with the types of abilities presumably measured in the Alverno model. However, as widely used as assessment centers are in business, there is still a need for more definitive research on these assessment processes. For example, there have been numerous findings of "exercise" factors. In other words, when ratings on assessment center performance are factor analyzed, the resultant factors represent performance on specific exercises, not the cross-exercise abilities or other constructs that the total assessment center was designed to measure.

In any assessment center it is essential that the exercises measure what they were designed to measure, and in contemplating national assessment, I believe it is essential that any system offered as a model be subjected to thorough research to ascertain what it is measuring.

It is suggested that one model may not fit all types of post-college career. The Alverno model appears to reflect only one model, the all-American stereotype. Yet the research literature is clear that scientists and engineers have personality (including ability) patterns that markedly depart from this stereotype. The creative geniuses of the world have not met the requirements of this stereotype. Furthermore, different communication styles are appropriate in different careers, and in business, accepted communications patterns may differ from organization to organization.

There are real problems with the research support for the Alverno model. Foremost is the problem that the long list of publications on the Alverno system includes few research publications in refereed journals. Furthermore, there are new terms introduced that have no referents in the professional literature. The term "conceptual validity" is an example.

The literature on the degree of stability of abilities and personality over time should have been acknowledged. For example, the 1989 issue of Journal of Personality that indicated the stability of personality should have been mentioned.

The literature on judgment of the characteristics of others and oneself should have been cited.

Finally, there is really not enough in the way of consideration of the practical difficulties in setting up on a national basis a system similar to Alverno's.

Summary, in terms of the evaluation criteria for these papers, I hold the following opinions:

a) The writing is well organized, but not so concise as it should be.

b) Much of the reasoning is sound, but sometimes suggestions are based on misconceptions.

c) The paper is not complete as it does not appear to acknowledge the practical difficulties of going from the Alverno model to a national system.

d) The supporting documentation in research is weak.

Response to Reviewers of DESIGNING A NATIONAL ASSESSMENT SYSTEM: ALVERNO'S INSTITUTIONAL PERSPECTIVE

Georgine Loacker

Alverno College
Milwaukee, Wisconsin

The degree of thought and attention that the reviewers of my paper clearly demonstrated deserves a response of like kind. I am grateful to them for further provoking my own thought. They tempt me to the luxury of an extended point-by-point dialogue. However, given the parameters of this particular context, I will limit myself to some questions that seem to be primary.

Elinor Greenberg enumerates the issues that constitute the feasibility argument. Margaret Miller expands them and Mary Tenopyr suggests that I should have considered them in the paper. Perhaps I did not sufficiently call attention to them, but it has been our experience at Alverno that to dwell on barriers to the feasibility of something that had not been tried would probably have kept us from ever trying it.

Despite Greenberg's noting of the difficulties of national implementation of a "development-based assessment process," she spends considerable space making some specific suggestions of how it might be done. Her review is a worthwhile supplement to the paper.

Miller's thoughtful review points out that "It seems to me that any national assessment system should preferably build on the good assessment being done on some campuses now, or at least not damage or supplant those assessment programs." I strongly support that observation and can easily see it as congruent with the recommendations I make in the paper.

Miller's position leads me to discuss issues of generalizability and transfer that I think underlie each of the reviews: To what extent can a single institution's assessment system generate information which serves that institution's purposes and still contributes to creating a national picture of how college students as a whole are meeting current and future goals and standards? This seems to be one of the conundrums we face, because it asks us to deal with preserving the diversity in our higher education system and, at the same time, contribute to and build on its coherence. One cannot assume that any specific set of practices at any one institution are generalizable to other contexts, including a national assessment system. However, it seems to me that the issue is not one of whether Alverno practices created in context would generalize to other campuses. Rather, the issue is whether Alverno and other institutions that practice effective assessment can contribute:

(1) A picture of how their students are doing that relates to the national goals. As I was writing my paper, I asked myself if Alverno's assessment systems could generate the kind of information that would be useful to creating this picture. I came to the conclusion that we could, it pressed, do so, and that it was part of our responsibility to make this kind of contribution to the national effort. As a test, are readers of this and my paper's companion piece *(Designing a National Assessment System: Assessing Abilities That Connect Education and Work* by Marcia Mentkowski), convinced by the evidence we report that Alverno students are meeting the expectations of the National Goals?

Would our trustees and employers of our graduates be convinced? If so, we could contribute, to a national assessment system, one kind of evidence for critical thinking, effective communication, and

problem solving in college students. Creating an institutional picture would make sense if a design for a national system could identify some strategies for enabling institutions to contribute to a national picture of how students are doing.

(2) Some illumination of whether any general principles yield insights that inform how a national assessment system should operate. The point of my paper was not to promote our particular practices at other campus, but to abstract *principles* that have proved effective in assisting students to learn. Such principles are likely to be informative, useful, and potentially shared. Such principles, which are virtually context-free, may be translated into a great variety of specific practices in specific contexts. Our experience suggests that institutions can learn from each other by sharing principles and then developing their own contextually valid systems. A good related example is that of writing across the curriculum. The widespread acceptance of the principles involved is apparent. Any possible arguments that they cannot be translated into some context disappear (no matter how difficult it might be to do so) in the face of their sound theoretical basis.

Perhaps the most important transfer issue, is not whether any practice of one campus can be replicated at another, but whether its graduate can transfer their abilities to life after college, and whether we can generate some performance-based evidence that the transfer is taking place (See Mentkowski's paper).

Tenopyr's perceptive observation that my paper did not deal with defining *abilities* and defining *knowledge* in depth is on target. I acknowledge it. Actually, defining *abilities* I left to my paper's companion piece (Marcia Mentkowski's paper). Defining *knowledge* has had advocates in the past (e.g., Bloom, social constructionists who develop the idea that knowledge is constructed), and those are important sources, as are the definitions and sequencing of content in the disciplines. But I had to set numerous priorities for the issues in this paper, and the more important issue that we needed to consider was how to integrate knowledge and abilities in practice. Appendix C gives some examples of outcomes as they appear in particular disciplines where knowledge and ability are reflected, both in the same statement. Given the difficulty of achieving this integration and its consequent issues (assess abilities generically or in the context of a discipline?), Tenopyr did well to focus on this issue.

In response to Tenopyr's argument that more research is needed to explore both the meaning of the abilities assessed and ways of assessing them, I concur that a national assessment system needs to have a research agenda, and a well-focused plan to assure critical review of the results and findings before they are used to inform the system. I might argue, however, that in our experience, the audience for such research is first of all *not* readers of professional journals in particular disciplines, but practitioners, persons to whom the system is directly accountable, and those who are working to improve their own campus-based systems. Further, it is our experience that interim in-house publications or reports to funding agencies, or papers presented at conferences are the most effective mode of dissemination. Further, initial results of longitudinal research are not published in professional journals until the final data point is analyzed, because one must demonstrate the validity of the scoring across time. Consequently, since we are just analyzing the final data from the fourth data point at this time, it would be premature to have published the findings in professional journals. Be assured, however, that our "university press," *Alverno Productions*, commissions reviews, and the modes of dissemination we use call for extensive external critique not only of the evidence but also its usefulness to other practitioners.