1-1-2000

Measuring assessment and testing cultures at a Reseach I University in the College of Engineering

Ann Kellogg Dieterich
Iowa State University

Measuring assessment and testing cultures at a Research I University in the College of Engineering

by

Ann Kellogg Dieterich

A thesis submitted to the graduate faculty

in partial fulfillment of the requirements for the degree of

MASTER OF SCIENCE

Major: Education (Research and Evaluation)

Major Professor: Mary E. Huba

Iowa State University

Ames, Iowa

2000

Graduate College
Iowa State University

This is to certify that the Master's thesis of

Ann Kellogg Dieterich

has met the thesis requirements of Iowa State University

Signatures have been redacted for privacy

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ACKOWLEDGEMENTS

For their patience and unfailing support while I pursued this degree, I acknowledge and appreciate my husband, Dennis, my children, Ben and Christy, and my parents, Bob and Lilla Kellogg. In more ways than I can count, all of you made it possible. At least half of this degree is yours.

For her guidance, mentoring, and magnificent editing ability, I acknowledge and appreciate Dr. Mary Huba. You guided me professionally with kindness and persistence. I hope I can continue to grow in the path you set for me.

For listening to me thrash about with new ideas and leading me to new insights about education and myself, I acknowledge and appreciate Professor Roland Jenison. I will miss our conversations. Your faith in me kept me moving toward my goals.

For extending yourself to give a neophyte the benefit of your experience, I acknowledge and appreciate the rest of my committee Dr. Barbara Licklider and Dr. Howard Shapiro.

For the continued interest you had in my work, for the ear you lent me when I needed it, and the base you provided me when I needed a break, I acknowledge and appreciate my extended family, ISU engineering faculty, and friends on and off campus.

I hope I can return to all of you the support with which I have been blessed.

CHAPTER I.

INTRODUCTION

Background for the Study

This is a time of change in most areas that touch our lives. Daily, the media announce the discovery of new scientific phenomena that force us to change our understanding of previously held theories. Periodically, politicians proclaim that the social fabric of our country is changing and that new policies must be implemented to accommodate the change. Frequently, the health-care industry reveals new treatments that change many patients' lives. It seems that nothing is constant except change itself; therefore, it should come as no surprise to find that education must change as well.

Changes in education are being proposed by industry, educational leaders, and even students. These groups have common motives about *what* education should address (e.g., curriculum content) (VanTassel-Baska, 1998) and *how* education should be addressed (e.g., the process of learning) (Black & Wiliam, 1998). They have common motives: 1) to increase accountability (Black & Wiliam, 1998; Ewell, 1991; Glaser & Silver, 1994; Kean, 1995) and 2) to improve student learning (Ewell, 1991; Tobias, 1990). Their proposals for change are largely supported by an evolving body of knowledge about the nature of learning that has been developed through cognitive psychology and educational research.

Specifically, the conversation about educational change is being guided by the evolution of psychometricians', educators', and psychologists' understanding of the nature of learning, which has led them from the "behavioral" model of the 1950s to an emerging "constructivist" model (Ewell, 1991; Glaser & Silver, 1994; Kim, McLean, & Iran-Nejad, 1996; Shepard, 1991). In the behavioral model of education, information is delivered to students, it is assumed that students absorb this information, and subsequent to instruction and absorption, the students are tested to see how much information they retain. Within this context, the teacher's job is to deliver information, and the role of the learner is to absorb the information. Assessment of student learning within the behavioral educational model involves testing how much delivered information the student retains. As stated by Shepard (1991), ". . . it is not considered possible in this low inference [testing of delivered information] system [for students] to function well on the test and not to have fully mastered the intended skills and concepts" (p. 7). The assumption underlying this practice is that passing the test proves student mastery of the content.

On the other hand, within the constructivist model of education, a quite different view of learning emerges. Learning is viewed not as a passive process which takes place through information absorption; instead, it is an active process through which the student "builds" knowledge. Shepard (1991) describes the process of learning: ". . . [Learning] require[s] reorganizing and restructuring as one learns, a more organic conception [than the behavioral model]" (p. 7). Also, instead of understanding learning as a linear, step-by-step process, the constructivist views "knowledge acquisition . . . [as] using semantic networks that show connections in many directions" (Shepard, 1991, p. 7).

As should be expected, the constructivist model of learning also includes a different view of the role of assessment in the educational process. It is an enlarged view requiring "... ways that capture the maximum capacity of students to perform" (VanTassel-Baska, 1998, p. 762). Proponents advocate "authentic assessment" which requires students to demonstrate the knowledge or skills they learned by using them in activities that "replicate the ways in which a person's knowledge and abilities are 'tested' in real world situations" (Wiggins, 1998, p. 22). This type of assessment ". . . blur[s] the distinction between testing and learning" (Airasian in Kleinsasser, 1995, p. 207). Authentic assessment is viewed as promoting as well as monitoring student learning (Huba & Freed, 2000), and it can be implemented throughout all levels of education.

How aware of the constructivist model of learning are faculty members in higher education? Much of the initial discussion of constructivism in education appeared in the educational literature at the elementary and secondary education level, but there is evidence that higher education increasingly finds merit in viewing student learning from this more complex, constructivist perspective. Johnson, Johnson and Smith recorded this trend in 1991: "College teaching is changing. We are dropping the old paradigm of teaching and adopting a new paradigm based on theory and research that has clear application to instruction" (p. 1:6). Further, Huba and Freed (2000) point out that many of higher education's "prominent leaders and theorists" have endorsed the shift to a learning-centered paradigm. As evidence that this paradigm shift has reached higher education faculty members, Campbell and Smith (1997) gathered current examples of college instructors' constructivist pedagogical methods, including: student management teams, use of academic controversy, developing community in the classroom, and using knowledge maps.

The fact that engineering education is participating in the movement toward constructivist education is revealed in the engineering education literature. Between 1995 and 1999, this literature increasingly presented both pedagogical and assessment methods which are grounded in the constructivist view of student learning as examples of exemplary teaching. These include: a) advocating cooperative learning (Lautenbacher, Campbell, Sorrows, & Mahling, 1997), b) connecting engineering concepts to real world problems (Kumar & Jalkio, 1999), c) encouraging qualitative assessment measures, and d) monitoring projects for formative assessment (Burton & White, 1999).

One motivation for engineering education to embrace the constructivist paradigm is the accountability that engineering educators have to their stakeholders to provide a quality education for their students (Holyer, 1999; Smith & Waller, 1997). The Accreditation Board for Engineering and Technology (ABET) is one such important stakeholder, and although ABET does not specify the use of constructivist methods of teaching and assessment, it has encouraged this approach. For example, ABET mandates that outcomes of student learning serve as the targets for evaluating an engineering program's educational effectiveness (American Society for Engineering Education, 1998). In other words, ABET now requires that its accredited educational programs focus on what students *learn* as opposed to their previous focus on "inputs" or what was *taught.* Learning outcomes then form the foundation on which a constructivist or learner-centered engineering educational system is centered (Huba & Freed, 2000); their achievement by students becomes the focus of accountability.

This is where the role of assessment asserts its significance. Without assessment, achievement of learning outcomes cannot be determined for either the individual student or the educational program. The engineering educational community understands this at the program level and acknowledges that "assessment of engineering education is a significant concern" (ASEE, 1998, p. 17) in terms of meeting accreditation standards. But assessment is also a concern for engineering faculty members at the classroom level since assessment "in a learner-centered paradigm is also an integral part of teaching . . . [and] through assessment, we not only monitor learning, but we also promote learning" (Huba & Freed, 2000, p. 8).

To understand how assessment influences learning in any one engineering education institution, an examination of the assessment practices within that institution must happen. This examination can provide insight into the "state of the alignment of instruction and assessment" (Haydel, 1997, p. 1) within the institution and can indicate if instruction is taking place within the teaching-centered / behavioral paradigm or the learning-centered / constructivist paradigm. In other words, assessment can be a weather vane regarding the state of faculty adoption and institutionalization of an educational paradigm.

Need for the Study

The Iowa State University College of Engineering is undergoing the accreditation process, and thus the faculty and administration are concerned about student learning within the college. An examination of the status of assessment in the college can be useful in identifying current educational practices as well as the prevailing instructional paradigm. This will reflect the degree to which faculty members are engaging in student-centered teaching which is so important to the institution's stakeholders.

This study revealed information about the College of Engineering's " evaluation culture." Assessed were the faculty members' adherence to a testing culture (a teaching-centered approach based in behavioral psychology) or an assessment culture (a learning-centered approach based in constructivist psychology). The instrument developed in the study will allow the college to assess for change in the evaluation culture of the faculty in the future.

Statement of the Problem

Three problems were addressed in the study: (a) developing a valid tool for measuring evaluation cultures in engineering courses, (b) measuring evaluation cultures of the engineering courses in the College of Engineering, and (c) revealing the relationships between engineering faculty characteristics and evaluation cultures within the College of Engineering. To address these problems, the following research questions guided the study:

1. Are the evaluation culture scores (assessment culture scores and testing culture scores) derived from the survey reliable and valid measures of the underlying testing culture or assessment culture in the engineering education setting?
2. How do engineering faculty members rate on a scale measuring the testing culture in their classrooms?
3. How do engineering faculty members rate on a scale measuring the assessment culture in their classrooms?

4. What is the relationship between testing culture scores and assessment culture scores?
5. Do engineering departments differ in terms of faculty members' (a) testing culture scores and (b) assessment culture scores?
6. Is there a relationship between the faculty members' opportunity to learn about student assessment (i.e., the number of college courses in classroom / student assessment taken for college credit, the number of semesters of participation in Project LEA/RN, and the number of faculty development sessions on the topic of classroom / student assessment other than Project LEA/RN) and their (a) testing culture scores and (b) assessment culture scores?
7. Is there a relationship between the number of years since the Baccalaureate degree / the number of years the engineering faculty spent in higher education and faculty members' (a) testing culture scores and (b) assessment culture scores?

Description of Variables

Dependent Variables

The first dependent variable in this study is engineering faculty members' testing culture scores. The testing culture score "measures the frequency with which . . . [faculty members'] classroom assessment practices and behavior align with those of a testing [culture]" (Haydel, 1997, p. 21). Using Shepard's (1991) model of "indirect means to study implicit theories" (p. 3), the higher the "testing culture" score of a faculty member, the more consistently that faculty member uses practices and behaviors reflecting beliefs about learning grounded in behavioral theories.

The second dependent variable in the study is the engineering faculty members' assessment culture scores which "measure the frequency with which . . . [faculty members'] classroom assessment practices and behaviors align with those of an assessment [culture]" (Haydel, 1997, p. 21). Again, based on Shepard (1991), the higher the "assessment culture" score, the more consistently a faculty member uses practices and behaviors reflecting beliefs about learning grounded in constructivist theories.

Independent Variables

Following are the independent variables for this study:

1. The department in which the faculty member teaches.
2. The number of college courses in classroom / student assessment taken for college credit.
3. The number of semesters of participation in Project LEA/RN.
4. The number of faculty development sessions on the topic of classroom / student assessment.
5. The number of years since receiving a Baccalaureate degree.
6. The number of years teaching in higher education.

Definition of Terms

The terms used in this study are defined as follows:

1. Assessment Culture: Assessment practices and behaviors indicating an adherence to beliefs about learning grounded in constructivist theories. Examples of these practices and behaviors include a) using assessment to promote as well as evaluate learning, b)

using direct methods of assessment such as projects, performance, and portfolios, c) developing assessments based on real-life problems in an "authentic" context, d) assessing the ability to use knowledge, f) using open-ended problems, g) using standards to determine achievement, h) including assessment in the learning process, i) operating in a cooperative culture, and j) including students and others as evaluators of achievement. (Huba & Freed, 2000).

2. Behavioral theory/model of learning: The "instructional frameworks that emphasize the teaching-testing-teaching relationships based on the principles of behavioral psychology. . . . This approach promotes "teacher control, sequential learning hierarchies, and learning outcomes. . . .Learning is seen to be sequential and linear." (Kim, McLean, & Iran-Nejad, 1996, p. 32).

3. Constructivist theory/model of learning: This understanding of learning recognizes that students reorganize and restructure knowledge as they learn. Therefore, learning does not occur in a linear hierarchy but can be depicted by using "semantic networks that show connections in many directions. . . . [as] learners construct and then reconstruct mental models that organize ideas and their interrelation. . . . Learning occurs by the individual's active construction of mental schemas" (Shepard, 1991, pp. 7-8).

4. Project LEA/RN: A program developed in the College of Education at Iowa State University to addresses faculty development needs in the areas of teaching and student

learning. This program reflects an "interactive model of faculty development based on . . . recent theoretical work in adult education and effective staff development research" (Licklider, Schnelker, & Fulton, 1997, p. 121). The program is built on the framework that: a) faculty need regular sessions over an extended time to have "ample opportunity to engage in the processes of both cognitive and behavior changes," b) faculty are involved in the "planning and implementation of the program," . . . c) "development activities are the result of faculty interest in their own professional development," d) the program occurs in a "trusting and collaborative environment," [and] e) "educational developers must help participants become self-directed learners" (pp. 125-126).

5. Testing Culture: Assessment practices and behaviors indicating an adherence to beliefs about learning grounded in behavioral theories. Examples of these practices and behaviors include a) assessing for mastery of facts and concepts, b) using objectively scored tests, c) using assessments with one right answer, d) sorting students through relative grading, e) monitoring learning, f) separating assessment from teaching, g) assessing outside of the real-world context, h) operating in a competitive culture, and i) utilizing the instructor as the primary evaluator (Huba & Freed, 2000).

Research Hypothesis and Rationale

Following are the hypotheses used to address research questions 4 through 7.

Hypothesis 1

Engineering faculty will rate higher on the testing culture scale than the assessment culture scale.

Rationale:

The emergence of the learning-centered paradigm of instruction is a relatively recent occurrence. The discussion of constructivist education in the engineering educational literature began almost a decade ago, but it did not become prevalent until about the middle of the 1990s. Because a learning-centered approach is relatively new to education, it is not unreasonable to assume that few engineering faculty have actually experienced it. Instead, most faculty members have probably experienced a more traditional "behavioral" educational system, and they have taught for many years under that same system. Since it is natural to fall back on what you know, it was expected engineering faculty members would rate more highly on the testing culture scale.

Hypothesis 2

There will be a negative relationship between the testing culture scores and the assessment culture scores.

Rationale:

Since the testing culture scores and the assessment culture scores have underlying constructs that are in opposition, it is reasonable that the two evaluation culture scores would also be in opposition. In other words, the more strongly a faculty member identifies with and applies the practices associated with one educational paradigm, the less that faculty member is likely to be engaged with the other paradigm. Therefore, as the scores go up in one evaluative culture scale, the scores go down in the other.

Hypothesis 3

Engineering departments will differ in terms of their (a) testing culture scores and (b) assessment culture scores.

Rationale:

The prediction for this hypothesis emerged from the observation that different engineering departments have addressed the accreditation process and its associated changes differently. These differences may be at least partially due to the differences in their educational paradigms, which would be reflected in their evaluation culture orientation and scores.

Hypothesis 4

There will be a positive relationship between assessment culture scores and the faculty members' opportunity to learn about the topic of classroom assessment: (a) the number of college courses in classroom / student assessment taken for college credit, (b) the number of semesters of participation in Project LEA/RN, and (c) the number of faculty development sessions on the topic of classroom / student assessment other than Project LEA/RN. A negative relationship will occur between the testing culture scores and the same educational experiences.

Rationale:

Faculty development in various forms has been institutionalized as a major tool for implementing change in the education setting. It is assumed that these development efforts make a difference, and the more exposure a faculty member has to faculty development sessions, the more apt that faculty member is to engage in proposed change. Since the change proposed was to shift to a learning-centered educational paradigm, as demonstrated by Project LEA/RN and other College sponsored faculty development sessions, it was projected a higher assessment culture score would be the result. For reasons discussed in Hypothesis 2, it was then reasonable to expect a lower testing culture score with more exposure to faculty development.

Hypothesis 5

There will be a positive relationship between the assessment culture scores and (a) the number of years since receiving a Baccalaureate degree, and (b) the number of years teaching in higher education. There will be a negative relationship between the testing culture scores and the same faculty characteristics.

Rationale:

The rationale for this hypothesis was based on research about faculty characteristics in which Fulton and Trow (as cited in Austin & Gamson, 1983) reported professors' interests shift from research to teaching with increasing age. Therefore, it was concluded that faculty members who have had the most time since Baccalaureate graduation and who have been teaching the most time would have the most interest in teaching. Consequently, they would be most inclined to take the time to adopt innovations associated with teaching, such as shifting educational paradigms to adopt assessment culture practices. Younger faculty or those with the fewest years of teaching would be less likely to adopt innovations or make a paradigm shift.

Assumptions of the Study

Following are the major assumptions of this study:

1. The respondents will be able to correctly interpret the questionnaires.
2. The respondents will be honest in their responses.
3. The respondents will adequately represent the faculty in the College of Engineering at Iowa State University in terms of their beliefs about teaching, learning, and assessment.
4. The questionnaire responses will accurately represent the higher education classroom culture of assessment and the classroom culture of testing.
5. Faculty development is an effective method for initiating educational change.
6. Recent faculty development within the College of Engineering, both on campus and from external sources (e.g., professional conferences), has been conceptualized within a constructivist, learning-centered paradigm.
7. Most engineering faculty are more familiar with a behavioral, testing-centered educational paradigm than with a constructivist, learning-centered educational paradigm.

Limitations of the Study

There were several limitations to the study. The first limitation is related to the correlational design of the study. In correlational research, causation cannot be determined. In this study, it could not be determined that the independent variables (department in the college, faculty development, the number years since a Baccalaureate degree, and the number of years teaching in higher education) caused the variation in assessment culture or testing culture scores. In non-experimental studies, random assignment cannot be used and this precludes determination of causation (Gall, Borg, & Gall, 1996).

Another limitation arose from the fact that participation in the study was voluntary. Because of the self-selection of faculty members to participate, it is possible that non-responding faculty members have different characteristics than responding faculty members. To minimize the possibility that this affected the study, several contacts were made with the study's population to maximize faculty response.

Finally, generalizability is another limitation to this study. Since the study was conducted in the College of Engineering on one campus, applying the results of the study to other settings in higher education should be done with caution. When considering whether the results may be generalized to another institution, consideration should be given to how similar the college is to Iowa State University's college in terms of size, cultural history, demographics, educational history, etc.

Significance of the Study

One purpose of this study was to develop a valid instrument to be used to determine the evaluation culture (assessment or testing) in an engineering education setting. Future use of the instrument will benefit the college by providing a set of data easily compared to this study's data, thereby facilitating a study of faculty change. Since ABET (2000) accreditation requires engineering programs to monitor educational improvement, including student learning and faculty development, a longitudinal study of this type seems beneficial (ASEE, 1998). This study serves as the baseline for such a longitudinal study.

This study contributed to Iowa State University's College of Engineering by providing information about the evaluation culture held by the faculty members in the college. This information can now be used by the faculty and administration within the college as they consider faculty members' needs for additional faculty development in the areas of teaching and assessment. Understanding the "mind models" of faculty members (Senge, 1990) will allow faculty development to address the disparity between faculty views of student / classroom assessment and the approach desired by the college and ABET.

Also, this study provided insight into factors related to faculty demographics and faculty development to determine what relationship, if any, those had with the evaluative culture scores. Faculty and administration in the college can now use that information to help them predict what strategies may help faculty members to meet current educational practice expectations. Also, they can more effectively commit the resources that would be most beneficial to the faculty for their improvement.

Finally, this study provided an instrument for other researchers and professionals in the field of higher education to use in conducting further studies in the educational culture in their settings.

CHAPTER II.

REVIEW OF RELATED LITERATURE

Introduction

A review of the current literature revealed that few, if any, studies have been done to identify the evaluation culture in higher education. Most of the published literature in this area focuses on specific assessment methods but does not examine evaluation in the larger context of culture. This chapter will present theories that address evaluation in the context of culture, and it will also present theories of change as they impact faculty development and faculty members' abilities to adopt to new evaluation cultures. First, general educational theories of learning, both behavioral and constructivist, and the role evaluation plays within those theories will be considered. Next, Haydel's study (1997) which provided the foundation for the development of this study's survey will be examined. Finally, a discussion of the literature associated with change and faculty development will be presented.

Educational Paradigms

Much has been written about educational paradigms over the last 20 years, especially as research in cognitive psychology and education have prompted new ways of thinking about student learning (Bergquist & Phillips, 1977; Caine & Caine, 1997; Huba & Freed, 2000). It is helpful to discuss educational theories in terms of paradigms since a paradigm frames our "deeply held beliefs and ideas that shape our grasp of reality" (Caine & Caine, 1977, p. 12), establishes boundaries around those beliefs and ideas, and sets the rules within which those ideas and beliefs operate (Huba & Freed, 2000). In the process of educational change, modifications can be considered in terms of what part of the paradigm is being stretched. For instance, if research produces a new understanding of neuronal development that in turn affects education professionals' beliefs about memorization, at some point the old boundary framing "teaching and learning" cannot be stretched far enough to accommodate the new beliefs. A paradigm "shift" must occur (Berquist & Phillips, 1977). This shift to the new paradigm not only changes the boundaries but also allows new rules to operate within those boundaries, creating new educational methods. It is possible for two different paradigms to coexist, each one embracing a different theory of learning and different rules of teaching. Currently education seems situated at the junction of two paradigms (Caine & Caine, 1997; Huba & Freed, 2000).

The educational paradigms operating in our current system of education have been well documented. The traditional paradigm is "grounded in the learning theory of behaviorism" (Shepard, 1991, p. 6) and is described as "mechanistic" (Caine & Caine, 1997, p. 12) or "teacher-centered" (Huba & Freed, 2000. p.2). Its boundaries encompass the beliefs that teachers deliver information, students are passive receivers of information, and evaluating student learning is carried out to monitor what students have learned (Huba & Freed, 2000; Kleinsasser, 1995; Shepard, 1991). Another more recent educational paradigm emerged from the field of cognitive psychology in the last 20 years, and it is identified as constructivist (Shepard, 1991). This is "a more organic conception" that recognizes that "learning requires reorganizing and restructuring as one learns" (p. 7).

This paradigm's constructs have become more refined over the last ten years; it has evolved into a "learner-centered" educational paradigm (Huba & Freed, 2000) whose boundaries include the beliefs that teachers are facilitators whose role is to guide students to process, analyze and examine educational experiences for meaning (Caine & Caine, 1997). Also, students are active participants in their learning, and student evaluation is used to "promote and diagnose learning" (Huba & Freed, 2000, p. 5).

Conceptions of Assessment and Evaluation

Because the behavioral and constructivist paradigms have such different belief systems, it follows that the rules that govern how educational practices operate in each system are also quite different. Kleinsasser (1995) investigated these rules as they relate to evaluation practices and grouped the rules into four categories: (a) roles of teachers, (b) roles of students, (c) rules of testing, and (d) rituals of testing. By examining the educational rules within each category, she then labeled the resulting evaluation culture as either a "testing culture" which aligns with the behavioral paradigm or an "assessment culture" which aligns with the constructivist paradigm.

The testing culture, as described by Kleinsasser (1995), is characterized by the following educational practices. First, the teacher's role is to control the test: The teacher keeps the test uncontaminated, determines what is on the test, and scores the test. Conversely, the student's role in the testing culture is one of "powerlessness, mystified by the testing process" (p. 206). Next, the "rules of testing" dictate that students must take the test alone, the test is a paper-pencil test, and the teacher is the "sole audience . . . [who]

owns the test" (p.206). Finally, "rituals of testing" in the testing culture require that testing and learning are separate and that learning is done for the test.

The assessment culture, as defined by Kleinsasser (1995) operates with different rules within each category. First the role of the teacher is to collaborate with the student: The "teacher and learner are in a conversation about learning" (p. 206), and the student and teacher build assessments together. The student's role is more active than in the testing culture since the student not only builds assessments with the teacher but also is a "self appraiser, [and] co-appraiser with teachers and other assessors" (p. 206). The rules of testing within the assessment culture include (a) students team or cooperate with each other, (b) assessments other than paper-pencil tests are used such as demonstrations and exhibitions, and (c) assessments are presented to an audience other than the teacher such as self, peers, and parents. Finally, rituals in the assessment culture are different from the testing culture in that "distinctions between learning and assessment are blurred" (p. 206), and the assessments are "demystified; student knowledge about assessments and the assessment process increases" (p. 206).

The evaluation culture classification developed by Kleinsasser was modified by Haydel in 1997. Although Haydel retained the evaluation domains of assessment and evaluation, she reorganized and expanded the evaluation practices within those domains. These reconceptualized evaluation practices were then labeled "dimensions of evaluation cultures" and included (a) "purpose, . . . [which] asks the question, 'Why assess?'" (b) "target, . . . [which] asks the question, 'What is assessed?'" (c) "method, . . . [which] asks the question, 'How do we assess?'" (d) "context, . . . [which] asks, 'What are the conditions under which assessment occurs?'" and (e) "communication of student learning.

. . . [which] asks the question 'How are assessment results communicated?'" (Haydel, 1997, pp. 7-8). Haydel's dimensions reflect important assessment issues discussed by many authors (e.g.; Black & Wiliam, 1998; Cain & Cain, 1997; Erwin, T. D., 1991; Ewell, 1991; Haydel, 1997; Huba & Freed, 2000; Kleinsasser, 1995; Shepard, 1991; Wiggins, 1998).

In the testing culture domain, there are several *purposes* for evaluation: (a) to make judgments about student performance, (b) to monitor learning, (c) to assign grades, and (d) to rank students. Conversely, the assessment culture domain uses evaluation for other purposes, including (a) to provide feedback to students about their learning, (b) to aid the instructor in planning further instruction, and (c) to monitor student progress in relation to standards.

The *target* of evaluation, or "what is assessed" in the testing culture, largely focuses on (a) the mastery of knowledge, (b) single discipline content, (c) instructor priorities, and (d) content that is decontextualized. In the assessment culture, the target is likely to (a) include the demonstration of the use of knowledge to solve problems, (b) involve affective characteristics of students, such as appreciation, (c) have interdisciplinary components, (d) have real-world applications, and (e) have clearly defined performance standards.

How evaluation is carried out, or the *method* of evaluation, in a testing culture is usually indirect and uses (a) selected response, or questions with one-right answer, (b) objectively scored tests, and (c) paper-pencil instruments for the majority of the assessment. When assessment culture evaluations are conducted, they rely largely on direct methods, using (a) student performance of a skill or project, (b) interviews or personal communication, (c) observation of students, (d) open-ended questions with more than one-right answer, and (e) portfolios and papers representative of student work.

The *context* of evaluation occurs under these conditions in the testing culture: (a) assessment is separate from teaching and learning, (b) it signals the end of a topic or unit, (c) each student is "on his/her own," and (d) the instructor is the sole beneficiary of the evaluation. In an assessment culture, (a) teaching and assessment are intertwined, (b) collaboration among students and between instructor and students may occur in the development or the completion of the evaluation, and (c) audiences other than the instructor may benefit from the evaluation.

Finally, the *communication* of student learning is also different for the testing and assessment cultures. In a testing culture, the dimension of *communication* of student learning focuses on (a) instructor control over what is communicated and how the communication occurs with the communication usually "emanating from them [instructors] and going to the students" (Haydel, 1997, p. 6), and (b) grades to indicate student performance, often posted or given on report cards with no chance for feedback. In an assessment culture, communicating student learning is expanded to include: (a) discussion between instructor and students about student performance, the level of achievement, and quality of work, (b) instructor responses to students' needs for improvement of performance, and (c) conferences with students or narrative reports to students.

As reflected in this list of criteria for the two evaluation cultures or domains, the assessment culture and testing culture spring from different belief systems. Therefore, it is expected that differences in these paradigms are bound to result in different evaluation practices.

Measuring the Assessment Culture and Testing Culture

The idea that one component of a paradigm, such as evaluation, can be used to make inferences about the nature of that paradigm has been used in previous research. In Shepard's study (1997) about psychometricians' beliefs about learning, she recognized that "teachers' classroom practices can be understood in terms of their beliefs or implicit theories about instruction and learning" (p. 2). Haydel (1997) restated that premise in her study of the culture of evaluation in the secondary education classroom: "Methods of assessment are determined by beliefs about learning" (p. 1). In both cases, the researchers focused their investigation on evaluation practices as the indicator of which educational paradigm the educators embraced since (a) those practices were observable and measurable, and (b) the evaluation practices could be classified into discrete paradigms, as discussed earlier.

Further, Haydel (1997) proposed that the consistency with which an instructor adheres to the rules of evaluation for an educational paradigm indicates how strongly the instructor adheres to the beliefs of that educational paradigm. This was based on the recognition that assessment plays a central role in the educational reform movement and "in schools across the nation where assessment and evaluation processes are being brought into congruence with changed classroom practice, a new culture of assessment must replace the old culture of testing" (p. 2). Subsequently, Haydel developed an instrument to measure evaluative culture in secondary education and to indicate teachers' placement within the educational paradigms.

Haydel's (1997) survey items described practices of classroom evaluation representing each dimension (purpose, target, method, context, and communication) for each domain. This survey was then administered to secondary education teachers. Results suggested that evaluative culture is multi-dimensional and can be "described in terms of scores on both the testing and the assessment scales" (p. 21). Haydel concluded that the testing culture scale and the assessment culture scales were construct valid as well as reliable. Therefore, the evaluative culture in an educational system can be operationally defined, and once so defined, it can be used as the basis for further study of the dynamics in an educational system, such as the relationship between the evaluative culture and student achievement.

The Process of Change and Faculty Characteristics

"Change" means different things to different people. It represents excitement for some, it is uncomfortable for many, and it is unmanageable for a few (Caine & Caine, 1997; Gray, 1997). Rogers (1995), in his work with educators and educational innovation, and Geoghegan (as cited in Gray, 1997) classified people in terms of their engagement with change using five categories: (a) innovators, (b) early adopters, (c) early majority, (d) late majority, and (e) laggards. These categories are based on the group members' receptivity toward innovation.

In this classification system, the innovators are the "first few people in a group to use an innovation . . . [and they] make up the first 2.5 % of the population" (p. 11). The early adopters consist of the "visionaries and risk takers, who do not fear failure . . . [who] make up the next 13.5 % of the population" (p. 11). Next, people who make up the following 34 % of the population are the early majority, and they characteristically "may be comfortable with the idea of innovation, but . . . are pragmatists, who are less willing to take risks" (p. 11). As their label suggests, the late majority can be convinced to use an innovation, but only after "it has become well established among the majority. . . . [They] have a high aversion to risk" (p. 12) and make up the next 34 % of the population. Finally come the laggards: who compose the last 16 % of the population and "most likely will never adopt the innovation . . . and may object to the innovations' use by others" (p. 12).

Rogers' classification system is a useful tool when leaders are promoting an innovation. The classification distinctions allow leaders to tailor efforts to motivate and meet the needs of each group (Gray, 1997).

Additional classification of faculty was conducted by Murray (as cited in Bergquist & Phillips, 1977). He identified faculty groups according to their motivational needs and labeled the groups: (a) "high-need-achievers", (b) "low-risk" takers and (c) "low-need-achievers" (p. 17). Examining the characteristics within each of the groups reveals a connection between the two classification systems of Rogers and Murray.

Murray's (as cited in Bergquist & Phillips, 1977) high-need-achievers group is similar to Rogers' (1995) innovators and early adopters. The high-need-achievers (a) desire to take on challenging tasks with some risk, (b) have a "realistic level of aspiration," and (c) maintain the ability to remain "persistent following initial failure" (p.17). Also, they are motivated by situations where (a) they consider *themselves* "responsible for the success or failure of an activity" and (b) they have "explicit knowledge of the results so [they] could determine when [they] succeeded" (p. 17). These characteristics are compatible with Rogers' innovators and early adopters and the characteristic of "risk taking" is explicitly common to both Murray's and Rogers' groups. Also it is apparent that the intrinsic motivation identified in the Rogers' innovators and early adoptors (Gray, 1997) can also be implied in Murray's high-need-achievers since it is the intrinsic motivation in the high-need-achievers that drives them to achieve.

A similar comparison can be made between the next two levels of Roger's population, the early majority and the late majority and Murray's (as cited in Bergquist & Phillips, 1977) "low-risk" faculty members. Both Rogers (1995) and Murray cite the over-riding faculty characteristic in these groups as a decreasing tolerance for risk-taking. Accompanying this risk aversion is also a requirement for proof of an innovation's effectiveness before the adoption of an innovation by the group: This population "want[s] proven practices that build on current processes . . . they look to central administrators, and local opinion leaders . . . for guidance and confirmation of the innovation's worth, [and] . . . they need more support" (Gray, 1997, p. 13).

Finally, reluctance to take a risk by adopting an innovation is increased even more for both Rogers' (1995) laggards and Murray's (as cited in Bergquist & Phillips, 1977) "low-need-achievers" (p.17). Bergquist & Phillips (1997) emphasized that to reach the low-need-achievers, an innovation's "meaning and impact on their teaching" must be demonstrated and the tasks presented must be easily be performed" (p. 14) in order to motivate adoption.

Rogers and Murray discussed faculty members' openness to change in terms of their personality characteristics. Caffarella and Zinn (1999) related openness to change to the demographic characteristics of faculty rank and tenure status. They suggest that it is probably the senior and tenured faculty who are most likely to initiate activities involving risk (e.g., innovation). Since these senior and tenured faculty are less likely to be injured if the innovation is a failure than is the junior or non-tenured faculty member, it appears that they are more apt to attempt it than their less seasoned peers.

Faculty Development

As mentioned previously, faculty developers facilitating the change process can benefit from understanding the different characteristics of faculty groups and their motivating forces. These benefits will be discussed in terms of the classifications of Murray and Rogers.

Faculty members who are innovators/early adopters/high-need-achievers are intrinsically motivated and have little fear of risk-taking. Little effort on the part of leadership is needed to facilitate change with this group. If leadership can give them access to ideas, the space to try the ideas, and recognition for their accomplishments, faculty development will occur with little external direction and effort (Bergquist & Phillips, 1977). They will take full advantage of faculty development programs that are "guided by the assumptions that faculty are autonomous, . . . seek to encourage faculty self-management and faculty decision-making about their learning, and . . . serve as a resource for extended learning beyond institutional pursuits (Fulton, Licklider, & Schnelker, 1997, p. 18).

On the other hand, lower risk tolerant faculty members, such as the early majority, late majority, and low-risk groups, require more of the traditionally discussed supports from administration to benefit from faculty development (Bergquist & Phillips, 1977; Caffarella & Zinn 1999; Fulton, Licklider, & Schnelker, 1997; and Walvoord & Pool, 1998). For instance, since these faculty members are less intrinsically driven, they have a higher need for the external support of their colleges' interaction and approval (Walvoord & Pool, 1998) than do other faculty groups. This support is especially needed when it is context of implementing a change – as is occurring in the educational paradigm shift explained by the Wright Institute (cited in Bergquist & Phillips, 1977):

> The current academic culture seems to be supportive of faculty so long as their developmental concerns are primarily associated with their discipline and research activities. When other developmental concerns associated with teaching and student development tend to dominate, then the faculty member is likely to feel isolated from his colleagues and estranged from the values and norms of academic culture. . . . Discussions about teaching can legitimate these developmental concerns, as well as provide a means whereby they can be jointly explored and clarified. (p. 21)

Another support that is especially helpful to faculty members facing change who have low-risk-tolerance is to provide them with evidence that the educational change is effective. Once the effectiveness has been established, the change is then more palatable since they can see that the risk of failure compared to the benefits is optimized (Bergquist & Phillips, 1977).

Finally, time emerges as a factor when considering faculty development for the entire faculty population (Gray, 1999, p.13). "Organizational theorists point out that it often takes some fifteen years for a set of new values to permeate an institution"

(Freedman, et. al., 1979, p. 158). This is consistent with the understanding that the late majority and laggards want proof of innovative effectiveness before they will engage in the process of change. Development of this proof takes time. Also, if faculty members are to engage in collaboration and support each other as previously suggested, time is needed to build the relationships necessary for a sustainable collaborative environment. (Fulton, Licklider, & Schnelker, 1997).

Because of all these factors, leaders who want to maximize faculty members' adjustment to changing educational paradigms should take a long-range perspective and offer faculty development opportunities throughout the period needed for the new paradigm to permeate the institution.

CHAPTER III

METHODOLOGY

Introduction

There are two purposes of this chapter. The first is to describe the procedures used to modify the Survey of the Culture of Classroom Assessment developed by Haydel (1997) for use in the engineering education setting. The second is to describe the procedures used in this study to examine the culture of assessment in the College of Engineering at Iowa State University and the relationships, if any, that exist between the culture and faculty characteristics. The faculty characteristics examined include: (a) the departments in which the faculty members teach, (b) the number or college courses in classroom / student assessment taken for college credit, (c) the number of semesters of participation in Project LEA/RN, (d) the number of faculty development sessions on the topic of classroom / student assessment other than Project LEA/RN, (e) the number of years since receiving a Baccalaureate degree, and (f) the number of years teaching in higher education.

Research Design

This study is a descriptive, "one point in time" study (Gall, Borg, & Gall, 1996), conducted to determine the culture of classroom assessment within the College of Engineering. It was designed to be an ex post facto, correlational study, (Gall, Borg, &

Gall, 1996) so that relationships between faculty characteristics and the culture of classroom assessment could be examined.

Population

The study's target population was the collection of tenured and tenure track faculty members of the College of Engineering employed at Iowa State University during the spring of 2000. The population was limited to the tenured and tenure track faculty since they are the faculty with which the College has made a commitment, and they represent a more continuous influence on students' classroom experiences than do temporary and adjunct faculty. The list of the faculty members was obtained from the College of Engineering dean's office. Subsequently, the list of faculty members from each engineering department was reviewed with the respective department's administrative assistant so that the names of faculty members on leave or otherwise unavailable could be removed. This resulted in a population of 215 tenured and tenure-track faculty members to be surveyed.

Of the 215 surveys distributed, 132 were returned, for a response rate of 61.40%. As can be seen in column 3 Table 3.1, the distribution of survey returns was uneven among the engineering departments. This uneven return resulted in some over-representation of engineering departments, specifically AEEM (Aerospace Engineering and Engineering Mechanics) and MSE (Material Science and Engineering) in the sample of returned surveys, but as seen in columns 3 and 4 of Table 3.1, that over-representation was not extreme.

Table 3.1. Survey Response by Engineering Department

Engineering Department	Number of Surveys Sent	% Returned By Dept.	% of Sample[(1)]	% of Population[(3)]
Aerospace Engineering and Engineering Mechanics	30	70.0%	15.8%	9.8%
Agriculture and Biosystems Engineering	24	45.8%	8.3%	5.1%
Chemical Engineering	18	61.1%	8.3%	5.1%
Civil and Construction Engineering	35	67.5%	17.3%	10.7%
Electrical and Computer Engineering	46	52.2%	18.0%	11.2%
Industrial and Manufacturing Systems Engineering	19	52.6%	7.5%	4.7%
Materials Science and Engineering	18	88.9%	12.0%	7.4%
Mechanical Engineering	25	60.0%	11.3%	7.0%
Total	215		98.5%[(2)]	61.4%[(4)]

[(1)] N of sample = 132; [(2)] Total % is less than 100% due to rounding; [(3)] N of population = 215;
[(4)] Total percent as shown is = 60.0 % due to rounding; actual total percent = 61.4 %.

Table 3.2, provides additional demographic information about faculty members who returned surveys. For both the number of years since Baccalaureate degree and the number of years teaching in higher education, a wide range was reported. Also, these two demographic variables were correlated with an r = .889. Since they were so highly correlated, the distribution of years across the ranges was similar, and those distributions were fairly even. The Baccalaureate degree distribution is reported here in more detail: a) The first quartile ranged from 6 to17 years; b) The next quartile ranged from 18 to 25 years; c) The third quartile ranged from 26 to 35 years; and d) The fourth quartile ranged from 36 to 51 years.

Table 3.2. Descriptive Statistics for Faculty Characteristics [1]

	Mean	Median	Mode	Standard Deviation	Range
No. Years Since Baccalaureate Degree	26.30	25.00	20	10.63	6 – 52
No. Years Teaching in Higher Education	17.81	15.50	1	11.59	1 – 43
No. of College Courses in Classroom / Student Assessment Taken for College Credit	.42	.00	00	1.47	0 – 12
No. of Semesters Participation in Project LEA/RN	1.46	.00	.00	2.18	0 – 8
No. Faculty Development Sessions on the Topic of Classroom / Student Assessment Other Than Project LEA/RN	2.47	2.00	.00	3.20	0 –20

[1] N = 132

Additionally, there was not much difference between the engineering departments in the reported number of years since Baccalaureate Degree. The mean number of years for the departments were all within 5 years of the College mean of 26.30 years.

Finally, expanding further on the demographic data reported in Table 3.2, investigating the relationship between the number of years since Baccalaureate Degree and faculty development revealed that there was little relationship. The number of years since Baccalaureate Degree correlated with a) the number of college courses in classroom / student assessment with an $r = .003$; b) the number of semesters of participation in Project LEA/RN with an $r = .219$; and c) the number of faculty development sessions on the topic of classroom / student assessment other than Project LEA/RN with an $r = .190$. None of these were statistically significant.

Instrumentation

The survey used for this study was based on a survey developed by Haydel (1997), "Measuring the Evaluative Culture of Classrooms" in a K-12 educational environment. Haydel first identified two evaluative cultures and classified them as a "testing culture," which is associated with a behavioral view of learning, and an "assessment culture," which is associated with a constructivist view of learning. Haydel identified five dimensions "that underlie sound classroom assessment practice and form the framework of evaluative culture" (p. 7). She used these dimensions as the foundation on which to write her survey items about classroom assessment practices in both the testing and assessment cultures. The teachers responding to Haydel's survey indicated the frequency of these cultural practices in their classrooms using a Likert-type scale from "never" to "always."

Although Haydel's (1997) survey was modified to determine the evaluative culture in the College of Engineering, the same cultures, or domains, (assessment and testing) were used as the focus of investigation, as were the five dimensions on which her survey was based: "(1) purpose of assessment, (2) targets of assessment, (3) methods of assessment, (4) assessment context, and (5) communication of student learning" (p. 7). (The relationship between the domains and dimensions is listed in Table 3.3) A review of the literature revealed that, although Haydel's survey was developed for secondary education, a change in these dimensions was not indicated for the higher education context. Only some minor modifications were necessary for adaptation to higher education. For example, Haydel's original question, "I used parent-teacher conferences to communicate student learning" was changed to "I use student conferences to communicate student learning." Additional items were modified or written to reflect the engineering discipline. For instance, the item, "The skills that I expect students to demonstrate are assessed by paper and pencil tests" was changed to "I use written tests to assess my students' abilities to apply engineering concepts." The complete list of survey items, by domain and dimension, are listed in Table 3.3.

Table 3.3. Survey Items by Dimension and Domain.

Dimension 1: Purpose of Assessment (*Why Assess*)	
Survey Items for Testing Domain	Survey Items for Assessment Domain
1. For me, a primary purpose of assessment is to assign grades to students.	8. For me, a primary purpose of assessment is to improve student performance.
32. When grading an assessment, I put the class's assessment scores in order and give each student a grade based on his / her rank in the group.	15. I review assessment results to guide further instruction in the areas of student weakness.
Dimension 2: Target of Assessment (*What to Assess*)	
Survey Items for Testing Domain	Survey Items of Assessment Domain
4. My assessments focus on measuring knowledge of concepts and facts.	11. My assessments require students to use their knowledge to generate solutions to real-world engineering problems.
17. I decide what will be included on each assessment without student input.	31. I involve students in deciding what will be included on each assessment.
	5. It is important to assess students' attitudes, interests, and motivations.
	36. My assessments evaluate students' abilities to express and defend their points of view.

Table 3.3, Continued

Dimension 3: Method of Assessment *(How to Assess)*	
Survey Items for Testing Domain	Survey Items for Assessment Domain
2. My assessments have one-right answer.	9. My assessments are structured so that there can be more than one right answer.
19. I use written tests to assess my students' abilities to apply engineering concepts.	25. I assess skills or abilities directly by observing performances or evaluating products / projects that use engineering concepts.
35. The skills that I expect students to demonstrate can adequately be assessed on paper.	29. My assessments require students to demonstrate skills or make products.
	22. I provide opportunities, such as portfolios or journals, for students to engage in self-reflection about how and what they learned.
	26. As part of my assessment, I question students orally in order to evaluate the depth of their understanding.
	33. I allow students the opportunity to justify answers marked wrong.

Dimension 4: Context of Assessment *(Conditions Under Which Assessment Occurs)*	
Survey Items for Testing Domain	Survey Items for Assessment Domain
13. I prefer that students work alone when being assessed.	7. I am comfortable using group work to assess student learning.
23. In my classes, I am the only one who evaluates student learning.	28. I provide opportunities for students to assist in the evaluation of their peers' learning.
37. I assess only after instruction has occurred.	18. In my classes, assessment occurs during as well as after instruction.
16. I require students to return tests so questions can be used again.	24. I review tests or other assessments with my students as a means of helping them improve their own learning.
34. In my classroom, I encourage students to compete with one another for the best grade.	14. In my classroom, I encourage students to support each other as they strive to achieve high performance standards.

Table 3.3, Continued

Dimension 5: Communication of Student Learning	
Survey Items for Testing Domain	Survey Items for Assessment Domain
6. Communication student performance in my class is adequately accomplished by posting grades.	12. When reporting student learning in my class, I find it necessary to write comments in addition to giving grades.
30. I use grades as the sole means of communicating learning to my students.	20. I use student conferences to communicate with students about their learning.

Dimension 6: Grading *(How assessments Translate Into Grades)*	
Survey Items for Testing Domain	Survey Items for Assessment Domain
10. The number of students in my class who receive the same grade (e.g., the number receiving As, Bs, Cs, Ds, or Fs; or the number receiving passing or failing) should be limited.	3. It is possible for all my students to receive the same grade (e.g., all As, Bs, Cs, Ds, or Fs).
27. I adjust my grading scale depending on how students perform on tests.	21. I inform my students about the performance standards and grading scale to be used when assessing their achievements.

After rewording survey items and developing some new items, it was discovered that some items did not fit well with the original response format of "never to always." For example, the survey item stating "It is important to assess students' attitudes, interests, and motivation" seemed to fit better with an agree / disagree response scale rather than a frequency scale. Therefore, a response scale with an agreement format was constructed, and each survey item was reviewed to determine if it required either a frequency or an agreement response scale. Further adjustments were then made to ensure that both assessment culture items and testing culture items were included in the sections with different response formats. As a result, in the final survey, items 1 through 13 constituted Part 1 using an agree / disagree response scale, and survey items 14 – 37 constituted Part 2 using a frequency response scale. (See the final survey in Appendix A, pp. 74-76.) Finally, a pilot survey was conducted, and items were revised based on participants' suggestions.

Survey Procedures

The Iowa State University Human Subjects committee approved the study in the summer of 1999.

On February 1, 2000, the survey was sent to 215 tenured and tenure-track College of Engineering faculty members. A cover letter from the College of Engineering Assistant Dean endorsing the survey and requesting faculty members' participation in the study was included. Copies of the letters can be found in Appendix A, pages 77-78.

A follow-up e-mail message was sent to faculty members who had not returned the survey by February 18^{th}, 2000. On February 29, 2000, 163 follow-up surveys, identical to the surveys initially distributed, were sent to faculty members who had not responded at that point. By March 22^{nd}, 132 faculty members had responded by returning the surveys. At that time, identification-coding sheets were destroyed, and data collection was considered complete.

Data Analysis

Survey Analysis

The survey, "Faculty Perceptions of Classroom Assessment," consisted of three parts. Parts one and two consisted of survey items reflecting testing and assessment cultures; they were printed on a computer-scored answer sheet. Participants responded to the items by darkening the circle corresponding to the appropriate response for each survey item. Part three consisted of demographic questions, which was printed on a sheet attached to the computer-scored answer sheet. Survey participants responded to the demographic questions by filling in a blank at the end of the question. Each survey was coded with the same identification number written on each of the two sheets.

When the surveys were returned, the computer-scored response sheets were separated from the demographic sheets, and the computer-scored response sheets were taken to the Iowa State University Computer Service's Testing and Evaluation Center. There the data from the computer-scored sheets were entered into a data set using the Statistical Package for Social Sciences (SPSS) Windows Version 9.0. At that point, the data set was transferred to a PC using SPSS. Finally, the survey's demographic sheet was matched, using the code numbers, to the assessment data, and the demographic information was entered into the data set by hand.

After all of the data were entered from the surveys, frequencies were run on all survey items, including demographics, to check for anomalies in the data set; none were found. The means for culture items 1 through 37 were calculated so missing values for an item could be replaced with the mean for that item. There were few missing values, and they were dispersed fairly evenly throughout the survey. For demographic items 38 through 42, missing data were not replaced with the means of the items because of the large range for each item.

Next, analysis of the data focused on evaluating the "psychometric quality of measures" of the survey (Green, Salkind, and Akey, 2000). This included running an exploratory factor analysis to "assess the dimensionality" (p. 291) of the assessment survey items with factor extraction and rotation as essential components. Reliability analysis using Cronbach's alpha was used to evaluate the scales' reliabilities. Following the reliability analyses, data analysis to address the study's hypotheses began. Hypotheses were tested using the Pearson product-moment correlation, the t-test, and analysis of variance. Details of the survey analysis are recorded in Chapter IV, Results.

Hypothesis Testing

Hypothesis 1: Engineering faculty will rate higher on the testing culture scale than the assessment culture scale.

The null hypothesis for the first hypothesis was "The engineering faculty members' mean testing culture score will be equal to the engineering faculty members' mean assessment culture score." Testing of the null proceeded using a one-tailed dependent *t* test because of the directional nature of Hypothesis 1. However, because it would be of interest if the testing culture scores were statistically significantly *less than* the assessment culture scores, a two-tailed test was also examined.

Hypothesis 2:There will be a negative relationship between the testing culture scores and the assessment culture scores.

The related null hypothesis was "There will be no relationship between the testing culture scores and the assessment culture scores." Testing of the null proceeded using the Pearson product-moment correlation with a one-tailed test for significance.

Hypothesis 3: Engineering departments will differ in terms of their (a) testing culture scores and (b) assessment culture scores.

The null hypothesis was "There will be no differences among engineering departments in terms of their (a) testing culture scores and (b) assessment culture scores." The test for this null proceeded using the analysis of variance (ANOVA).

Hypothesis 4: There will be a positive relationship between the assessment culture scores and the amount of exposure to experiences outside the classroom with classroom assessment: (a) the number of college courses in classroom / student assessment taken for college credit; (b) the number of semesters of participation in project LEA/RN; and (c) the number of faculty development sessions on the topic of classroom/student assessment other than Project LEA/RN. A negative relationship will occur between the testing culture scores and the same faculty characteristics.

Six null hypotheses were considered here. "There will be no relationship between assessment culture scores or testing culture scores and:

(a) The number of college courses in classroom / student assessment taken for college credit,

(b) The number of semesters of participation in project LEA/RN,

(c) The number of faculty development sessions on the topic of classroom / student assessment."

All null hypotheses were tested using the Pearson product-moment correlation with a one-tailed test of significance since a directional relationship was hypothesized.

Hypothesis 5: There will be a positive relationship between the assessment culture scores and (a) the number of years since receiving your Baccalaureate degree, and (b) the number of years teaching in higher education. There will be a negative relationship between the testing culture scores and the same faculty characteristic.

Four null hypotheses were considered here. First, “There will be no relationship between the assessment culture scores and (a) the number of years since receiving the Baccalaureate degree, and (b) the number of years teaching in higher education.” Also, “There will be no relationship between the testing culture scores and (a) the number of years since receiving the Baccalaureate degree, and (b) the number of years teaching in higher education.” These null hypotheses were tested using the Pearson product-moment correlation with a one-tailed test of significance since a directional relationship was hypothesized.

Details of the analysis are recorded in Chapter IV: Results.

CHAPTER IV

RESULTS

Introduction

The study was conducted with three purposes in mind: (a) to develop a valid tool for measuring testing and assessment cultures within engineering education, (b) to measure testing and assessment cultures in engineering education at one institution, and (c) to discover relationships between engineering faculty characteristics and the evaluation cultures at one institution. In this chapter, the results of the data analysis are presented in two parts:

Part 1: Evaluating the psychometric quality of the survey

Part 2: Testing the study's hypotheses.

Analysis of the data utilized the computer statistical program, Statistical Package for Social Sciences (SPSS) Windows Version 9.0.

Part 1: Psychometric Quality of the Survey

Evaluation of the psychometric quality of the survey was carried out to answer two questions: (a) Did the survey measure the testing and assessment cultures? and (b) Would the scales measured by the survey be reliable enough to be used when testing hypotheses? The first question was answered using factor analysis, and the second question was answered using reliability analysis, specifically Cronbach's alpha.

Factor Analysis

Because the survey items were developed to measure testing and assessment constructs (Haydel, 1997), it was anticipated that factor analysis would sort the survey items into these two domains. Investigation of this expectation occurred "in two stages: factor extraction and factor rotation" (Green, Salkind, & Akey, 2000). Factor extraction was conducted in order to "make *initial* decisions about the number of factors underlying" the survey items (p. 294). It was performed by using the "principal components solution" to determine the eigenvalues of components or extracted factors. Because an eigenvalue is an indicator of "the amount of variance of the variables accounted for by a factor" (p. 296), this determination allowed for an accounting of "the largest amount of the variability among the measured variables" (p. 294). Thirteen components had eigenvalues greater than 1, which is a standard cut-off point when using this analysis (p. 297). Table 4.1 lists the 13 factors with their related eigenvalues.

Table 4.1. Factor Extraction for Survey Components with Eigenvalues >1

Component (Factors)	Eigenvalues	% of Variance	Cumulative %
1	6.114	16.88	16.88
2	3.195	8.63	25.51
3	2.412	6.67	32.18
4	1.834	4.74	36.92
5	1.691	4.56	41.48
6	1.648	4.47	45.95
7	1.568	4.33	50.28
8	1.319	3.62	53.89
9	1.197	3.28	57.17
10	1.168	3.17	60.34
11	1.119	3.07	63.41
12	1.031	2.86	66.27
13	1.001	2.79	69.07

Because of the large number of factors with eigenvalues greater than 1, the scree test was used as another basis for determining the major factors underlying the survey (Green, Salkind, & Akey, 2000). In this procedure, the eigenvalues are plotted and the factors retained as significant are "all factors with eigenvalues in the sharp descent part of the plot before the eigenvalues start to level off" (p. 297). Using this procedure, the number of factors were reduced from 13 to 3, as seen in Figure 4.1.

In order to more closely examine the three factors extracted in the first step of factor analysis, it was necessary to do a factor rotation, which yields a correlation between each survey item and each factor (Green, Salkind, & Akey, 2000). The specific method of rotation used was the "maximum likelihood, VARIMAX" procedure. Results of this factor rotation gave insight into the "relative importance of each factor" (p. 298) by eliciting the percentage of variable variance accounted for by each factor: a) Factor one accounted for 10.62%; b) Factor two accounted for 9.77%; and c) Factor three accounted for 5.57%. Therefore, the first two factors accounted for roughly equal amounts of variance, whereas the third factor accounted for much less. Factor rotation also provided information for evaluating the nature of the factors. Table 4.2 reports the factor loadings and the correlations between survey items and factors.

In this table, the survey items are listed in the left-hand column and the loadings are recorded in the columns to the right. Any loading above $|.35|$ was considered to be a meaningful contributor to the factor and was shaded. Using this criterion, six survey items (1,6, 16, 21, 24, and 33) did not contribute to any factor and were eliminated from further analyses. The next step was to determine the underlying construct represented by each

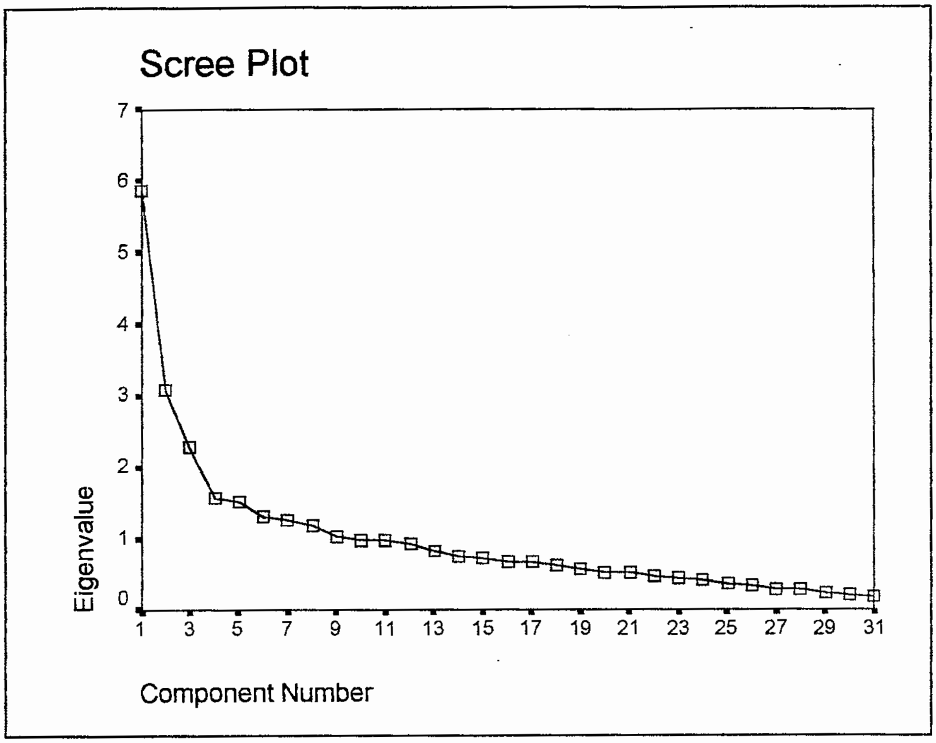

Figure 4.1. Scree Plot of the Eigenvalues.

factor by reviewing the patterns of content in the factor's shaded items. Although consideration was given to the possibility that constructs other than the ones initially designed into the survey were present, two of the factors appeared to be familiar; factor number one was an "assessment" factor, and factor number two was a "testing factor." The third factor was not as easily categorized, but "grading" emerged as the most consistent theme. Note that some items that were originally written to measure the assessment culture loaded more heavily on the testing factor than on the assessment factor—but in a negative direction.

Table 4.2. Rotated Factor Matrix

Survey Items	Factor		
	1	2	3
1. For me, a primary purpose of assessment is to assign grades to students.	-.223	.283	.085
2. My assessments have one right answer.	.008	.590	.152
3. It is possible for all my students to receive the same grade.	.002	-.013	-.420
4. My assessments focus on measuring knowledge of concepts and facts.	.120	.320	.418
5. It is important to assess students' attitudes, interests, and motivation.	.417	-.179	.075
6. Communicating student performance in my class is adequately accomplished by posting grades.	-.212	.178	.231
7. I am comfortable using group work to assess student learning.	.290	-.501	-.206
8. For me, a primary purpose of assessment is to improve student performance.	.418	-.093	-.116
9. My assessments are structured so that there can be more than one right answer.	.150	-.636	-.076
10. The number of students in my class who receive the same grade should be limited.	-.185	-.047	.379
11. My assessments require students to use their knowledge to generate solutions to real-world engineering problems.	.511	.146	-.084
12. When reporting student learning in my class, I find it necessary to write comments in addition to giving grades.	.468	.019	-.139
13. I prefer that students work alone when being assessed.	-.186	.553	.153
14. In my classroom, I encourage students to support each other as they strive to achieve high performance standards.	.429	-.291	-.265

Table 4.2, Continued

15. I review assessment results to guide further instruction in the areas of student weakness.	.485	-.113	-.079
16. I require students to return tests so questions can be used again.	.081	.170	-.060
17. I decide what will be included on each assessment without student input.	-.325	.476	-.182
18. In my classes, assessment occurs during as well as after instruction.	.403	-.070	-.175
19. I use written tests to assess my students' abilities to apply engineering concepts.	.025	.458	-.115
20. I use student conferences to communicate with students about their learning.	.614	-.127	-.041
21. I inform my students about the performance standards and grading scale to be used when assessing their achievement.	.047	.202	-.084
22. I provide opportunities, such as portfolios or journals, for students to engage in self-reflection about how and what they learned.	.543	-.206	.064
23. In my class, I am the only one who evaluates student learning.	-.238	.393	.072
24. I review tests or other assessments with my students as a means of helping them improve their own leaning.	.178	.183	.024
25. I assess skills or abilities directly by observing performances or evaluating products / projects that use engineering concepts.	.425	.092	-.017
26. As part of my assessment, I question students orally in order to evaluate the depth of their understanding.	.544	-.038	.085
27. I adjust my grading scale depending on how students perform on tests.	.064	-.103	.434
28. I provide opportunities for students to assist in the evaluation of their peers' learning. 1	.481	-.353	-.047

Table 4.2., Continued

29. My assessments require students to demonstrate skills or make products.	.353	-.015	.184
30. I use grades as the sole means of communicating learning to my students.	-.061	.424	.198
31. I involve students in deciding what will be included in each assessment.	.338	-.394	.314
32. When grading an assessment, I put the class's assessment scores in order and give each student a grade based on his / her rank in the group.	-.017	.025	.680
33. I allow students the opportunity to justify answers marked wrong.	-.053	.013	.202
34. In my classroom, I encourage students to compete with one another for the best grade.	.321	.231	.437
35. The skills that I expect students to demonstrate can adequately be assessed on paper.	-.159	.613	.056
36. My assessments evaluate students' abilities to express and defend their points of view.	.505	-.209	.190
37. I assess only after instruction has occurred.	-.136	.418	.289

The next step was to form scales to be used as dependent variables in hypotheses testing and evaluate their reliability. Two issues were addressed. First, one item loaded above .35 on both the testing and assessment factors. It was decided to include this item in the assessment scale because of its higher loading on the assessment factor (.481) when compared to the testing factor (-.353). Second, the content of each item that loaded negatively on a factor was examined to determine if reversing the wording—and the coding—of the item would be compatible with the underlying construct. For example, the loading of survey item number seven in Table 4.2 was -.501. Reversing the wording in that

that item produced the statement: " I am not comfortable using group work to assess student learning." This reversed-worded item is consistent with the construct of a testing domain. Therefore, to reverse-word the item and leave that item in the testing scale should strengthen that scale. Four survey items were reverse-worded in this manner: 3, 7, 9, and 31. Consistent with the reverse-wording of the survey items, re-entering the database was necessary, and those items were also reverse-coded so that there was consistency between the reworded item and the item scores.

Reliability Analysis

The reliability for each scale was evaluated using Cronbach's alpha. Consistency among the items in each scale was assessed, and the ". . . greater the consistency, in responses among items, the higher the coefficient alpha" (Green, Salkind, & Akey, 2000). Cronbach alpha was the analysis of choice since the underlying assumptions of the analysis were met: "For coefficient alpha, every item is assumed to be equivalent to every other item. All items should measure the same underlying dimension" (p. 305). The resulting alpha for each scale is listed in Table 4.3.

The removal of items from the scales did not increase the alpha scores: therefore, each scale was left with all of the items intact. Both the assessment scale and the testing scale had alpha scores greater than .80, and thus it was concluded that they were sufficiently reliable to be used in hypotheses testing. However, the alpha of the grading scale was considerably lower (.64). In addition, the use of the grading scale was not necessary to investigate any of the originally posed hypotheses. Therefore, the decision was made to eliminate the grading scale from further data analysis.

Table 4.3. Reliability Coefficients (Alpha) for Scales

Scales	Alpha
1. Assessment Scale	.8096
2. Testing Scale	.8062
3. Grading Scale	.6423

The survey items remaining in the assessment and testing scales were evenly distributed throughout the dimensions in spite of the reversal and removal of some items. Of the items removed, one survey item was removed from each of the dimensions of purpose, method, communication, and grading. Of the items reversed-coded, there was a similar distribution, with one item removed from grading, context, method, and target.

Part 2: Research Questions 2 and 3

Research Question 2: How do engineering faculty members rate on a scale measuring the testing culture in their classrooms?

Research Question 3: How do engineering faculty members rate on a scale measuring the assessment culture in their classrooms?

As reported in Table 4.4, the mean testing culture score for the sample of the engineering faculty was 3.16 and the mean assessment culture score for the sample was 3.45. The testing culture scores were more disperse than the assessment culture scores, as shown by the broader range of the testing culture scores which extend from a low of 1.36 (indicating little use of testing culture methods) to a high of 5.00 (indicating a consistent use of testing culture methods) and the larger standard deviation of the mean testing culture score (.578). In comparison, the assessment culture scores ranged from a low of 1.43 to a high of 4.79, with a standard deviation of .501.

Table 4.4. T- Test of Means for Assessment Culture Scores and Testing Culture Scores

	Mean	Range	Standard Deviation	$t^{(1)}$	Sig. (one-tailed)	Sig. (two-tailed)
Assessment Culture	3.45	1.43 – 4.79	.501	3.60*	.000	.000
Testing Culture	3.16	1.36 – 5.00	.578			

* Indicates a statistically significant "t" value
(1) Tests for significance were one-tailed at .05 and two-tailed at .025

Part 3: Testing the Study's Hypotheses

Hypothesis 1: Engineering faculty will rate higher on the testing culture scale than the assessment culture scale.

The null hypothesis, "The engineering faculty members' mean testing culture score will be equal to the engineering faculty members' mean assessment culture score," was tested using a one-tailed dependent *t* test for the means. As can be seen in Table 4.4, the mean testing culture score of 3.16 was not higher than the mean assessment culture score of 3.45. Therefore, the null hypothesis could not be rejected, and the research hypothesis could not be supported. This result prompted the question: Would the mean assessment culture score be significantly higher than the mean testing culture score? By examining the results of the t-test with a two-tailed perspective, this new proposition could be supported because p = .000; the mean assessment culture score was higher than the mean testing culture score. (In fact, as Table 4.4 shows, this difference would have been significant even at a one-tailed level.)

Hypothesis 2: There will be a negative relationship between the testing culture scores and the assessment culture scores.

The null hypothesis, "There will be no relationship between the testing culture scores and the assessment culture scores," was tested using the Pearson product-moment correlation. The correlation between the scales was -.442, and this value had a probability of occurrence under the null hypothesis of .000 using a one-tailed test. Therefore, the null hypothesis was rejected. A negative relationship was found between the testing culture scores and assessment culture scores. As faculty members' assessment culture scores increased, their testing culture scores decreased. Figure 4.2 illustrates this relationship.

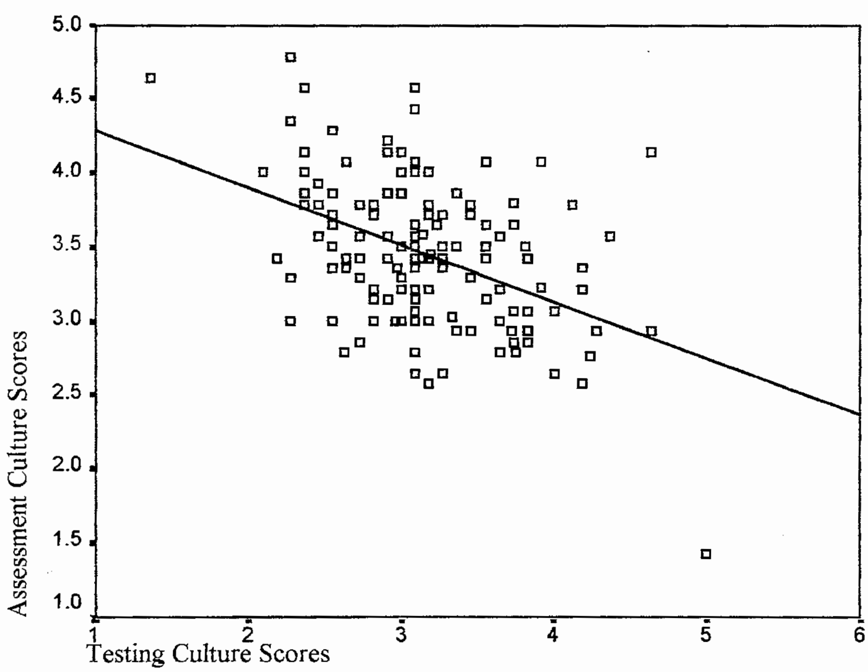

Figure 4.2 Scatterplot of Assessment Culture Scores and Testing Culture Scores

Hypothesis 3: Engineering departments will differ in terms of their (a) testing culture scores and (b) assessment culture scores.

The null hypothesis for this hypothesis stated, "There will be no differences among engineering departments in terms of their (a) assessment culture scores and (b) testing culture scores." The means for assessment and testing cultures for engineering departments are presented in Table 4.5. Analysis of variance (ANOVA) was the statistical test used to determine if any of the means were statistically significantly different. The results are presented in Table 4.6.

Table 4.5. Comparison of Departmental Means for Assessment Culture and Testing Culture

		Assessment Culture Score		Testing Culture Score	
Engineering Department	N[(1)]	Mean	Standard Deviation	Mean	Standard Deviation
College of Engineering	132	3.45	.501	3.16	.578
Aerospace Engineering and Engineering Mechanics	21	3.35	.635	3.56	.785
Agriculture and Biosystems Engineering	11	3.44	.388	2.94	.475
Chemical Engineering	11	3.44	.469	3.19	.542
Civil and Construction Engineering	23	3.38	.462	3.17	.421
Electrical and Computer Engineering	24	3.49	.479	3.22	.464
Industrial and Manufacturing Systems Engineering	10	3.65	.602	2.84	.719
Materials Science and Engineering	16	3.47	.457	3.07	.466
Mechanical Engineering	15	3.53	.519	2.93	.504

[(1)]N for the College of Engineering does not equal the sum of the departments due to non-identification of the department on a survey.

Table 4.6. ANOVA for Departmental Means

	Sum of Squares	df	Mean Squares	F	Sig.
Assessment Scale:					
Between Groups	.862	7	.123	0.475	.852
Within Groups	31.913	123	.259		
Total	32.775	130			
Testing Scale:					
Between Groups	5.374	7	.768	2.494*	.020
Within Groups	37.858	123	.308		
Total	43.232	130			

* Indicates a statistically significant "f" value
(1) Tests for significance were two-tailed at .025

The null hypothesis for the assessment culture score could not be rejected since p = .852. Therefore, the research hypothesis could not be supported. There were no apparent differences among the engineering departments in the assessment culture scores.

The null hypothesis for the testing culture was rejected with a p = .020, so a post hoc test was required to determine where the differences were among departments. Because the variances could not be assumed to be equal among the means (see Table 4.7), Dunnett's C was chosen as the appropriate post hoc test (Green, Salkind, & Akey, 2000). In spite of the previous indication that a difference was present, this more stringent post hoc test failed to support a difference. The research hypothesis that there would be a difference among departments' testing culture mean scores could not be supported.

Table 4.7. Test of Homogeneity of Variances

	Levene Statistic	df1	df2	Sig.
Assessment Culture Scale	0.733	7	123	.644
Testing Culture Scale	2.296*	7	123*	.031

*Indicates a statistical significance with a two-tailed test at .025

Hypothesis 4: There will be a positive relationship between assessment culture scores and the number of educational experiences related to the topic of classroom assessment: (a) the number of college courses in classroom / student assessment taken for college credit (Q40), (b) the number of semesters of participation in project LEA/RN (Q41), and (c) the number of faculty development sessions on the topic of classroom/student assessment other than Project LEA/RN (Q42). A negative relationship will occur between the testing culture scores and the same educational experiences.

Three null hypotheses for each scale were needed to test this hypothesis; all statistical tests used the Pearson product-moment correlation. First, the null hypotheses associated with the assessment culture scores were considered. "There will be no relationship between the assessment culture scores and (a) number of college courses in classroom / student assessment, (b) the number of semesters of participation in project LEA/RN, and (c) the number of faculty development sessions." Results of the tests are listed in Table 4.8.

The correlation between the assessment culture scores and the number of assessment college courses reported by the faculty was .265, the correlation between the assessment culture scores and the number of semesters in Project LEA/RN was .264, and the correlation between the assessment culture scores and the number of faculty development sessions was .224. All of these were significant at the one-tailed level with p values of .002, .002, and .0005, respectively. Therefore, the null hypotheses were rejected. The research hypotheses stating there were positive relationships between assessment culture scores and exposure to faculty development experiences were supported.

Next, the null hypotheses related to testing culture scores were considered. "There is no relationship between the testing culture scores and (a) number of college courses in classroom / student assessment, (b) the number of semesters of participation in project LEA/RN, and (c) the number of faculty development sessions." Results of the tests are listed in Table 4.8. The correlation between the testing culture scores and the number of college courses reported by the faculty was -.159; the correlation between the testing culture scores and the number of semesters in Project LEA/RN was -.220; and the correlation between the assessment culture scores and the number of faculty development sessions was -.267. All of these were significant at the one-tailed level with p values of .038, .007, and.001, respectively. Therefore, the null hypotheses were rejected. The research hypotheses were supported revealing there were negative relationships between testing culture scores and exposure to faculty development experiences.

Table 4.8. Correlation of Faculty Development with Assessment Culture & Testing Culture Scores

Independent Variable	**Assessment Culture Scores**		**Testing Culture Scores**	
	$r^{(1)}$	p	$r^{(1)}$	p
40. Number of assessment college courses	.265*	.002	-.159*	.038
41. Number of semesters in Project LEA/RN	.264*	.002	-.220*	.007
42. Number of faculty development sessions other than Project LEA/RN	.224*	.0005	-.267*	.001
40. With cases of "0" removed.	.348	.072	-.195	.212
41. With cases of "0" removed	.138	.155	-.223	.045
42.With cases of "0" removed	.249*	.009	-.218*	.021

* Indicates a statistically significant "r" value.
(1) Tests for significance were one-tailed at .05.

Examination of the frequency distribution of responses to questions 40, 41 and 42 revealed that a large number of faculty members reported no faculty development experiences. See Tables 4.9, 4.10, and 4.11 for distributions in each of the questions.

To see if a tighter relationship could be established between the dependent and independent variables for these measures, all cases that reported no experiences for each of the variables were deleted and correlations were recalculated. As reported in the bottom portion of Table 4.8, there was a slight increase in the magnitude of some correlations.

Table 4.9. Frequencies of Classroom/Student Assessment College Courses

Number of College Courses	Frequency	Percent	Valid Percent	Cumulative Percent
0	107	81.1	84.9	84.9
1	8	6.1	6.3	91.3
2	5	3.8	4.0	95.2
3	1	0.8	0.8	96.0
4	3	2.3	2.4	98.4
8	1	0.8	0.8	99.2
12	1	0.8	0.8	100.0
Total	126	95.5	100.0	
Missing	6	4.5		
Total	132	100.0		

Table 4.10. Frequencies of Semesters in Project LEA/RN

Number of Semester	Frequency	Percent	Valid Percent	Cumulative Percent
0	69	52.3	54.3	54.3
1	19	14.4	15.0	69.3
2	10	7.6	7.9	77.2
3	3	2.3	2.4	79.5
4	11	8.3	8.7	88.2
5	5	3.8	3.9	92.1
6	5	3.8	3.9	96.1
7	1	0.8	0.8	96.9
8	4	3.0	3.1	100.0
Total	127	96.2	100.0	
Missing	5	3.8		
Total	132	100.0		

Table 4.11. Frequencies of Faculty Development Sessions Other Than Project LEA/RN

Number of Sessions	Frequency	Percent	Valid Percent	Cumulative Percent
0	41	31.1	31.5	31.5
1	19	14.4	14.6	46.2
2	26	19.7	20.0	66.2
3	21	15.9	16.2	82.3
4	4	3.0	3.1	85.4
5	2	1.5	1.5	86.9
6	4	3.0	3.1	90.0
8	1	.8	0.8	90.8
9	1	.8	0.8	91.5
10	10	7.6	7.7	99.2
20	1	.8	0.8	100.0
Total	130	98.5	100.0	
Missing	2	1.5		
Total	132	100.0		

The correlations between assessment culture scores and Q 40 and Q 42 increased slightly with values of.348 and .249 respectively. But, only the correlation with Q 42 remained significant at the one-tailed level of significance at p = .009. The magnitude of the correlations between testing culture scores and Q 40 and Q 41 were also slightly higher, with values of -.195 and -.223, respectively; however, only the correlation with Q 41 both improved and remained significant at the one-tailed level of significance at p = .045. The failure to remain significant with an increase in the magnitude of the correlation could be attributed to the relatively small number of cases (n = 19) in the analysis with Q 40.

Due to the very small improvements in the correlations and a decrease in the significance of the correlations, the judgement was made to work with the results inclusive of the "no exposure" cases.

Finally, an additional analysis procedure, multiple regression analysis, was conducted to see which faculty development experiences were the best predictors of the assessment culture or testing culture scores. The results are listed in Table 4.12.

For both assessment culture and testing culture scales, the number of college courses added the most prediction and was statistically significant. The number of semesters of Project LEA/RN also added to the predictability of the assessment culture scale, with a significance of p = .008, but it did not add to the predictability of the testing culture scale since p = .070. The number of faculty development sessions did not add to the predictability of either scale.

Table 4.12. Multiple Regression of Evaluation Culture Scores on Faculty Development Experiences

Predictor	Assessment Culture Scale			Testing Culture Scale		
	Standardized Coefficients Beta [1,3]	Zero-Order Correlation[2,3]	Signifi-cance	Standardized Coefficients Beta [1,3]	Zero-Order Correlation[2,3]	Signifi-cance
Number of college courses in classroom / student assessment	.278*	.266*	.001	-.214*	-.215*	.015
Number of semester of Project LEA/RN	.247*	.262*	.008	-.171	-.222	.070
Number of faculty development sessions other than LEA/RN	.095	.219	.304	-.172	-.260	.069

* Indicates a statistically significant value.
[1] Correlation between each predictor and scale controlling for all other predictors.
[2] Correlation between each predictor and scale.
[3] Tests for significance were one-tailed, at .05.

Hypothesis 5: There will be a positive relationship between the assessment culture scores and (a) the number of years since receiving a Baccalaureate degree, and (b) the number of years teaching in higher education. There will be a negative relationship between the testing culture scores and the same faculty characteristic.

The null hypotheses were "There will be no relationship between the assessment culture scores and (a) the number of years since receiving a Baccalaureate degree, and (b) the number of years teaching in higher education." Also, "There will be no relationship between the testing culture scores and the same independent variables. The Pearson product-moment correlation technique was used to test the null hypotheses and the results are listed in Table 4.13.

It can be seen from this table that the correlations between the independent and dependent variables were low, and none of the null hypotheses could be rejected. The research hypotheses were not supported. There is no evidence that there is a correlation between assessment culture or testing culture scores and years since receiving a Baccalaureate degree or years teaching in higher education.

Table 4.13. Summary of Correlations Between Evaluation Culture Scores and Both (a)Years Since Receiving a Baccalaureate Degree, and (b) Number of Years Teaching in Higher Education [(1)]

Independent Variable	Assessment Culture Scores		Testing Culture Scores	
	r[(2)]	p	r[(2)]	p
Number of Years Since Receiving Baccalaureate Degree	.027	.387	.093	.145
Number of Years Teaching in Higher Education	.051	.286	.039	.328

[(1)]For Both Q38 & Q39, N = 130; [(2)]Tests for significance were one-tailed, at .05.

Summary of Findings

- Testing culture scores and assessment culture scores and assessment culture scores were negatively related.
- Faculty rated higher on the assessment culture scale than on the testing culture scale.
- There was a positive relationship between the assessment culture scores and the amount of exposure to faculty development regarding student and classroom assessment.
- There was a negative relationship between the testing culture scores and the amount of exposure to faculty development regarding student and classroom assessment.
- Assessment culture scores and testing culture scores were not related to department, the number of years since receiving a Baccalaureate degree, or the number of years teaching in higher education.

CHAPTER V.

DISCUSSION OF RESULTS AND RECOMMENDATIONS

Summary of the Study

The purpose of this study was: (a) to develop a valid tool for measuring testing and assessment cultures within engineering education, (b) to measure testing and assessment cultures in engineering education at Iowa State University, and (c) to discover relationships between engineering faculty characteristics and the evaluation cultures at Iowa State University. The results of this study will be discussed in relationship to these purposes.

Part 1: Survey Validity

The first purpose of this study was to develop a tool to measure the evaluative culture for engineering education. Ensuring the validity of this tool was the initial concern, and this was handled by reviewing the literature, using a panel of experts in the fields of education and engineering, and piloting the survey.

The survey developed for this study was based on a study done by Haydel (1997) for secondary education teachers. Through the panel of experts, it was discovered that some of her survey items required rewording to better represent the nature of engineering education. One result of this improvement effort was the addition of a "grading" dimension to the initial five domains proposed by Haydel. Factor analysis of this study's survey items discounted this addition. The items that aligned with grading emerged as an unanticipated third factor and that factor's scale did not prove to be very reliable ($\alpha = .64$). Therefore, it was concluded that adding grading practices as a separate dimension would not improve Haydel's survey.

On the other hand, the domains of the survey -- testing and assessment -- did line-up with the other two significant factors in the factor analysis. Since those factors were sufficiently reliable ($\alpha = .81$ for each), support was bolstered for the view that scores derived from the survey were valid representations of underlying constructs.

Construct validity was also supported by finding a negative relationship between the testing and assessment culture scores. Since the two scales represent contrasting paradigms, the finding of a negative correlation is consistent with their construct validity. The scales should be negatively correlated since it would not be logical that a faculty member could operate at high or low levels in both paradigms at the same time. Based on these findings, it was concluded that the survey instrument was a valid tool for discovering information about the evaluative culture in the College of Engineering at Iowa State University.

Part 2: The Evaluation Culture – Testing and Assessment

The first three hypotheses of the study predicted outcomes related to the evaluation culture in the College of Engineering at Iowa State University.

The first hypothesis predicted that engineering faculty members would rate higher on the testing culture scale than on the assessment culture scale. The rationale for the hypothesis was based on (a) the understanding that the behavioral educational paradigm has been part of the educational system for a much longer time than the constructivist paradigm and (b) research acknowledging that cultural change does not happen rapidly (Freedman, et al., 1979). Therefore, since more faculty members would identify with the entrenched behavioral model of learning, the mean testing culture score should be higher than the mean assessment culture score. Results of this study did not confirm this expectation. Instead, the assessment average was significantly higher than the testing average.

There are several possible explanations for the unanticipated finding of a higher average assessment culture score for engineering faculty. First, it could be argued that the higher assessment average was due to the peculiar nature of engineering education which is traditionally a hands-on, product-oriented discipline, especially in the engineering design courses. In that environment, evaluation methods associated with the assessment culture (e.g., direct assessment through student projects, using assessment projects with real-world applications) would be part of the traditional engineering culture and would not necessarily reflect a shift toward a constructivist educational paradigm. However, reliability analysis of the scales did not support this explanation. The project-oriented and direct observation survey item responses were highly correlated with the items on the assessment scale only. If those items reflected the general engineering culture, they would have been more evenly distributed between the two scales because, for some individuals, these assessment practices would have been expected to co-exist with traditional testing practices.

Another explanation for the higher mean assessment score might be found in the type of faculty members who were surveyed. It is possible that the population for this study was composed of a larger percentage of faculty members willing to engage in risk taking due to their tenured status (Caffarella and Zinn, 1999) as discussed in Chapter 2. Caffarella and Zinn (1999) proposed that senior and tenured faculty members are more likely to be involved in activities where a higher risk of failure is present, and one such high risk activity would include shifting to a new paradigm in education and adopting assessment culture activities. Since the population identified for this study was tenured and tenure-track faculty members, it is possible they took more risk than the general faculty population in their educational practices, thus producing the unanticipated elevation in mean assessment culture scores and the lowered mean testing culture score.

Also there is the possibility that the sample of faculty members returning the survey did not represent the entire population in that they were more likely to be "innovators" or "early adopters" (Rogers, 1995; Geoghegan, as cited in Gray, 1997) or representatives of Murray's (as cited in Bergquist & Phillips, 1977) "high-need-achievers". If so, they may have been more likely to attend "risky" faculty development targeted at shifting the educational paradigm, and they also would have been more likely to adopt those associated evaluation culture practices as reflected in the higher than expected mean assessment culture score. Therefore, again, the study's results may have overestimated the college faculty members' identification with the assessment culture.

Finally, there is one other explanation. It is possible that the engineering faculty members surveyed do represent the College faculty population and that the mean assessment culture score actually reflects the dominant educational paradigm within the

College. This explanation is plausible when the percentage of those engaging in faculty development is considered. In the literature of "change" by Geoghegan (as cited in Gray, 1997) 16 % of a population (innovators) are the most receptive to innovation. The "early adopters," who are a little more resistant to change, account for the next 34%. Those less willing to take risks and innovate make up the remaining 50% of the population. Since 64% of faculty members responding to this study's survey had been engaged in at least one form of faculty development addressing classroom and student assessment, it is probable that the population most likely to adopt new methods of assessment have been reached.

The third hypothesis predicted that the engineering departments would differ in terms of their testing culture scores and their assessment scores. This also was not supported by the study results. The original hypothesis was based on personal observations from contact with the engineering departments during committee work for accreditation, but a limited population of faculty members from each department was involved in this contact. No empirical data were available for corroboration of the observations. The study supports the conclusion that the observations were not accurate.

Part 3: Relationships Between Faculty Characteristics and the Evaluation Cultures

The next two hypotheses of the study predicted outcomes related to the characteristics of faculty members in the College of Engineering at Iowa State University and the evaluation culture in the College.

It was predicted that there would be a positive relationship between faculty exposure to faculty development involving student / classroom assessment and the assessment culture, and a negative correlation between the same exposure and the testing culture. This prediction was supported. As exposure to faculty development increased, assessment scores increased and testing scores decreased. There is one caution; as mentioned previously, causality cannot be implied from the correlation, so although it is tempting to report that faculty development makes a difference and raises assessment culture scores, it would be imprudent.

An examination of other factors associated with faculty development might lend insight into the relationship between the amount of faculty development experience and the culture scores. The fact that 64% of faculty members have had at least one experience in faculty development implies that innovators and early adopters have probably already engaged in the innovation process of shifting to the assessment paradigm. Since they were the "first on board," they would have had more time to engage in faculty development sessions. Also, as Murray (as cited in Berguist & Phillips, 1977) described, since they are persistent and not averse to failure, it is probable that they also engaged in more experimental assessment behaviors.

Conversely, those faculty members who have had relatively few faculty development experiences have higher testing culture scores and are probably in the late majority cohort of innovators – they will not adopt an innovation until it is proven. Since they are unwilling to take risks (Gray, 1997), it is possible that they also minimize exposure to risk by minimizing participation in faculty development that forces them to confront educational change. This lack of exposure to faculty development further entrenches their resistance to adoption of a new educational paradigm, and assessment culture practices are not implemented. Also, to further complicate adoption of change, this group requires more support from their peers and administration when trying an innovation (Gray, 1977), and because of their isolation from faculty development, they are the least likely faculty members to receive that support. This cycle of resistance to change, and self-imposed isolation, which leads to more resistance, may cause them to further retreat to the familiar paradigm of the teaching-centered culture (Berguist & Phillips, 1977).

The implication of this is that further adoption of assessment culture behaviors by faculty may become increasingly difficult because the first three cohorts (innovators, early adopters, and early majority) have already engaged the assessment culture. The rest of the faculty members not already operating in the learning-centered paradigm will need larger amounts of administrative support, peer support, and external validation since they are less intrinsically driven (Bergquist & Phillips, 1977; Caffarella & Zinn, 1999; Fulton, Licklider, & Schnelker, 1997; Walvoord & Pool, 1998). Administrative support also would be valuable in documenting that change in assessment methods is beneficial to the educational system, the students, and especially the faculty members. Finally, since pressure is mounting for accountability to stakeholders, increased friction might occur as the distance between the innovators and the laggards widens. Leadership and colleague support will be needed to bridge this gap.

The results of the multiple regression analysis of evaluation scores on faculty development variables (number of college courses taken in classroom / student assessment, number of semesters in Project LEA/RN, and number of faculty development sessions other than Project LEA/RN) are also relevant to the discussion. No specific faculty development experience was shown to be overwhelmingly predictive of either the assessment culture or testing culture scores. In the multiple regression, the number of college credit courses and the number of semesters in Project LEA/RN contributed about equally to the assessment culture and testing culture scores. The faculty development sessions outside of Project LEA/RN and number of college credit courses added no significant extra prediction to either testing or assessment culture scores. From this it can be concluded that the *experience* of participating in "assessment oriented" faculty development was related to an increase in assessment culture scores. The *type* of assessment oriented faculty development was not the important factor.

The last hypothesis predicted that a positive relationship would exist between the assessment culture scores and the number of years since receiving the Baccalaureate degree and the number of years the faculty members spent teaching in higher education. A negative relationship would exist with the testing culture scores. None of these predictions were supported. If it is assumed that the rationale on which this hypothesis was based is valid (i.e., that professors' interests shift from research to teaching with increasing age, Fulton and Trow – as cited in Austin & Gamson, 1983), the lack of support for the hypothesis could be due sampling error. Repeating the study using the same demographic information and other exploratory information (e.g., the number of years faculty members worked in industry, etc.) should yield more insight into this result.

On the other hand, this result may also lend credibility to the assumption that the older a person becomes the more resistant they become to change. Additional research may allow definitive conclusions about age and responsiveness to adopting a learner-centered educational paradigm.

The results of this study suggest that, in order to promote change, administration and faculty members must know where faculty members fit in Roger's model of change. Leadership cannot rely on demographic information to make inferences about faculty adaptation to change since this information does not give a true picture of faculty members' propensity to engage in change. To continue to operate without the knowledge of where faculty fit within the model of change will raise frustration levels for both leadership and faculty members as each does not meet the others' expectations.

Summary in Brief

This study provided the opportunity to develop an instrument to measure the evaluative culture in the College of Engineering. The instrument was found to be valid and reliable. The engineering faculty members' beliefs and practices are somewhat more in line with an assessment culture than a testing culture. Of the faculty characteristics surveyed, the one that demonstrated the strongest relationship to an increasing assessment culture score and a decreasing testing culture score was the amount of exposure to faculty development.

Since this study was conducted in one location with a limited population, generalizing the study to another population should be done with caution.

Recommendations for Further Research

1. To provide further validation of the survey and improve it as a measurement instrument, a qualitative study with engineering faculty members should be conducted

to refine the underlying dimensions and also to see if there is continuity between the responses of the faculty and their actual classroom practices.

2. A study using all engineering faculty members in the college should be conducted to see if there are significant differences between the faculty members with whom the college has a commitment, the tenured and tenure-track, and those faculty who have a more tenuous relationship with the college, temporary and adjunct professors.
3. A longitudinal study should be conducted in the College of Engineering, using the same instrument and variables, in order to document shifts in the evaluation culture and educational paradigms.
4. A study examining how students perform in different evaluative cultures would provide insight into relationships between the evaluative culture and student learning.
5. To provide additional insight into the relationships between the evaluative cultures and faculty development, further examination should be made of how the evaluative culture changes when faculty development is varied.
6. The same study should be conducted at other engineering education institutions to improve the generalizability of the findings.

APPENDIX : SURVEY INSTRUMENT

Faculty Perceptions of Classroom Assessment

The following survey has three parts; two parts are on this survey sheet and one part is on an attached sheet. Please read the directions to each part and complete the associated survey items. There are no right or wrong answers so please answer truthfully and complete all items.

Part 1
Directions: Read each statement and, using the following response scale, fill in the corresponding oval in the answer area.
1= Strongly Disagree 2= Disagree 3= Neither Agree or Disagree 4=Agree 5= Strongly Agree

1. For me, a primary purpose of assessment is to assign grades to students.
2. My assessments have one-right answer.
3. It is possible for all my students to receive the same grade (e.g., all As, Bs, Cs, Ds, or Fs).
4. My assessments focus on measuring knowledge of concepts and facts.
5. It is important to assess students' attitudes, interests, and motivation.
6. Communicating student performance in my class is adequately accomplished by posting grades.
7. I am comfortable using group work to assess student learning.
8. For me, a primary purpose of assessment is to improve student performance.
9. My assessments are structured so that there can be more than one right answer.
10. The number of students in my class who receive the same grade (e.g., the number receiving As, Bs, Cs, Ds, or Fs; or the number receiving passing or failing) should be limited.
11. My assessments require students to use their knowledge to generate solutions to real-world engineering problems.
12. When reporting student learning in my class, I find it necessary to write comments in addition to giving grades.
13. I prefer that students work alone when being assessed.

Part 2
Directions: Read each statement, and using the following response scale, fill in the corresponding oval in the answer area.

1 = Never 2 = Infrequently 3 = Sometimes 4 = Usually 5 = Always

14. In my classroom, I encourage students to support each other as they strive to achieve high performance standards.
15. I review assessment results to guide further instruction in the areas of student weakness.
16. I require students to return tests so questions can be used again.
17. I decide what will be included on each assessment without student input.
18. In my classes, assessment occurs during as well as after instruction.
19. I use written tests to assess my students' abilities to apply engineering concepts.
20. I use student conferences to communicate with students about their learning.

Continue on the Next Page

Faculty Perceptions of Classroom Assessment, Part 2 Cont.

1 = Never 2 = Infrequently 3 = Sometimes 4 = Usually 5 = Always

21. I inform my students about the performance standards and grading scale to be used when assessing their achievements.
22. I provide opportunities, such as portfolios or journals, for students to engage in self-reflection about how and what they learned.
23. In my classes, I am the only one who evaluates student learning.
24. I review tests or other assessments with my students as a means of helping them improve their own learning.
25. I assess skills or abilities directly by observing performances or evaluating products / projects that use engineering concepts.
26. As part of my assessment, I question students orally in order to evaluate the depth of their understanding.
27. I adjust my grading scale depending on how students perform on tests.
28. I provide opportunities for students to assist in the evaluation of their peers' learning.
29. My assessments require students to demonstrate skills or make products.
30. I use grades as the sole means of communicating learning to my students.
31. I involve students in deciding what will be included on each assessment.
32. When grading an assessment, I put the class's assessment scores in order and give each student a grade based on his/her rank in the group.
33. I allow students the opportunity to justify answers marked wrong.
34. In my classroom, I encourage students to compete with one another for the best grade.
35. The skills that I expect students to demonstrate can adequately be assessed on paper.
36. My assessments evaluate students' abilities to express and defend their points of view.
37. I assess only after instruction has occurred.

Continue with Part 3 on the attached sheet.

Faculty Perceptions of Classroom Assessment, Continued.

Part 3
Directions: For each question, please fill in the blank and return this paper with your completed answer sheet.

38. How many years has it been since you received your Baccalaureate degree? ______ yrs.

39. How many years have you been teaching in higher education? ______ yrs.

40. How many college courses in classroom / student assessment have you taken for college credit?

______ courses

41. How many semesters have you participated in Project LE/ARN? ______ semesters

42. Besides project LE/ARN, in how many faculty development sessions on the topic of classroom / student assessment (e.g., workshops, training seminars) have you participated?

______sessions

43. In which department do you teach? _______
A. Aerospace Engineering and Engineering Mechanics
B. Agriculture and Biosystems Engineering
C. Chemical Engineering
D. Civil and Construction Engineering
E. Electrical and Computer Engineering
F. Industrial and Manufacturing Systems Engineering
G. Materials Science and Engineering
H. Mechanical Engineering

Thank You!

Please return both survey sheets in the enclosed envelope.

TO: Iowa State University, College of Engineering Faculty
FROM: Loren Zachary, Assistant Dean
RE: Assessment Survey
DATE: January 18, 2000

I am aware of the many challenges all of us face as we try to provide our students with the best engineering education possible. It is sometimes difficult to balance the different demands on our time and energy.

With this in mind, I am requesting that you take the time to participate in the college-wide study that will help us determine what is happening in the College in the arena of student assessment. Student assessment is one of the pedagogical issues emerging as having a profound impact on our students' learning and it is also of interest in the accreditation process. Your participation will ensure that we get a complete picture of faculty perceptions in the College and will therefore give us a better tool to work with as we strive to meet your developmental needs. Also, the honesty in your response will help us establish a baseline of our teaching and assessing practices so we can track changes in these practices over the years.

To participate in the study, please read the enclosed letter from Ms. Dieterich and fill out and return the attached survey, "Faculty Perceptions on Classroom Assessment." If you have any questions concerning this survey, please contact Ms. Dieterich at dieteann@iastate.edu.

As you are a valued faculty member, I appreciate your consideration in this research effort. Together we can continue to improve the already high-value education we provide for our students.

L. Zachary

January 19, 2000

Iowa Sate Engineering Faculty,

As you know, engineering faculty today are being challenged to examine their mission, set goals, develop student outcomes, measure them, develop assessment plans and teach students. In an effort to understand one aspect of this challenge, I am investigating in my Master's Thesis the way in which engineering faculty approach student assessment in their classrooms. This knowledge can then be used in faculty development as the engineering profession continues to address issues of pedagogy and how it affects student learning.

This is where I need your help. With the knowledge and approval of the College of Engineering ABET committee, I am writing to ask you to complete the enclosed survey about your assessment in your courses. Your participation in this investigation is voluntary and individual responses will remain confidential. The surveys are numbered, but this is only to allow me to contact those who do not respond. The numerical coding will be removed after any follow-up is completed.

Completion and return of the attached survey is your consent to participate in this investigation. After you have answered the survey questions, please return the instrument in the enclosed envelope.

I appreciate your assistance in expanding our understanding of the culture of student assessment in higher education in the ISU College of Engineering. If you have questions or are interested in any aspect of this project, please contact me at: 1200 Howe Hall, Iowa State University, Ames IA 50011.

Sincerely,

Ann Dieterich
Graduate Student in Education
dieteann@iastate.edu

BIBLIOGRAPHY

Accreditation Board for Engineering and Technology (ABET) 2000. (2000). ABET Homepage. [on-line], Available: http://www.abet.org.

American Society for Engineering Education. (1998). How do you measure success?: Designing effective processes for assessing engineering education. Washington, D.C.: ASEE Professional Books.

Austin, A. E., & Gamson, Z. F. (1983). Academic workplace: New demands, heightened tensions, ASHE-ERIC Higher Education Research Report No. 10, Washington, D.C.: Association for the Study of Higher Education.

Bergquist, W. H., & Phillips, S. R. (1977). In G. H. Quehl (Ed.), A handbook for faculty development (Vol. 2). Washington, D.C.: Council for the Advancement of Small Colleges, published in cooperation with Cardiff-by the-Sea, CA: Pacific Soundings Press.

Black, P., & Wiliam, D. (1998). Inside the black box: Raising standards through classroom assessment. Phi Delta Kappan, 80, 139-148.

Burton, J. D., & White, D. M. (1999). Selecting a model for freshman engineering design. Journal of Engineering Education, 88, 327-332.

Caffarella, R. S., & Zinn, L. F. (1999). Professional development for faculty: A conceptual framework of barriers and supports. Innovative Higher Education, 23, 241-254.

Caine, R. N., & Caine, G. (1997). Education of the edge of possibility. Alexandria, VA: Association for Supervision and Curriculum Development.

Campbell, W. E., & Smith, K. A. (Eds.). (1997). New paradigms for college teaching. Edina, MN: Interaction Book Company.

Erwin, T. D. (1991). Assessing student learning and development: A guide to the principles, goals, and methods of determine college outcomes. San Francisco: Jossey-Bass, Inc.

Ewell, P. T. (1991). To capture the ineffable: New forms of assessment in higher education. Review of Research in Higher Education,17, 75-124.

Freedman, M., Brown, W., Ralph, N., Shukraft, R., Bloom, M., & Sanford, N. (1979). Faculty development: Theory and action. In Academic culture and faculty development (pp. 155-165). Berkeley, CA: Montaigne Press.

Fulton, C., Licklider, B., & Schnelker, D.L. (1997). Revisioning faculty development: Improving teaching and learning. Journal of Staff, Program, and Organizational Development, 15, 17-28.

Gall, M. D., Borg,W., & Gall, J. P. (1996). Educational research: An introduction. (6th ed.). White Plains, NY: Longman.

Glaser, R., & Silver, E. (1994). Assessment, testing, and instruction: Retrospect and prospect. Review of Research in Higher Education, 20, 393-419.

Gray, P. J. (1997). Viewing assessment as an innovation: Leadership and the change process. In P.Gray. & T. Banta (Eds.), New Directions for Higher Education, 100, 5-15.

Green, S. B., Salkind, N. J., & Akey, T. M. (2000). Using SPSS for windows: Analyzing and understanding data (3rd ed.). Upper Saddle River, NJ: Prentice Hall.

Haydel, J. B. (1997). Measuring the evaluative culture of Classrooms. Paper presented at the annual meeting of the American Education Research Association, Chicago.

Holyer, R. (1999, July). The road not taken. Change [On-line serial], Available: http://www.aahe.org/change/so982.htm

Huba, M. E. & Freed, J. E. (2000). Learner-centered assessment on college campuses: Shifting the focus from teaching to learning. Needham Heights, MA: Allyn & Bacon.

Kean, M. H. (1995). The national testing movement, *redux*. The Clearing House, 71, 201-204.

Kim, J., McLean, J. E., & Iran-Nejad, A.(1996). The integration of teaching, testing, and learning. Educational Research Quarterly, 19, 29-51.

Kleinsasser, A. M., (1995). Assessment culture and national testing. The Clearing House, 71, 205-210.

Kumar, S. & Jalkio, J. (1999). Teaching mathematics from an applications perspective. Journal of Engineering Education, 88, 275-279.

Lautenbacher, G. E., Campbell, J. D., Sorrows, B. B., & Mahling, D. E. (1997). Supporting collaborative, problem-based learning through information system technology. Proceedings of the ASEE 1997 Frontiers in Education Conference, USA, 1252-1256.

Licklider, B., Schnelker, D., & Fulton, C. (1997). Revisioning faculty development for changing times: The foundation and framework. Journal of Staff, Program, and Organizational development, 15, 121-134.

Rogers, E. M. (1995). Diffusion of innovations. New York: Free Press.

Senge, P. M. (1990). The fifth discipline: The art and practice of the learning organization. New York: Doubleday.

Shepard, L. A. (1991). Psychometricians' beliefs about learning. Educational Researcher, 20, 2-16.

Smith, K. A., & Waller, A. (1997). A new paradigm for engineering education. Proceedings of Frontiers in Education Conference, 1423-1427.

Tobias, S. (1990). They're not dumb. They're different: A new 'tier of talent' for science. Tuscon, AZ: Research Corporation [A Foundation for the Advancement of Science].

VanTassel-Baska, J. (1998). The development of academic talent: A mandate for educational best practice. Phi Delta Kappan, 80, 760-763.

Walvoord, B. E., & Pool, K. J. (1998). Enhancing pedagogical productivity. New Directions for Higher Education, (103) 35-39.

Wiggins, G. (1998). Educative assessment: Designing assessment to inform and improve student performance. San Francisco: Jossey-Bass

Zahorik, J. A. (1995). The emergence of constructivist teaching. [Bulletin]. Bloomington, IN: Phi Delta Kappa Educational Foundation.