

**PLATFORM UNITY/ALIGNMENT BETWEEN COURSE PLANNING AND TESTING
DECISIONS IN CRITERION-REFERENCED SITUATIONS**

by

Thomas Charles Leitzel

Dissertation submitted to the Faculty of the
Virginia Polytechnic Institute and State University
in partial fulfillment of the requirements
for the degree of

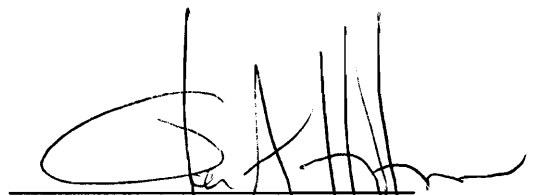
DOCTOR OF PHILOSOPHY

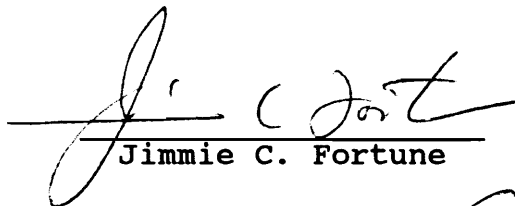
in

Community and Junior College Administration

APPROVED:


Don G. Creamer


Glen A. Holmes


Jimmie C. Fortune


Richard A. Pointer


Daniel E. Vogler
Chairman

April, 1993

Blacksburg, Virginia

C.2

LD
5655
V856
1993
L4958
C.2

**PLATFORM UNITY/ALIGNMENT BETWEEN COURSE PLANNING AND TESTING
DECISIONS IN CRITERION-REFERENCED SITUATIONS**

by

Thomas Charles Leitzel

Committee Chairman: Daniel E. Vogler
Community and Junior College Administration

(ABSTRACT)

The purpose of this study was to document the effects of applying the platform unity concept, a concept based on the principle of curriculum alignment. This principle states that the planned, the delivered, and the tested curricula are congruent. Specifically, platform unity aligns planned, domain-referenced content with appropriate test types. Mathematical formulae were created to determine numerically if planned and tested content were congruent. In addition, four other constructs were examined. They included overtesting and undertesting of course content and, effectiveness and efficiency of test item type selection.

The purpose was fulfilled through a two-phase survey. Thirty, full-time computer faculty in 15 randomly selected Virginia community colleges were given content goal statements and asked if the content were planned and tested. If content were tested, faculty indicated the type of test they used. Numerical indices, based on formulae developed in this study, were used to analyze the degree of platform unity, effectiveness, and efficiency, as well as overtesting and

undertesting of content. The calculated indices provide useful data and measurement tools to help faculty evaluate teacher-made, criterion-referenced tests.

A Chi-square Goodness of Fit Test was used to compare faculty responses to planning and testing decisions as determined by an outcome-based curriculum model. The model served as the expert system. Based on the expert system, a Table of Test Type Specifications provided the recommended test item types used to test domain-referenced content plans.

The hypothesized proportions in the Chi-square test allowed for a five percent margin of error. Data indicated significant differences ($p < .01$) between content plans and the test types used to test content. Based on the analysis, it was determined that faculty do not plan and test content congruently across three levels of cognitive content. The levels included fact, understanding and application. Also, faculty tended to overtest content; they were effective in their selection of test types, but not efficient.

Most of the faculty agreed with the expert system test type recommendations when informed of planning and testing decisions that followed curriculum alignment principles. Based on these findings, it is proffered that faculty planning and testing decisions could be improved if information about platform unity, effectiveness, and efficiency were provided through professional development.

ACKNOWLEDGEMENTS

The author wishes to express his sincere gratitude to numerous individuals who provided support, guidance, and assistance in the completion of the doctoral degree and dissertation.

First of all, I am grateful to members of the doctoral committee for their support and encouragement. Dr. Jimmy C. Fortune planted the seed for the study with his insight on academic management. Dr. Glen A. Holmes agreed to serve on the committee when a vacancy existed. Dr. Richard A. Pointer traveled many miles from Minnesota to personally demonstrate his support. Dr. Don G. Creamer's professional interest in community colleges provided a dimension that kept the study focused and conceptually sound.

I consider my advisor and chairman, Dr. Daniel E. Vogler, a wonderful friend and mentor. His personal contributions to the field are underrated. The outcome-based curriculum model he developed is a revolutionary tool for the promotion of effectiveness and accountability in the academic community. I value the opportunities and challenges Dr. Vogler provided. His prompt, thorough, and thoughtful responses to my manuscript drafts were very much appreciated.

In addition to the dissertation committee members, I recognize with appreciation other professors at Virginia Tech

who dramatically shaped my academic career. In alphabetical order, this esteemed group included, Drs. Aaron, Atwell, Belli, Clowes, Creamer, Cross, Eaton, Fortune, Garrison, Hoerner, Holmes, Hunt, Lane, Lichtman, Morgan, Stalcup, Sullins, Williams, Vogler, and Wolf. In addition, Dr. David Hutchins provided valuable input to the study during its formidable stages.

The author is appreciative of respondents in the VCCS who provided input through the survey instruments. I'm also grateful to colleagues from Maryland and Virginia who provided help during the field-test phases.

While in residence at Blacksburg, I had the privilege to work as a graduate research assistant for Dr. Samuel D. Morgan. Dr. Morgan's professional insights were valuable in shaping my personal ambitions. Opportunities for publication and presentations were made possible through his guidance.

My fellow classmates and graduate assistants who shared in the experience included Abbot, Art, Aubrey, Barbara, Becky, Bruce, Carol, Finn, Frank, Fred, Gary, Jack, Jim, Liz, Lynda, Marcia, Margaret, Patrick, Peg, Sandra, Susan, Terry, and Victor. Maintaining your friendship will always be important.

A special word of thanks goes to Ronnie Bartley, a fellow graduate student. Ronnie provided assistance in dBase III Plus, helped refine the formulae used in the study, and put me

in touch with key VCCS computer personnel.

I appreciate Dr. Steve Culver's thorough editing of this manuscript. His insight on assessment added many substantive changes.

In previous experiences, there were individuals who shaped my career aspirations. My first mentor was Paul W. Goldfeder who inspired my quest for knowledge. My former bosses, Drs. Don Bergerstock and Stuart Bounds encouraged my schooling and provided challenges that made my academic work meaningful.

A special thanks goes to my dear friends, Pauline, and, Luke and Theda who supported me from a distance with their thoughts and prayers. Curious family members seemed interested in my progress. My mother-in-law, Patricia, was always anxious for the latest rate of return figures. In-laws, Bret and Karyn, and Jim showed support by inquiring about my progress. Occasional visits to them yielded necessary distractions from reading and writing obligations. Cards, letters, and phone calls from cousins Ceylon and Karen demonstrated the sustaining love of family. Also, I appreciate my dear friend Phil who was faithful with cheerful phone calls.

To my parents, please accept my deepest appreciation for everything. Words cannot describe my gratitude for their many life sacrifices. Daily, I felt their love and concern; their

constant prayers provided inspiration. Their Christian love and dedicated service reminds me of the sacrificial, loftier goals one needs to maintain in life.

Most of all to my wife, a special word of thanks for allowing me to pursue my academic interests. I depended on her extensive knowledge of computer hardware and software to get me through all kinds of situations. On occasion, she served as principal editor and critic of my work. Living separately was difficult. Thank goodness for Tobie, our little puppy, who kept me company. Without Valerie's patience, understanding and love, I would not have been able to complete this degree. Together, I trust we will greet the future and all that it holds.

No human undertaking is without Divine inspiration. I therefore dedicate this work to the glory of God. His unfailing guidance and intervention in my life significantly contributed to this minor accomplishment.

TABLE OF CONTENTS

ABSTRACT	ii
ACKNOWLEDGEMENTS	iv
TABLE OF CONTENTS	viii
LIST OF TABLES	xiv
LIST OF FIGURES	xvi
Chapter 1: INTRODUCTION	1
Conceptual Framework	2
Assumptions	4
Statement of the Problem	5
Purpose	7
Research Questions	7
Delimitations	8
Limitations	9
Definitions	10
Need for the Study	15
Organization of the Study	17
Chapter 2: REVIEW OF THE LITERATURE	18
Basis of Curriculum Alignment	19
Terminology	19
Unity	21
Platform Unity	22
Content Validity	24
Sameness of Process	25

Delivery	27
Implications of Curriculum Alignment	29
Conditions Not Conforming to Unity	30
Overtesting and Undertesting	31
Effectiveness and Efficiency	32
Misalignment	33
Planning	34
Goals	34
Objectives	36
Stating Instructional Objectives	38
Principles of Performance Instruction	41
Content Goals	43
Building an Exam	44
Taxonomy of Educational Objectives	45
The Cognitive Domain	46
Criticisms of the Classification Scheme	47
Simplifying Faculty Course Planning	48
Testing	50
Test Categories	51
Teacher-Constructed Tests	53
Classroom Testing	57
Purposes of Classroom Tests	58
Planning the Test	60
Table of Test Type Specifications	61
Selecting Appropriate Test Types	63

True-False Test Items	65
Matching Test Items	66
Multiple Choice Test Items	66
Completion Test Items	68
Short Answer Test Items	68
Essay Test Items	69
Summary	70
Chapter 3: RESEARCH DESIGN AND METHODOLOGY	76
Developing Formulae	77
Testing Formulae	78
Research Methodology	79
Sampling Frame And Survey Sample	80
Steps in Constructing Survey Instruments	80
Field Test of Instrument	81
Data Collection Procedures	83
Mode of Gathering Data	83
Receipt Control	84
Callbacks	86
Data Processing of Results	87
The Analysis Plan	87
Table of Test Type Specifications	88
Using Goodness of Fit To Test For Differences	89
Methodological Assumptions	91
Analytical Considerations	92
Content Level Sorts in the Cognitive Domain	92

Planning and Testing Values	96
Calculating Unity	97
Effectiveness and Efficiency	99
Summary	102
Chapter 4: RESEARCH FINDINGS	105
Questionnaire Responses	106
Phase I Instrument	106
Phase II Instrument	107
Respondent Demographics	108
Phase I Findings	112
Planning and Testing Decisions	114
What Are Faculty Test Type Choices?	121
Faculty and Expert System Test Type Differences	127
Analysis of Unity, Effectiveness, And Efficiency	129
Platform Unity	130
Overtesting Content	131
Undertesting Content	131
Effectiveness of Test Type Selection	134
Efficiency of Test Type Selection	134
Disunity	138
Phase I Summary	140
Phase II Findings	142
Agreement With Expert System	144
Disagreement With Expert System	145
Phase II Summary	148

Summary of Results	149
Chapter 5: CONCLUSION AND RECOMMENDATIONS	152
Study Overview	152
Summary of Findings	155
Demographics	155
Phase I Findings	155
Summary of Calculated Indices	156
Differences Between Expert System and Faculty Test Types	156
Phase II Findings	156
Purposes of Testing	157
Differences Between Phase I and Phase II Data	158
Conclusions	159
Recommendations	171
Summary of Recommendations	177
Researcher's Commentary	177
REFERENCES	180
APPENDICES	
A. Phase I Survey: Cover Letter, Instructions, and Instrument	191
B. Phase II Survey: Cover Letter, Instructions, and Instrument	200
C. Rate of Return Graphs (Phase I and II)	209
D. Survey Followup Letters (Phase I and II)	212
E. Participant Thank You Letter	215

F.	Questionnaires for Field-Test	217
G.	Virginia Community Colleges	230
H.	Formulae Calculation Spreadsheets	233
	For Level One Content	234
	For Level Two Content	238
	For Level Three Content	242
VITA	246

LIST OF TABLES

Table	Page
1 Inappropriate and Appropriate Planning Verbs	40
2 Table of Test Type Specifications	89
3 Cognitive Domain Content Goal Verbs	93
4 Performance Instruction Domain Level Codes	95
5 Test Types and Associated Point Values	96
6 Cognitive Test Item Types	97
7 Unity, Effectiveness and Efficiency Formulae	98
8 Effectiveness Matrix	100
9 Efficiency Matrix	101
10 Highest Academic Degree of Faculty	109
11 Academic Rank	109
12 Content Goals and Domain Level	113
13 Cognitive Test Item Types	114
14 Planning and Testing Decisions by Content	
Goal	115
15 Faculty Test Type Selections	118
16 Ranked Preference of Test Types	120
17 Faculty Selected Test Types To Test Facts	121
18 Faculty Selected Test Types To Test	
Understanding	122
19 Faculty Selected Test Types To Test	
Application	123

LIST OF TABLES CONTINUED

Table		Page
20	Differences Between Faculty and Expert System Test Type Selection	128
21	Range of Values Produced By Formulae	130
22	Unity, Effectiveness, and Efficiency Means	135
23	Unity, Effectiveness, and Efficiency Means For Levels of Cognitive Content	135
24	Unity, Effectiveness, and Efficiency Indices Per Content Goal	137
25	Disunited Planning and Testing Decisions	139
26	Faculty Responses To Expert System Recommendations	144
27	Respondents Agreeing With Expert System For Cognitive Domain Level	145
28	Reasons For Disagreeing With Expert System Test Types	146
29	Reasons For Disagreeing With Expert System Test Types For Factual, Understanding, and Application Content	147

LIST OF FIGURES

Figure		Page
1	Matrix of Scholar Contributions	72
2	Faculty Professional Development	111
3	Faculty Planning and Testing Decisions	116
4	Phase I Test Type Choices	119
5	Faculty Selected Test Types To Test Facts	124
6	Faculty Selected Test Types To Test Understanding	125
7	Faculty Selected Test Types To Test Application	126
8	Overtested and Undertested Content Per Content Goal	133

CHAPTER I

INTRODUCTION

Educators have always been concerned with measuring and evaluating the progress of their students. Yet, as the goals of education become more complex and as the accountability movement pervades more institutions, the task of measurement becomes more difficult. The accountability movement in higher education causes certain practices that deviate from conventional principles of instruction. Two such practices are overtesting and undertesting of course content. Another practice is measuring student achievement through standardized tests, a problem that can diminish faculty autonomy by controlling curriculum planning and delivery elements.

Studies on teacher-made tests are scant in comparison to studies on achievement tests. One purpose of testing is translating student performance to a letter grade. Faculty ignore measurement issues like reliability, validity, and item analysis when they use their own tests as indicators of student performance. At a more basic level, faculty may not even know if their tests are congruent with their plans (Vogler, 1991). Consequently, faculty may overtest or undertest course content. The type of tests they select may also be ineffective and inefficient.

A conventional curriculum principle, curriculum alignment, holds that the written curriculum ought to be the delivered curriculum as well as the evaluated curriculum. The

formation of an instructional program originates with written plans. Content delivery and testing become functions of content planning.

This descriptive study quantified and simplified faculty course planning and testing decisions. The extent to which these decisions were consistent was termed platform unity and quantified through a numerical index. In addition, four constructs, including, overtesting and undertesting of course content, and effectiveness and efficiency of test type selection, were analyzed and simplified through the unity concept.

CONCEPTUAL FRAMEWORK

This study has three primary anchors to ground it: (a) English's (1988) concept of curriculum alignment; (b) Gronlund's (1985) selection of appropriate test types used in instruction; and, (c) Vogler's (1991) classification of cognitive content planning and testing, presented through an outcome-based computer software package. The relationship between faculty content planning and testing decisions was termed platform unity and determined by comparing domain-referenced planning decisions with the types of tests used to evaluate student performance.

English (1988) stated that three elements must be present in any situation for it to be functional: (a) a standard or goal/objective, (b) work directed towards the standard or

goal/objective, and (c) feedback about the level of attainment or performance (p. 110). Sound curriculum management, according to English, ensures that the written curriculum is related to the taught curriculum, and both are related to the tested curriculum (p. 37).

When the outcome of instruction is to have students perform in a specified manner, Gronlund (1985) indicated there is a direct relationship among objectives, teaching, and testing. One way to ensure that these elements be in harmony is to prepare a planning chart that includes all three. This Table of Specifications provides an overall plan to assure that each objective receives the proper share of attention in the instructional process and that the methods of teaching and testing will be more relevant to the attainment of the desired learning outcomes in terms of what is planned (Gronlund, 1985).

In this study, the concepts of instructional planning and testing unity were analyzed through Instructional Performance System's PEAKS performance software, employing the Principles of Performance Instruction. Performance Instruction, as defined by Vogler (1991), consists of three phases: planning, delivery, and evaluation. The key to performance-based instruction is communicating to the learner the material addressed and the level at which it must be mastered (DeBrunner, DeBrunner, & Vogler, 1991).

The planning and evaluation of course content guided by the expert system embedded in the software employs a logic train prompting faculty to develop course plans that are delivered and tested at the same levels. The expert system within the software assures unity among planned, delivered and tested content. Following the Principles of Performance Instruction, the instructor records in advance the material to be covered. Specific decisions, such as identifying content to be taught, determining how it will be taught, when it will be taught, and how it will be evaluated, are made and documented with the assistance of the software.

Assuring unity in an outcome-based mode is a way to provide accountable instructional practices. Instead of relying on assessment testing to determine effectiveness, outcome-based planning through Performance Instruction is effective and efficient due to the content-driven, student-oriented system of documenting faculty course plans. Determining effectiveness and efficiency through the unity in instruction ensures that the content planned is also the content delivered and the content on which evaluations are constructed. The key to unity is the level at which faculty content planning decisions are congruent with the level at which content is delivered and tested.

ASSUMPTIONS

In order for this study to progress in an efficient

manner, several broad assumptions served as starting points.

1. Faculty possess subject matter expertise to make necessary planning, delivery and testing decisions. The decisions made by faculty regarding planning and testing as they applied to the course Introduction to Computers were used solely for analytical purposes. These decisions were not challenged as to their importance, inclusion, or overall contribution to student learning outcomes.

2. This study assumed the instructional content planned by faculty is delivered.

3. It was assumed that the Vogler Curriculum-Pedagogy-Assessment (C-P-A) model and Principles of Performance Instruction contained in the software's expert system have sound pedagogical underpinnings based on principles of sound instructional design used for planning, delivery and evaluation of course content by faculty. Further, the tenets established by Vogler (1991), replicated heuristically in the software, were assumed to be correct.

4. It was assumed that the course, Introduction to Computers, was the same course across all community colleges in the Virginia Community College System.

STATEMENT OF THE PROBLEM

Curricula exists in three, general forms: the planned, the delivered and the evaluated. Although these curricula are examined separately to determine quality, they are seldom

examined to determine unity.

Public concern about the effectiveness of education has led to a reliance on testing and test results. Until recently, test and measurement experts tended to limit their study of testing in the schools to the role of large-scale standardized testing programs, largely ignoring the nature or quality of teacher-developed classroom instruments (Stiggins & Bridgeford, 1985). Ebel (1979) indicated that it is difficult to make a single, standardized test suitable for students who have been taught by different teachers using different textbooks and learning materials in courses having different orientations.

Typical standardized tests consist of items that are intended to reflect important aspects of widely used curriculum materials. Alternatively, classroom tests are attempts to measure the impact of student learning in a particular course. These tests should measure clearly defined learning outcomes that are in harmony with the instructional objectives (Gronlund, 1982). Linn (1983) noted that no claim is made that classroom tests accomplish this.

Faculty at all levels of education too often instruct and evaluate without giving much attention to educational objectives. Often, objectives are undefined and vague (Ammons, 1969). This study involved examining faculty course planning decisions to see if they were consistent with their

testing decisions. The procedural problem was to develop and analyze a numeric platform unity formula.

PURPOSE

The general purpose of this study was to document the effects of applying the platform unity concept. This concept aligns planned content with appropriate test types. Based on the expert system in Vogler's (1991) Performance Instruction model, a table of test type specifications was created. This table provided a list of consistent test-types based on faculty-coded content planning decisions. From the list, test types were selected to structure content-based test questions, thus helping assure a consistent match or unity between course planning and testing.

To accomplish this documentation, five ancillary purposes were addressed:

- (a) synthesize the extant literature,
- (b) create an expert system table of test type specifications,
- (c) analyze course planning and testing decisions,
- (d) classify unity dimensions, and
- (e) apply unity to planning and testing decisions.

RESEARCH QUESTIONS

The following research questions were addressed in this study.

1. Chapter Two consists of a literature review addressing

several supporting research questions. The research questions include: (a) How is alignment related to unity?, (b) What is overtesting and undertesting?, (c) What is effective and efficient testing?, (d) What is the rationale to support domain level planning decisions as predictors of test item types?, and, (e) What is the connection among taxonomy classifications?

2. Is there a difference between expert system and faculty test type selection? The results of the analysis to this question appear in Chapter Four.

3. How can a unity index be used? Calculated indices of unity, effectiveness and efficiency appear in Chapter Four. The implications are discussed in Chapter Five.

4. What test types do faculty use to test facts, understanding, and application? The findings are presented in Chapter Four.

DELIMITATIONS

The following delimitations were made to narrow the focus of this study.

1. Although curriculum consists of the planned, the delivered, and the evaluated, in this study, platform unity was analyzed by delimiting the curriculum to include the planned and the evaluated curriculum.

2. The planned curriculum in this study was delimited to include only cognitive content.

3. The tested curriculum in this study was delimited to teacher-developed classroom tests, rather than a standardized final exam.

LIMITATIONS

The following limiting factors may affect the internal and external validity of the study.

1. The study was limited to analyzing planning and testing decisions within the context of formal instruction. As this study used formal curriculum constructs, such as course planning, testing, effectiveness and efficiency, it was less likely to be biased toward interpretation due to the documentation of course plans. It was also less likely to capture informal instruction or what might appear as the "hidden curriculum." The drawback of using documented decisions was that it directs attention to the formal curriculum and not on the presentation made by the teacher.

2. In developing a method of testing the platform unity concept, responses from faculty were limited to six test item types used to test content. No other judgment was made about differences in mission, curriculum or institution.

3. Although there were various functions of classroom tests, such as using test results to provide student feedback, to determine student rank and placement, etc., this study was limited to examining unity among planned, instructional content goals and test items. The results must be cautiously

generalized to the body of literature on testing and measurement.

4. As presented in this study, the concept of platform unity was limited. Although there may be unity in test-type selection, there is no assurance that written test items contain reliable and valid indicators of student performance.

5. The terms overtesting, undertesting, effectiveness, and efficiency were operational constructs developed and designed for specific application in this study. Use of these terms beyond this study is therefore limited.

DEFINITIONS

The following definitions refer to terms as they were used in this study.

Accountability - a condition in which individual role holders are liable to review and the application of sanctions if their actions fail to satisfy those with whom they are in an accountable relationship (Kogan, 1986).

Alignment - the congruence of all elements of a curriculum, including goals, the instructional program, and the tests used to judge outcomes (Crowell & Tissot, 1986).

Behaviorial objective - a statement of what a student will learn in terms of the behavior that will be exhibited when learning occurs, the condition under which the behavior will be exhibited, and the criteria to be used in evaluating the behavior (Salvia & Hughes, 1990).

Cognitive - a domain of instruction that is knowledge or mind based (Vogler, 1991).

Cognitive content levels - Level one, (fact level), used with a single concept; level two, (application level), combining two or more concepts; level three, (understanding level), combining two or more concepts to form something new (Vogler, 1991).

Content goal - a statement that communicates learner orientation, specifies tangible product, includes one present tense action verb, aligns with learning exit point, conforms to three hours learning time, ends with direct object, and uses few parts of speech (Vogler, 1991).

Criterion-referenced - the comparison of a measurement with some objective standard (Mager, 1973, p. 9).

Curriculum - (1) all written or intended, academic and nonacademic instructional objectives for a student or group of students (Salvia & Hughes, 1990); (2) that output of a curriculum development process intended for use in planning instruction (Schiro, 1978, p.28).

Educational program evaluation - the process of delineating, obtaining, and providing useful information for judging decision alternatives (Stufflebeam, Foley, Gephart, Guba, Hammond, Merriman, & Provus, 1971).

Effectiveness - selecting test item types that match or exceed the planned content level. These test items require

more complex student responses. (See Table 2, p. 89.)

Efficiency - selecting test item types that match or fail to reach the planned content level. It represents the easiest way of soliciting a student response. (See Table 2, p. 89.)

Evaluation - (1) the process of delineating, obtaining, and providing useful information for judging decision alternatives (Stufflebeam, et al, 1971). (2) the systematic process of determining the extent to which instructional objectives are achieved by students (Gronlund, 1981).

External test - tests imposed by authorities from outside a particular instructional system (Nitko, 1989, p. 447).

Grade - a label representing the outcome of an evaluation (Mager, 1973, p. 9).

Instructional objective - a written statement describing what a student will learn (Salvia & Hughes, 1990).

Item - a single question or exercise in a test (Payne, 1992).

Matching item - a test item calling for the correct association of each entry in one list with an entry in a second list (Payne, 1992).

Measurement - a process of determining the extent of some characteristic associated with an object or person (Mager, 1973, p. 8). It is concerned with the systematic collection, quantification, and ordering of information (Payne, 1992) about student behavior (e.g., test scores) (Gronlund, 1981).

Norm-referenced - comparing performance of students with reference to each other (Mager, 1973, p. 9).

Objective test - a test for which there is only one best answer (Vogler, 1991).

Objective-referenced test - a test with which student performance is evaluated in terms of the specific objectives attained when the objectives are derived from individual test items (Salvia & Hughes, 1990).

Overtesting - faculty selection of test item types that test content at a higher level than the domain level on which content plans were established.

PEAKSolutions software - curriculum development software created by Daniel E. Vogler. The complete package includes three integrated components: CourseBuilding, for course development; LessonBuilding, for lesson plan creation; and ExamBuilding, for test construction and evaluation (Vogler, 1991).

PEAKS suite - the combination of all three PEAKSolutions software packages, each containing the Vogler C-P-A model, for use by an institution.

Performance objective - a type of course objective that contains outcome statements, measurable standards, and a condition statement (Vogler, 1989).

Planning instruction - the selection and sequence of activities to facilitate learning (Salvia & Hughes, 1990).

Platform unity - the specific alignment of planned, domain-referenced content with associated test types.

Student assessment - the evaluation of cognitive knowledge, specific abilities, designated skills, and attitudes of individual students (Deutsch, 1991).

Test - an event during which someone is asked to demonstrate some aspect of his or her knowledge or skill (Mager, 1973, p. 8).

Testing - any systematic procedure for observing and classifying students' performance for purposes of obtaining instructionally relevant information (Nitko, 1989, p. 447).

Test item - a question that calls for a single response or set of responses to a single stimulus or stimulus pattern (Mager, 1973, p. 7).

Undertesting - faculty selection of test item types that test content at a lower level than the domain level on which content plans were established.

Unity - (see also platform unity) the specific alignment of planned, domain-referenced content with associated test types.

Validity - the extent to which the results of an evaluation procedure serve the particular uses for which they are intended (Gronlund, 1981).

Vogler C-P-A Model - (Curriculum-Pedagogy-Assessment) -a method of "performance instruction" that results in a

communicative, content-driven course. It utilizes PEAKSolution software (Vogler, 1991).

NEED FOR THE STUDY

Specific studies dealing with planning and testing are scant, at best. Traditionally, studies on alignment have focused on broad curriculum issues, such as making sure centrally-developed objectives are taught in the program (Crowell & Tissot, 1986). These studies are typically descriptive; for example, knowing the degree of unity or disunity is touted as "desirable" (Savard & Cotton, 1982). There is no mention of a quantitative index for faculty or administrators to conveniently obtain information on the degree of unity between planning and testing.

On the other hand, there are numerous studies dealing with testing and measurement. Nearly all reports on achievement measurement have focused on topics more relevant to large-scale standardized tests (Stiggins & Bridgeford, 1985). Teacher-made tests have yet to be examined in depth (Lazar-Morris, Polin, May, & Barry, 1980; Stiggins, Conklin & Bridgeford, 1986). Nitko (1989) found that traditional approaches to disciplined inquiry in test design have tended to focus on optimizing the measurement efficiency of tests rather than on optimizing their instructional efficacy: "The testing procedures used in the instructional process must measure the student behaviors and cognitive processes that

have been stated as desired outcomes of instruction" (p. 453). This coherence is central in English's (1986-87, 1988, 1989, 1992a, 1992b) concept of alignment.

Savard and Cotton (1982) saw curriculum alignment as "new". They stated rather convincingly:

"This topic of curriculum alignment is important. It is a new topic--but it sounds like an old, common sense idea--especially to someone outside the educational profession. While the idea of curriculum alignment is simple, its implementation as an educational practice is not easy--yet, the potential it holds for improving the quality and effectiveness of education demands that we address the topic" (1982, p. i).

An ERIC search revealed only one citation when "curriculum-alignment" was cross referenced with "research" (Savard & Cotton, 1982). During a recent search, the researcher noticed two citations when crossreferencing the same two descriptors from a 1966 - 1981 ERIC database and 53 citations using a 1982 - September, 1992 ERIC database. Despite the increased number of citations, Crowell and Tissot (1986) cited a general lack of attention to alignment. They concluded "very little practical research exists to guide the efforts of local school districts" (p. 2). "For curriculum alignment to be achieved, there must be more widespread discussion and demonstration of the importance of alignment" (Crowell & Tissot, 1986, pp. 6-7). They advocated the implementation of practical procedures to aid in understanding

and determining alignment (p. 7).

ORGANIZATION OF THE STUDY

This study is presented in five chapters. The first chapter contains a proposal of the study. Chapter Two contains a review of the literature related to faculty course planning and evaluation. The theoretical underpinnings of curriculum alignment and Principles of Performance Instruction are detailed in this section.

The research design and methodology is presented in Chapter Three. The instruments and means of analysis used in the study for data collection and assessment are described.

The results and analysis of the data are presented in Chapter Four. Factors presented in the review of the literature were supported with the findings of this study. The main research question--Is there a difference between expert system and faculty test type selection--is answered in Chapter Four. In addition, data show how the unity index can be used to determine alignment or misalignment of content planning and testing decisions.

Chapter Five presents an overview of the study, the findings of the study, conclusions drawn from the study, and recommendations for future study.

CHAPTER II

REVIEW OF THE LITERATURE

This research study concerned the unity of planned and tested curriculum. The unity construct, in its purest form, aligns with the Vogler Curriculum-Pedagogy-Assessment (C-P-A) model, the concept on which Principles of Performance Instruction is built. Performance Instruction holds that planning, delivery and evaluation of course content must be consistent for instructional outcomes to be effective and efficient.

The organization of the literature review occurs around the three anchors of the study. Curriculum alignment is reviewed first. This establishes the terminology used in the study. Performance Instruction is reviewed next. Topics related to planning instruction are reviewed. Finally, literature relating to testing is reviewed. This section ties together the concepts of planning and testing as they relate to principles of curriculum alignment, as defined by a table of test type specifications.

This chapter addresses five research questions. These questions are: (a) how is alignment related to unity? (b) what is overtesting and undertesting? (c) what is effective and efficient testing? (d) what is the rationale to support domain level planning decisions as predictors of test item

types? and (e) what is the connection among taxonomy classifications?

The literature review is presented in eight major sections: (a) Basis of Curriculum Alignment, (b) Implications of Curriculum Alignment, (c) Planning, (d) Principles of Performance Instruction, (e) Taxonomy of Educational Objectives, (f) Testing, (g) Table of Test Type Specifications, and (h) Selecting Appropriate Test Types.

Basis of Curriculum Alignment

Curriculum alignment is a term used to denote the conscious congruence of three educational elements: curriculum, instruction, and assessment (Savard & Cotton, 1982). Scott (1983) believed student learning is most successful when these three curricular elements come together. To align the curriculum, faculty first must determine what they intend to teach, then they teach it. Finally, they test what they have taught through different measurement methods. Savard and Cotton (1982) stated that alignment is something new, and its implementation as an educational practice is not widespread.

Terminology

Various terms are used in the literature to describe the condition whereby three components of the curriculum, the written, the delivered, and the evaluated, have explicit relationships to one another. Among these terms are

congruence (Brickell, 1976), overlap (Leinhardt, 1983; Leinhardt & Seewald, 1981), curriculum test overlap (Hartzell, 1984), match, linking (Airasian & Madaus, 1983), and alignment (English, 1986-1987). Considering these terms, curriculum alignment, or just simply alignment, is most closely related to the purpose of this study.

There are variations of the term, "curriculum alignment," mentioned in the literature. Although Crowell and Tissot (1986) claimed the Southwest Regional Educational Laboratory (SWRL) was credited with coining the term, the most notable advocate of curriculum alignment has been Fenwick W. English. For English, alignment consists of the written, the taught, and the tested curriculum. In his National Academy for School Executives, English (1992a) developed a thorough inquiry into alignment elements. Among terms he used to describe the written element are "curriculum" and "the work plan". Terms used to describe the taught element include "delivered" and "the work". The tested element included such terms as "the measured," "evaluated," and "the work measurement".

In developing an all-encompassing definition Crowell and Tissot (1986) referred to alignment as the congruence of all elements of a school curriculum, including the curriculum goals, the instructional program, and the tests used to judge outcomes. Savard and Cotton (1982) considered alignment to be the congruence between state curriculum guidelines and

locally-developed instructional practices. Although the terminology used is similar, it may be confusing as a research construct. In this study, curriculum alignment was defined as congruence between the planned, the delivered, and the tested curricula. Analytic constructs used in this study, including platform unity, overtesting, undertesting, effectiveness and efficiency, were all based on the alignment among the planned, the delivered, and tested curriculum.

Research Question A: How is alignment related to unity?

The next two sections address this question. There are not differences between alignment and unity. They both refer to the congruence of curriculum elements. In this study the term platform unity was introduced to specifically refer to the planned and tested elements of the curriculum.

Unity

A central construct in this study, unity was first mentioned by English (1979) to represent oneness between the written, the taught and the tested parts of the curriculum. In all his work, English (1979, 1986-1987, 1988, 1992a, 1992b) used a triangle to depict this concept and it became the universal symbol to represent alignment in the literature. Savard and Cotton (1982) referred to it as the "alignment triangle," with objectives, instruction, and testing forming its corners.

Although English's concept of alignment was developed

primarily for use in secondary school management through a process he termed an audit (English, 1988), it is generic to instruction at all levels. For instance, Pautler (1989, 1990, 1992) advocated using alignment in community colleges. He believed that all institutions, regardless of level, are concerned with the improvement of instruction. According to Pautler:

"curriculum alignment should be in the vocabulary of community college faculty and administrators. In this day and age of so many part-time instructors being employed by community colleges, the need for alignment is even more critical. The process of alignment is directly related to the development and use of a functional course of study" (1989, p. 178).

Platform Unity

The three elements of the curriculum were placed at the three points on the triangle. Along one plane, or platform, of the triangle is the relationship between the planned and tested curriculum; hence the origination of the term, "platform unity." This term was selected to simplify the analysis of these two constructs. In addition to the planning-testing platform, one finds a planning-delivery platform and a testing-delivery platform.

The reasonable place to begin a study of alignment is along the planning and testing platform. The other two platforms involve delivery that is difficult to measure quantitatively. This study examined the relationship between

the written and the tested curriculum, delimiting the delivered or taught curriculum. Specifically, the researcher examined platform unity between the planned, domain-referenced content and associated test types.

English (1992a) distinguished two types of curriculum alignment, design alignment and delivery alignment. Design alignment is the relationship between the curriculum and the test. Delivery alignment is the relationship of what is taught to the test and to the curriculum. Platform unity most closely resembles English's concept of design alignment.

Platform unity is a curriculum-based estimate of congruence between planning and testing. It includes four additional constructs: overtesting, undertesting, effectiveness and efficiency. However, Leinhardt and Seewald (1981) distinguished a curriculum-based approach of analysis from an instruction-based approach. A curriculum-based approach includes written documentation; it excludes information covered by faculty.

The conditions of content validity and sameness of process must be met before platform unity is analyzed. In this study, content validity was addressed by delimiting to teacher-made tests. Sameness of process was addressed through the assumption that the expert system in Vogler's model establishes sameness of process or cognitive symmetry between faculty planning and testing decisions.

Content Validity

The validity of a measure is how well it fulfills the function for which it is being used (Stanley & Hopkins, 1972). If a measure lacks validity, the information it provides is useless. Content validity is particularly relevant to this study because it is based on how adequately it samples a domain of behavior or knowledge about which an inference is based (Jacobs & Chase, 1992). Content validation, unlike "face" validity, requires systematic, logical analysis. Typically, a content-valid test also has face validity (Ebel, 1965). Tests should have face validity, otherwise students may feel they are being unfairly assessed.

For a test to have high content validity it should be a representative sample of both the topics and the cognitive processes of a given course or unit (Stanley & Hopkins, 1972). Cureton (1969) believed a test should possess topic and process validity. Tuckman (1985) said a test is appropriate, and its content is valid, when it represents the skills, competencies, aims and objectives established in the planning stages. Content validation is a logical process. Stanley and Hopkins (1972) believed that examining test items in relation to the objectives and instruction requires the following judgments:

1. Does the test content parallel the curricular objectives in content and process?

2. Are test and curricular emphases in proper balance?
3. Is the test free from prerequisites that are irrelevant or incidental to the present measurement task (p. 102)?

It is important to note that this study was delimited to teacher-made, classroom tests. Stanley and Hopkins (1972) stated, "it is impractical to conduct an extensive content-validity analysis on most or even all teacher-made tests" (p. 102). Instead, they recommended the instructor systematically examine each classroom test before its administration to assure a good sampling from the content universe. An instructor should assess whether important instructional topics represented in the exam are in correct proportion. Stanley and Hopkins (1972) believed such questions, if raised systematically by all instructors at all educational levels, would greatly improve the quality of teacher-made tests.

In this study, content validity was enhanced by using actual course content goals from an introductory computer literacy course at a local community college. These content goals were developed through a task analysis by faculty and practitioners in the computer field. Therefore, since this study was delimited to teacher-made tests, content validity of tests was assumed.

Sameness of Process

Gagne (1975) called learning a process comparable to

other human organic processes such as digestion and respiration. Learning is an enormously intricate and complex process. Cognitive strategies govern the learner's own thinking, remembering and learning behavior (Gagne & Briggs, 1979; Gagne, 1975). They represent ways the learner manages the processes of learning.

In educational systems, cognitive strategies are of equal importance to content goals that are statements of educational outcomes (Gagne & Briggs, 1979). Learning objectives are derived from learning outcomes (Gagne, 1975). The most essential part of the objective statement, the outcome performance, follows the use of an appropriate outcome verb (Diamond, 1989; Gagne, 1975; Mager, 1973; Vogler, 1991). Learning objectives provide a view of learning from the back forward. The main reason for this view is to keep firmly in mind what the ends of learning are (Gagne, 1975).

Gagne (1975) observed that the educational literature posits cognitive strategies as desirable learning outcomes. For example, learning to learn and learning to think are often cited as educational goals of high priority. However, measuring the attainment of such highly regarded goals is ambiguous. Ambiguity can be avoided when courses are described in terms of objectives (Baker & Popham, 1970; Diamond, 1989; Mager, 1975; Popham, 1973, 1978; Vogler, 1991).

Basing this study on Vogler's (1991) Performance

Instruction model inherently ties together the planning, delivery and evaluation of course content. His recommended use of the content goal verb assures unity of process among the three elements of instruction. The present tense action verb used in planning instruction unites the process from delivery to evaluation, thus promoting sameness of process or instructional symmetry.

After the outcome is decided, Vogler (1991) recommended setting the cognitive standard to measure content attainment. In Performance Instruction, this standard becomes part of the instructional objective by changing the planning verb to a noun. Since the content goal verb is domain-referenced, cognitive symmetry is maintained through the delivery and testing phases of instruction. Setting the standard assures the process is consistent, from planning the content to testing it.

Delivery

As previously noted, this study was delimited to include the written and evaluated curricula. Consequently, these two aspects are reviewed more thoroughly in this chapter. Nonetheless, below is a brief examination into the relationship between teaching and testing to provide a context for this discussion.

Several researchers focused their attention on the relationship between instruction and testing (Airasian &

Madaus, 1983; Baker & Herman, 1981; Leinhardt & Seewald, 1981; Linn, 1981). Leinhardt and Seewald (1981) used the term, "overlap" to mean the extent to which the content of what is taught matches the content of the test. Although the domain of instruction may be identical to the domain sampled by the test, it may partially overlap or share the domain of the test, or it may be totally different. When a set of test scores is used to help evaluate the impact of instructional programs, knowledge about the extent of overlap is critical to interpretation of the results (Leinhardt & Seewald, 1981).

Classroom instruction varies according to time on task, timing of instruction, quality of instruction, reinforcement, sequencing, emphasis, and other such factors. Testing methods include format, materials, organizations, administration methods and procedures, scoring procedures, and specific uses of results (Nitko, 1989). Achievement test results are a consequence of instructional variation (Airasian & Madaus, 1983). To properly investigate linkage between instruction and testing, Airasian and Madaus favored Bickell's (1976) claim that it was necessary to observe teachers in practice.

In making judgments about school performance, Airasian and Madaus (1983) reported that items distinguishing schools and programs are as likely to come from criterion-referenced tests as from norm-referenced tests. However, criterion-referenced tests tend to be constructed without the benefit of

empirical tryout and item-analysis procedures. Airasian and Madaus pointed to the apparent conflicts in linking instruction with testing in standardized, norm-referenced situations. With standardized testing, individuals who might or might not be associated with a particular instructional effort apply different social criteria for judging the importance of content. Because there are no uniform standards of importance and because many standardized examinations are created by individuals unfamiliar with the specific day-to-day operations of instruction, students are receiving what is judged to be important. Thus, a particular test may or may not match the content coverage and emphasis delivered to students (Nitko, 1989). Since one cannot assume congruence between the objectives used to produce a commercially developed standardized test and the elements of what a teacher plans to teach, analysis of linkage is difficult to measure (Airasian & Madaus, 1983).

Implications of Curriculum Alignment

English (1987) believed curriculum alignment begins with the written curriculum. English and Steffy (1983) called the written curriculum a "prescription" or a set of specifications. Ideally, it comes first in a sequence of events that also includes delivery and evaluation.

English and Steffy (1983) observed that problems in developing the written curriculum are usually design problems

that stem from errors made in planning the curriculum. For example, a design problem occurs when what is planned is not linked to test construction. Consequently, students are tested on skills or knowledge for which they have not been taught. A faulty design, no matter how capably or uniformly implemented, prevents optimization until it is corrected so that plans and tests are congruent (English & Steffy, 1983).

Pautler (1990) believed that developing written curriculum documentation of courses of study may be one of the least understood activities within community colleges. According to Pautler (1990), most college-level courses lack a consistent format. The design of functional courses of study parallels curriculum alignment as a worthwhile effort to improve instructional effectiveness of the college (Pautler, 1990). Pautler (1990) advocated instructional design principles to develop courses of study. He believed if community college staff were given more preparation in instructional design principles, then improved institutional effectiveness would result. Therefore, alignment should be an explicit part of the faculty development process and be implemented in the instructional design process (Pautler, 1990).

Conditions Not Conforming to Unity

In a situation where the type of test selected to test content does not match the level at which course content was

planned, platform unity will not be absolute. The level of planning serves as the basis for comparison. In this study, content tested above the level at which content was planned creates a situation termed overtesting. Content tested below the level at which the content was planned creates a situation termed undertesting.

Research Question B: What is overtesting and undertesting?

Overtesting and Undertesting

Overtesting or undertesting of course content occurs when there is not absolute unity between the level of content planned and the level of content tested. In situations where instructors elect to overtest or undertest course content, the purpose of the test has to be examined. It is perfectly legitimate to overtest or undertest content as long as the purpose of the test is upheld. The purposes of testing are explored later in this chapter.

Bloom, Madaus, and Hastings (1975) favored the unity concept. They believed material should **not** be tested at higher levels than at which it is taught. The unity principle holds that if instructional objectives are taught at the knowledge level, faculty should neither test nor expect students to know the material at a higher level. However, instructors do not always test at the level at which content was planned. An inquiry into the effectiveness and efficiency

of test type selection is beneficial.

Research Question C: What is effective and efficient testing?

Effectiveness and Efficiency

Effectiveness and efficiency refer to faculty selection of specific test types to test planned content. Indices of effectiveness and efficiency are calculated separately. One finds the selected test type effective when overtesting results. The inverse is also true: one finds the selected test type efficient when undertesting results.

Effective test types are those that a faculty member uses to thoroughly test student knowledge of concepts. An effective test type requires the student to respond to test questions created at a level which equals or exceeds the level at which content plans were written. For example, a multiple-choice test is an effective selection if faculty want to test level one, level two or level three planned content.

Test type selection is efficient when faculty select a test type that easily solicits responses from students. A test is not efficient if factual level content were tested by issuing an essay examination. It is efficient if faculty test application content, the highest level of planned content, through one or more of the six test types identified in this study.

Misalignment

Misalignment is a condition when content plans and testing decisions are not congruent. Another term for this is disunity. Savard and Cotton (1982) observed that there are some individuals who claim that conscious alignment efforts restrict spontaneity. Alignment is rooted in the belief that the success of instructional goals are determined by establishing an outcome-based content goal when planning a course and assuring that the delivery and assessment are all congruent. However, there are conditions where misalignment is legitimate. Such situations relate to the purpose of testing.

A way to see the importance of curriculum alignment is to view what happens when misalignment occurs. Most students can tell stories about instructors whose lectures did not follow the course syllabus and whose final exam had nothing to do with either (Pautler, 1989). Nitko (1989) stated that tests created by classroom teachers often inadequately measure the teachers' espoused instructional goals. When misalignment occurs, the written, taught, and tested curriculum function independently. The result is that what faculty plan exists apart from what students encounter on tests (Savard and Cotton, 1982). According to Nitko (1989), there are several undesirable consequences of using tests that are inadequately linked to instructional plans. These include the following

four:

"(a) teachers and students might be inappropriately informed about students' learning progress and learning difficulties;

(b) students' motivation for learning could be reduced;

(c) critical decisions about students (e.g., whether to award a student a degree) might be made unfairly; and

(d) the effectiveness of instruction may be evaluated incorrectly" (p. 447).

Systematic planning, delivery and evaluation of course content is a way to uphold the principles of curriculum alignment.

Planning

Planning instruction is a necessary faculty activity, though it is rarely given proper attention. According to Cross (1987), the processes involved in classroom teaching in higher education should be taken much more seriously than they have been in the past. Recent advances in cognitive psychology indicate that students learn better if teachers clearly communicate course goals, objectives, and discipline structure (Stark, Lowther, Sossen & Shaw, 1991).

Goals

Establishing goals for instruction is a common method of planning. These goals are the philosophic intentions of education. Goal-referenced instructional models focus on desired, observable learner behavior produced as the result of

instruction (Baker & Popham, 1970). Baker and Popham frame the basic planning decision. They suggest, instead of asking what shall I do, the instructor needs to ask, what do I want my learners to become. By introducing this question, the emphasis shifts to a student-oriented planning model. There is a shift occurring in education away from course-centered and content-centered approaches of course development to more learner-centered approaches (Menges & Mathis, 1988, p. 14).

Diamond (1989) believed the process of determining whether or not academic programs are successful begins by stating what is expected of students. This requires describing goals and objectives in performance terms and creating evaluation instruments that assess the abilities of students to meet specific criteria. Developing a course plan that optimizes each aspect of instruction is a challenging faculty activity requiring expertise and informed decision-making.

Educational goals must be stated in more precise and observable form to give direction to the important tasks of curriculum development and evaluation (Stanley & Hopkins, 1972). According to Cook (1978), a clear, useful, verifiable statement of student learning outcomes should contain at least three ingredients. First, it should specify the learner; second, it should describe an unambiguous and observable action; third, it should specify the conditions under which

these outcomes will be assessed. Others believe it should contain a description of the minimal level of acceptable response (Gagne, 1975; Gerhard, 1971; Mager, 1975; Popham, 1973).

Objectives

While goals communicate curricular intent, objectives are communication devices for students (Vogler, 1991). Objectives provide instructors with a method of specifying instructional goals in a curriculum. They are, by definition, intended outcomes of instruction. At the course level, objectives provide a framework and a guide for the specific instructional decisions that follow (Gunter, Estes & Schwab, 1990). Vogler (1991) stated that "objectives describe student outcomes, the standards used to measure outcomes, and the conditions under which the outcomes will be produced" (p. 29). The terms "behaviorial", "performance", and "measurable" objectives have been used interchangeably to denote the end behaviors desired as a result of instruction (Gaff, 1975). Course objectives written in behavioral, observable and measurable terms state precisely what students will be able to do as they interact with specific course content; they specify what students will do or produce when the instructional goal has been met or after completing the instructional unit (Bloom, Madaus & Hastings, 1981; Copperud, 1979; Gerhard, 1971). These statements also describe how behavior will be measured

(Gerhard, 1971). In the planning process, faculty take their plans and make a communication statement to students, usually in the form of a course syllabus, which contains the course objectives. The syllabus paves the way to student acceptance and understanding of the instructional system (Kibler, Cegala, Watson, Barker, & Miles, 1981).

There has been considerable professional attention directed toward the nature of objectives, the way they should be formulated, and the functions they serve in the design of instructional sequences (Baker & Popham, 1970; Diamond, 1989; Eisner, 1969; Gunter, Estes & Schwab, 1990; Kibler, Cegala, Watson, Barker, & Miles, 1981; Mager, 1962, 1973, 1975; Popham, 1973, 1978; Salvia & Hughes, 1990; Vogler 1991). Although the professional literature recommended that objectives should describe precisely what is expected of students, Diamond (1989) believed most faculty conceptualize their teaching in terms of their content area rather than with reference to student outcomes (p. 127). Although they strive to plan and teach courses in ways that help students learn effectively, faculty members seldom have received specific training for these tasks (Stark, Lowther, Sossen, Shaw, 1991).

When planning a course, it is necessary to plan instructional outcomes, not instructional activities. It is important to focus behavioral objectives on student performance from a student's point of view rather than on

teacher performance (Feldhausen, 1980; Salvia & Hughes, 1990). For example, reading Plato's Republic and discussing portions of it in class describes a planned activity. It does not indicate what students are supposed to gain from the activity. It is more appropriate to state the learning task in behavioral terms, such as the identification of major features of Plato's theory.

In stating objectives, teachers "seek to clarify in their own mind and to communicate to others the sought-for change in thoughts, actions, or feelings that a particular unit or educational program should help a student realize" (Bloom, Madaus & Hastings, 1981, p. 17). Mager (1975) and Diamond (1989) posited that a complete objective has three elements: (a) an action or behavior that the student is to perform; (b) the conditions under which these actions are to be performed; and (c) the criteria or standards for saying whether these actions have been performed satisfactorily. The most important element is the behavior (Mager, 1975).

Stating Instructional Objectives

In preparing a list of instructional objectives for a course of study, there are two immediate goals in mind. One is to obtain as complete a list of objectives as possible. The other goal is to state objectives so that they clearly indicate the learning outcomes that one expects from instruction (Gronlund, 1981, p. 45). Statements of specific

learning clearly convey instructional intent if each statement begins with an action verb that indicates definite observable responses (Copperud, 1971; Gronlund, 1981, 1985; Tuckman, 1985; Vogler, 1991). Verbs that have discrete, definite meanings should be used rather than those open to many interpretations (Stanley & Hopkins, 1972). For example, the words "to identify", "to solve", or "to construct" are preferable to "to understand" and "to appreciate" (Mager, 1962).

Table 1 provides several other examples of appropriate and inappropriate verbs to use in planning instruction.

Table 1

Inappropriate and Appropriate Planning Verbs	
Inappropriate Verbs Open To <u>Many Interpretations</u>	Appropriate Verbs Open To <u>Few Interpretations</u>
to know	to write
to understand	to recite
to appreciate	to identify
to enjoy	to sort
to believe	to construct
to internalize	to compare
to improve	to define
to comprehend	to name
to learn	to classify
to think	to predict

Research Question D: What is the rationale to support domain level planning decisions as predictors of test item

types? This research question covers a broad area that includes both planning and testing. Essentially, decisions made when planning content establish test item types. If alignment is upheld, then test item types are selected from the same domain level at which planning occurred. English (1992b) used a concept called "frontloading" to assure that content plans precede content testing.

The first step in a systematic approach to instruction is specifying objectives in operational terms (Baker & Popham, 1970). To formulate useful statements, educators must choose words that are interpreted in the same sense by readers (Bloom, Madaus & Hastings, 1981). Objectives that are tied to performance levels provide a basis for evaluating a range of performance (Diamond, 1989). English and Steffy (1983) stated that objectives used in planning instruction should focus on the knowledge, skills and attitudes a student should master, lend themselves to measurement, and reflect research knowledge about the subject.

Cognitive objectives specify that learners will do something with knowledge. Gunter, Estes, and Schwab (1990) specified two different kinds of cognitive objectives, declarative and procedural. Declarative objectives represent knowledge that can be expressed in true or false statements. Procedural knowledge results from the reasoning process. The most common instructional objectives in postsecondary

education focus on information recall (Babbin, 1987).

Gaff (1975) stressed the most effective courses are those in which the objectives, learning experiences, and evaluation of students are consistent. A model can be used to help bring together faculty planning and testing decisions. The model advocated in this study for assuring this unity is Performance Instruction.

Principles of Performance Instruction

The conceptual model on which this study is based is Vogler's (1991) Performance Instruction. Performance Instruction is the planning, delivery, and evaluation of learning and teaching. It is based upon the Vogler Curriculum-Pedagogy-Assessment (C-P-A) model (Vogler, 1991). The model is the foundation of the expert system for three independent or integrated software packages. The model works without the software; however, the software guides, monitors and aggregates user decisions. The information presented below briefly explains some of the features of the expert system in the curriculum model.

The expert system in the software tutors the user in correct course design, lesson planning, and the creation of test items and criterion referenced examinations. It provides a standard that promotes consistency within the curriculum. The system is dynamic by providing the infrastructure to develop some functional output--a syllabus, a lesson plan, or

an exam. On the other hand, the expert system is passive by providing a prescriptive manuscript that can be accessed by the user. The software flexibility is enhanced by allowing the user to override default-based decisions.

The expert system in Performance Instruction guides, monitors, and aggregates the decisions of an instructor in planning, delivering and evaluating instruction. The functions of the expert system are simply literature-based, approved practices that have been harnessed by the modern technology of a computer.

Performance Instruction holds that content included in a course should be planned, delivered, and evaluated. The integration of these elements is often neglected in traditional approaches to instruction. It is not uncommon to find no or scant planning documentation, an unfollowed syllabus, or testing that goes beyond or does not relate to what was taught.

The C-P-A model replicated in the expert system is based on taxonomy classifications. Vogler amalgamated the works of Bloom, Englehart, Furst, Hill, and Krathwohl (1956), Krathwohl, Bloom, and Masia (1964), and Simpson (1966) to produce simplified levels in three domain classifications: the cognitive, the psychomotor, and the affective. The expert system relies on a syntax for writing content goals that helps the user manipulate simple decisions to form syllabi with

internal consistency and external appeal as communicative devices.

Content Goals

Careful selection of content goal action verbs produces content communication across the planning, delivery, and evaluation phases of instruction. In Performance Instruction, a present tense verb such as "define" is used in the planning phase. The verb can be changed to the gerund form "defining" to serve as the instructional topic in delivering the instruction. Finally, the same verb can be changed to the singular form "defines" and the evaluating paradigm is created.

Content goals are the communication devices that specify the outcomes of Performance Instruction. The syntax of a content goal is controlled to maximize clarity and to provide the basis for aggregating information. The final format of a content goal is a sentence five to nine words in length. For example, a content goal from an introductory economics course might appear as follows: The student will identify equilibrium points.

Here, the present tense action verb "identify" is linked to a direct object that may be modified by up to four adjectives. The verb will later be classified as cognitive and at a factual level. The choice of the verb, object and adjectives sets the stage for delivery and evaluation.

Through guided steps in Performance Instruction, the user identifies the domain as either cognitive, psychomotor or affect. Within each domain, there are three levels moving from simple to complex. Content with a higher level will require more time to teach and to learn. Domain and level are linked to the verb within the content goal. As a consequence, the expert system recommends the decision for the user. The decisions made at this stage affect delivery and evaluation decisions.

Building an Exam

Performance Instruction is a model built around theoretical principles. Planning instructional outcomes set the stage for other instructional activities, such as evaluation. The expert system in the software assures consistency between planning and testing by recommending test types that coincide with the domain level used in the planning stage. The verbs from various taxonomies are coupled with alternative objects to form a hierarchical set of stem templates. These same verbs are linked to six test item types so that true-false and completion items will normally be used with low level cognitive content goals; matching and short answer will normally be used with middle level cognitive content goals; and multiple-choice and essay items will normally be used with high level cognitive content goals. The software recommends the type of test items based upon the

domain and level of the verb in the content goal.

Taxonomy of Educational Objectives

A step toward recognizing the complexity of educational objectives and the difficulties involved in measuring student achievement was taken when Bloom, Englehart, Furst, Hill, and Krathwohl (1956) worked to develop a taxonomy of educational objectives. The main reason for developing the taxonomy was to facilitate communication among educational researcher, curriculum developer, and evaluator (Payne, 1992). The taxonomy matched educational goals and test items in the cognitive domain.

According to Bloom, Englehart, Furst, Hill, and Krathwohl (1956), learning occurs in three different domains: cognitive, affective, and psychomotor. There are six major classifications of this taxonomy, arranged in hierarchical order according to complexity and sophistication: knowledge of factual information; comprehension of knowledge; application of knowledge and understanding to specific situations; analysis of complex ideas into their component parts; synthesis of disparate items into a whole; and evaluation of the consequences of actions. Payne (1992) promoted the taxonomy's usefulness for analyzing instruction, as in comparing the emphases in course planning with those in test questions.

The Cognitive Domain

The cognitive domain, which includes instructional objectives that deal with recall or recognition of learned material and the development of intellectual abilities, is at the core of curriculum test development (Stanley and Hopkins, 1972). The largest portion of educational objectives fall into this domain (Gaff, 1975; Krathwohl, et al., 1964; Payne, 1992).

Moving up the hierarchy, each category is assumed to include behaviors at the lower level; higher-level skills utilize lower-level ones and demand greater intellect from the learner (Jacobs & Chase, 1992). For example, comprehension includes knowledge because students must have to comprehend material before they can apply or analyze it. The application level includes knowledge and comprehension, and so on. Evaluation is the most demanding of the cognitive skills because it requires judgments using criteria recalled by the learner (Jacobs & Chase, 1992).

Two other domains in the taxonomy include the affective and psychomotor. The affective domain includes those objectives that emphasize feeling and emotion, such as interests, attitudes, appreciation, and methods of adjustment. The psychomotor domain includes those objectives that emphasize motor skill, such as handwriting, typing, swimming, and operating machinery (Gronlund, 1985).

Criticisms of the Classification Scheme

Though Bloom's Taxonomy is certainly a noteworthy contribution to educational classification, it is not without its problems. Krathwohl and Payne (1971) summarized the available research on the taxonomy of the cognitive domain and concluded that the rationale has been generally supported for the first three categories. Support for the order of the more complex categories has largely failed to develop. According to Stahl and Murphy (1981), this taxonomy has been and is being used even though it is not consistent with any presently accepted theory, model, or approach to human memory, thinking, or learning (Stahl & Murphy, 1981). Several researchers (Kropp, Stoker & Bashaw, 1966; Metfessel, Michael, & Kirsner, 1969; Stanley & Bolton, 1957) have reported that persons frequently disagree on the taxonomy level represented by many items, except those at the knowledge level.

In the taxonomy, the same term may be used at several levels. The verb "identify", for example, appears at the knowledge, understanding, application, and analysis levels. Although this may be appropriate, the overlap of terms causes confusion. Diamond (1989) believed many faculty were turned off because many of the early advocates of stating goals in performance terms focused on minutiae and on complex classification systems that faculty did not understand (p. 127). According to Diamond (1989), faculty spend a great deal

of time analyzing the type or level of their objectives. He believed it is far more essential that efforts are made to ensure that useful statements be written, that they include all of the elements that should be addressed, and that they be measurable within the context of the course. No matter how the objectives are conceived or categorized, if they are clearly stated, then it is easier for the instructor to plan content, deliver content, and to prepare tests to assess the extent to which objectives are achieved (Diamond, 1989).

Simplifying Faculty Course Planning

After studying over eighty instructional projects, Bergquist and Armstrong (1986) reported that the inability to state course goals in performance terms was a major problem. According to Diamond (1989), faculty resent stating instruction in behavioral terms and deem the task a waste of time. Consequently, they write objectives at a trivial level, which are probably the easiest to write.

Gaff (1975) believed that by stating objectives with a high degree of specificity, it is possible for everyone involved to recognize the behavior that meets or fails to meet them when evaluating. With this structure, the specific objectives inform instructors what their tests should be like. They tell what will be covered and with what emphasis (Jacobs & Chase, 1992). Considering the importance of planning and evaluation, Greeno (1976) believed the development of

instructional objectives begins with considering the kinds of tests used to assess whether students have acquired the knowledge intended as the outcome of learning.

Research Question E: What is the connection between taxonomy classifications? Bloom's Taxonomy is a well known classification scheme, and Vogler's classification of fact, understanding and application in the cognitive domain is tied to Bloom. Vogler (1991) simplified the task of planning. So, its applicability to planning and testing decisions is less confusing. While theoretically maintaining Bloom's three major classifications, Vogler presented each major category with three, distinct sub categories he termed levels. This is far less confusing than having to select verbs from among six categories using Bloom's classification. In the cognitive domain, Vogler's three sub categories or levels are "fact," "understanding," and "application." These levels also create an association that implicitly relates to testing. Two other domains, psychomotor and affective, each have three sub categories or levels. Faculty select appropriate verbs pertaining to each of the levels within the appropriate domain to plan instruction. The verb is used to plan instruction by creating an outcome-based content goal statement.

The type of design Vogler promoted in Performance Instruction relates to similar beliefs about planning and testing held by Gronlund (1981, 1985), Diamond (1989), Nitko

(1989), and Popham (1978). The domain-referenced verb in Vogler's (1991) model directs the instruction. Although a verb in Vogler's model may be referenced to more than one sub domain level (e.g., fact, understanding, application), it is appropriate as long as the phases of instruction are consistent. If a verb is selected to plan instruction at the application level, the same verb should also be used when delivering and evaluating instruction to maintain consistency at the application level.

Testing

Tests are an integral part of college courses, typically constructed to measure whether objectives have been met (Tuckman, 1975). Another purpose is to obtain valid, reliable, and useful information concerning student achievement so that learning can be improved (Gronlund, 1988). The instructor attempts to determine if students can behave as intended when the instruction was planned.

Lozak (1987) estimated that between one fourth to one third of an instructor's time is spent on measurement. According to Stanley and Hopkins (1972), constructing a test is one of the hardest jobs a teacher has to perform. It demands an understanding of the objectives being assessed and of the examinees and their test-taking behavior. In constructing classroom tests, a major concern is that test items call for particular types of performance indicated in

the specific learning outcomes pertinent to each instructional objective or content goal (Gentry, 1989; Gronlund, 1988; Heywood, 1989; Popham, 1978; Vogler, 1991). Baker and Popham (1970) stated that objectives and evaluation should be identical; test items should be drawn from the class of behavior specified in the objectives.

Test Categories

Tests are either criterion-referenced or norm-referenced. Criterion-referenced exams compare an individual's score to some specified level of performance or criteria; norm-referenced exams compare the individual's score to those of individuals in a group (Jacobs & Chase, 1992). In criterion-referenced testing situations, the meaningfulness of an individual score does not depend on a comparison with other testees. One seeks to know what the individual can do, not how he or she stands in comparison to others. When testing to a criterion, the score indicates within error limits exactly what an individual can and cannot do (Popham, 1973).

With criterion-referenced tests, scores are compared with some specific standard of acceptable performance. In criterion-referenced testing situations, instructors are not concerned with how difficult the items are or how well the items discriminate between students. The key element is how well the items reflect the specific learning tasks (Jacobs & Chase, 1992). While Tuckman (1985) indicated that criterion-

referenced tests in education are used to monitor student progress and evaluate instructional program effectiveness, Popham (1973) argued that few analyses have ever been made of the practical implications of these tests.

Typical norm-referenced measures, such as item-analysis procedures and internal-consistency estimates, as well as the standard notions of validity and reliability are designed to facilitate comparisons among individuals. It is not possible to tell a norm-referenced test from a criterion-referenced test by looking at it. Although a criterion-referenced test could also be used as a norm-referenced test, the reverse is not easy to accomplish (Popham, 1973).

The contrast between norm-referenced and criterion-referenced measurement demonstrates how current test construction and test improvement has purposes other than measuring the quality of instructional programs. "Mechanisms such as item-analysis procedures and internal-consistency estimates, as well as the standard notions of validity and reliability are not pertinent to the kinds of measurement procedures that must be used to assess the attainment of measurable instructional objectives" (Popham, 1973, p. 7-8). Popham (1973) believed a great deal of sophisticated work needs to be done on how to devise measures that satisfactorily serve to sample the behavior and content domains delimited by a well-stated objective. To go along with this, Popham

suggested using efficient ways to produce satisfactory measures of criterion-referenced objectives so that they can be economically assessed.

Teacher-Constructed Tests

A measure more closely aligned with taught content is the teacher-developed test. Ebel (1965) believed classroom tests prepared by the faculty member are likely to fit content and objectives of a particular course better than a test prepared by anyone else. Dorr-Bremme (1982) reported that over 75 percent of high school faculty create their own tests. MacCuish (1986) reported that 91 percent of college faculty create their own tests. However, the process of constructing a good test item is time consuming. It demands an understanding of the objectives being assessed and of the examinees and their test-taking behavior (Stanley & Hopkins, 1972).

Tests differ in what type of interpretations are made from the scores, how the content of the test is developed, and the appropriateness of the tests for making various educational decisions (Salvia & Hughes, 1990). Publisher-prepared tests include items that have not been carefully prepared or critically reviewed by other experts in educational measurement or in the subject field (Ebel, 1965). Further, the content of published tests often does not match the content that is taught. While these tests may reflect

students' intellectual abilities, they are useless for evaluating what a student has learned from school instruction.

Unfortunately, some instructors, outstanding in their scholarship and teaching ability, possess naive notions about the requirements for effective measurement of educational achievement (Ebel, 1965). Findings suggest teaching and testing often occur to a considerable extent at the lowest level (Stanley & Hopkins, 1972; Ball, Doss & Dewalt, 1986). Students are not encouraged to use their minds at higher levels. One of the most common defects of teacher-constructed tests is the lack of relevant, difficult items. Errors that occur in testing include writing too many questions for a certain topic, including only a few questions from a certain part of the course, and completely ignoring another part of the course (Copperud, 1971).

Some faculty are more inclined to ask about the specific, incidental details than about the important general principles (Ebel, 1965). Jacobs and Chase (1992) reported that tests:

"include too many questions requiring only knowledge of facts and trivia and provide no intellectual challenge to the students. Students often complain that test content does not reflect the material discussed in class or what the professor seemed to indicate was most important. They often feel somewhat cheated when they have put forth a great deal of effort and then take a test that does not permit them to show what they have learned" (p. 13).

Teachers display weaknesses in test construction.

"Content-oriented teachers can become so preoccupied with the importance of students' learning specific facts and terms that their students fail to gain a sense of direction and a level of understanding, which comes from seeing the interrelatedness of concepts and principles and applying them to unfamiliar problems" (Adams, 1966, p. 329).

Fleming and Chambers (1983) analyzed over 400 teacher-developed tests and thousands of test items. They found that (a) teachers use short answer questions most often; (b) teachers avoid essay questions; and (c) teachers used more matching items than multiple-choice items. They also found that teachers devise more test questions to sample knowledge of facts than any of the other behavioral categories; teachers develop few questions to test application behaviors.

Proper balance improves test reliability and validity (Gentry, 1989). It also lessens the chance that the test will be filled with items dealing with the instructor's "pet topics" or items easy to construct. It also makes it less probable that topics or content will be ignored. Stanley and Hopkins (1972) recommended using a table of specifications to guide the test maker so that the test essentially mirrors the instructional objectives.

Most college instructors feel poorly prepared to construct tests in their classes because they have never received any kind of formal training in this area (Jacobs &

Chase, 1992). Lozak (1987) blamed the higher education system for not preparing its graduates for the assessor role. Goslin (1967) found only 40 percent of elementary and secondary teachers had as little as one class in testing and measurement. Yeh (1978), in a study of teacher use of test results, reported that only 50 percent of the teachers sampled were able to correctly interpret percentile ranks and grade equivalents. Consequently faculty are uncomfortable with test construction and view it as a difficult chore. Milton and Associates said, "testing is perhaps the most neglected feature of good instruction" (1978, p. 101). Stanley and Hopkins (1972) concluded that if evaluation procedures are poor, then the quality of the information on which judgements are based cannot be adequate.

Students study what they think they are going to be asked in the instructor's tests (Anderson, 1987). If students believe instructors will emphasize factual recall, students will learn facts. "If instructors are going to ask them to apply material to new situations, students will study and learn how to apply principles and theories. If students believe they will be required to analyze, synthesize, and make judgments, then they will learn to use higher-level thinking skills" (Jacobs & Chase, 1992, p. 5-6). Instructors need to gain a perspective on their purposes of testing since tests and examinations profoundly influence what students study

(Anderson, 1987, p. 40).

Classroom Testing

Testing is an integral part of instruction. Unlike the demand for expanded use of test scores for accountability, largely the creation of legislators, administrators, and policy boards, Linn (1989) believed the renewed interest in testing was to make better day-to-day instructional decisions. The movement to make better instructional decisions is spurred by educational measurement specialists and cognitive psychologists. The instructional use of test results is not to predict who will succeed or to report results to satisfy accountability demands. Rather, it is to help individual students gain the most from instruction (Linn, 1989).

According to Isaac and Michael (1990), the purpose of testing is not on theory building but on product delivery and mission accomplishment. Dyer (1970) stated that the measurement process is usually thought of only in terms of numerical test scores or other quantitative descriptions of the phenomena being measured. Nitko (1989) observed that traditional inquiry into test design has focused on optimizing measurement efficiency rather than on optimizing instructional efficacy.

Tests given during the course assess the extent to which students are achieving the stated objectives. Jacobs and Chase (1992) believed:

"classroom tests provide feedback on what students are learning in the class and thus help instructors to determine the effectiveness of their teaching of a particular segment of the course. By analyzing the test results, faculty can identify problem areas that perhaps should be reviewed or retaught before students move on to new material" (p. 3).

The idea that testing should help students learn better is not really new. Cook (1951) concluded that all the functions of educational measurement are concerned either directly or indirectly with the facilitation of learning. Tyler (1951) argued that this facilitation requires linking testing and instruction. He viewed educational measurement not as a separate process, but an integral part of instruction.

Purposes of Classroom Tests

Instructors commit a common error in testing when they fail to communicate to students the purpose of their classroom tests and how the results will aid learning (Jacobs & Chase, 1992). According to Linn (1983), there are three purposes of classroom tests: (a) to provide feedback to students and to the teacher; (b) to flag facts or concepts that are considered important; and, (c) to determine grades.

According to Adams (1966), tests provide students with tangible indicators of the outcomes expected from a course, even to a greater degree than do the textbook or syllabus. It is largely the in-class, teacher-made test that provides

students with confirmation or feedback concerning the effectiveness of their efforts to learn (Adams, 1966). Feedback from tests also helps teachers provide more appropriate instructional guidance for individual students and the class as a whole. Beyond providing feedback, tests are used as a means of demonstrating mastery or certification (Fortune, 1985; Stanley & Hopkins, 1972).

A test is often given to students before admission to a program of studies to increase the quality of selection and classification (Stanley & Hopkins, 1972). If the admission policy of the institution changes, the purpose of the test can change. For example, if an admission policy were liberal and encouraged admission to a certain program, placement testing requirements might be less stringent. For such purposes, a condition of undertesting content may be legitimate.

Pretests also provide a baseline to evaluate the changes over a semester's time and hence the effectiveness of the instruction that has taken place (Jacobs & Chase, 1992). Tests at the beginning of the course serve as a pretest of content knowledge so faculty can measure students' knowledge of prerequisite material or concepts to be covered in the course and can plan their presentations accordingly (Jacobs & Chase, 1992).

When faculty review, interact with, or practice skills and concepts after they have been mastered, they engage in

what psychologists call overlearning. In certain situations, faculty may purposely want students to overlearn content. Examinations are a useful means of overlearning (Stanley & Hopkins, 1972). Overlearning contributes to long-term retention. When tests are designed for such purposes, a condition of overtesting may be legitimate.

Planning the Test

Before any test items are written, Jacobs and Chase (1992) urged instructors to develop a plan that specifies clearly what they are going to measure. A test plan enables the instructor to create an instrument that tests what has been planned and taught. Typically, a test plan has two components: the content to be covered and the cognitive skills to be measured (Jacobs & Chase, 1992). Test item content thus relates back to some plan. In the second component, instructors are concerned with measuring cognitive skills or the intellectual processes that students use to construct answers to questions.

Boersma (1967) found that teachers systematically involved in applying evaluative criteria had a clearer perception of the curriculum. Bloom (1961) noted that when teachers participated in the construction of tests in a systematic way, not only were objectives clarified, but more relevant instruction also occurred. Ebel (1965) insisted that classroom tests be prepared by the faculty member. He said,

"if testing and teaching are in the hands of the same person, they are likely to be more effectively integrated in the total educational process than if the testing were separated from the teaching" (Ebel, 1965, p. 9). To insure that the curriculum drives the tests, test construction should follow and be dependent upon curriculum development (English & Steffy, 1983). Professional development for faculty should focus on test construction and measurement techniques.

Table of Test Type Specifications

Gronlund (1985) believed systematic procedures to obtain a representative sample of student performance is the only assurance that a classroom test validly measures instructional objectives and course content. A device widely used for this purpose is the two-way, table of specifications chart. Adams (1966) believed there was sound rationale for building a table of specifications.

"If a teacher accumulates test items without a plan, they will unduly represent informational learning, especially knowledge of specific facts. Moreover, teachers are likely to overemphasize certain areas of content in which items are easily constructed. As a way to improve the test's representativeness, or its content validity, one should first develop a blueprint for the test" (p. 327).

The chart relates instructional objectives to course content and specifies the relative emphasis to be given to each type of learning outcome. On the chart, content topics are listed on one dimension and cognitive skills on the other.

According to Adams (1966), a two-way table of specifications gives adequate consideration to both the content and cognitive abilities of the course. He believed it was important that test coverage be adequate from both points of view.

To build a table of specifications one must (a) obtain a list of instructional objectives, (b) outline the course content, and (c) prepare the two-way chart. Gronlund (1981) argued that effective evaluation of student learning relates evaluation procedures as directly as possible to the intended learning outcomes. This is more easily accomplished if the instructional objectives and the specific learning outcomes have been clearly stated in performance terms. It is then a matter of constructing or selecting evaluation instruments that provide the most direct evidence concerning the attainment of the stated outcomes (Gronlund, 1981).

Preparing test items directly relevant to the instructional objectives is primarily a matter of matching the performance measured by the test items to the types of performance specified by the intended outcomes. If the intended learning outcomes call for identifying a procedure, the test items should be concerned only with the process of identifying, rather than more complex outcomes. If the intended learning outcomes call for performing a procedure, the test items should require actual performance, rather than

a verbal description of how to do it.

Selecting Appropriate Test Types

Items used in classroom tests are generally divided into two general categories: (a) structured, objective items that require students to fill-in-the-blank or to select the correct answer from among a limited number of alternatives, and (b) subjective items that permit students to select, organize, and present answers in written form (Gronlund, 1981). Powell and Gillespie (1990) used the term selected-response to refer to what Gronlund called objective test items and constructed-response for subjective test items.

Selected-response tests require much more time to create, but scoring them is relatively quick. While questions for a constructed response test are relatively easy to prepare, they are much more difficult to grade. Despite scoring complexities, the use of the constructed-response test is rising (Powell & Gillespie, 1990).

The specific type of item selected depends on the nature of the objective to be measured (Gronlund, 1981; Stanley & Hopkins, 1972; Vogler, 1991). If faculty want students to learn facts, they test them on their ability to recall information; if they want to promote critical thinking, they write test items that emphasize understanding, application, and other higher-level skills (Jacobs & Chase, 1992). The process of relating test items as directly as possible to

specific outcomes provides us with greater assurance that the test is a valid measure of instructional objectives (Gronlund, 1981).

Advantages and disadvantages of subjective and objective test items have been cited by authorities in the field of teaching and measurement (Cunningham, 1986; Ebel & Frisbie, 1986; Green, 1975; Jacobs & Chase, 1992; Mehrens & Lehmann, 1984; Payne, 1974; Roid & Haladyna, 1982; Swezey, 1981; Thorndike & Hagen, 1969; Vogler, 1991; Wesman, 1971). Each type of test item is efficient for measuring some learning outcomes and inefficient for measuring others. Popham (1978) stated that for measuring knowledge of factual information, the selected-response test is more efficient. If the aim is to measure originality, the ability to synthesize ideas, write effectively, or to solve problems, constructed-response tests are better (Powell & Gillespie, 1990).

PEAKS ExamBuilding Software (tm), which is a manifestation of evaluating in Vogler's (1991) C-P-A model, uses six test item types for the cognitive domain. They include true-false, matching, multiple-choice, completion, short answer, and essay. It is important to note that the level of the test item used to test content must be consistent with the level at which the content was planned. As Gronlund (1981) suggested, a table of test type specifications is a useful way to ensure unity between testing and planning. The

concept of unity holds that the domain level used to plan student learning outcomes be identical to the level used when testing those outcomes. The expert system preserves unity through its recommendations. Unity is preserved if the test type level matches the level at which content was planned.

True-False Test Items

A true-false item is a declarative statement that the test taker must decide is correct or incorrect. Of the various types of questions found in teacher-made tests, the true-false item is undoubtedly the most widely used and the most severely criticized. True-false items are popular with teachers because they seem relatively easy to construct (Adams, 1966).

A true-false item seeks to assess cognitive information at the lower levels of the domain. It is efficient for measuring fact acquisition; it does not effectively measure understanding or application (Vogler, 1991, p. 217). Moreover, Adams (1966) observed, true-false items are very well adapted to testing (a) understanding of principles or generalizations; (b) persistence of popular misconceptions; and (c) situations in which there are only two logical responses (e.g., north, south; right, left; larger, smaller). Jacobs and Chase (1992) indicated that a true-false test can sample many more bits of information about a topic than any other test format. The biggest drawback associated with a

true-false item is that it inherently has a 50 percent guess factor (Vogler, 1991).

The expert system in Performance Instruction recommends true-false be used for testing content at the factual level, or level one. The expert system deems a true-false test effective with level one content; it is efficient with level one, level two, and level three content.

Matching Test Items

In a matching exercise, there are several problems or questions; the answer to each one is to be chosen from a single list of options. The matching test item is characterized by two columns of topics. The left column usually contains the stem that is matched with a response appearing in the right column. A matching item is used to test facts and understanding (Vogler, 1991). The guess factor is eliminated if the alternatives are greater than the number of stems to be matched.

The expert system in Performance Instruction recommends a matching test be used for testing content at the understanding level (level two). The expert system deems a matching test effective with level one content and level two content. It is efficient with level two and level three content.

Multiple-Choice Test Items

Multiple-choice test items are the most widely used and

versatile selection-type item in the college classroom (Jacobs & Chase, 1992). A multiple-choice item consists of an incomplete statement followed by several possible completions. These tests are designed to test cognitive information at the fact, understanding, and application level (Vogler, 1991). Adams (1966) believed multiple-choice items can be designed to require reasoning and judgment. Properly constructed multiple-choice items are especially good for measuring higher-level cognitive objectives (Jacobs & Chase, 1992). They are also used to assess the level of cognition by looking at the responses for second and third best choices (Vogler, 1991). Well-constructed multiple-choice items tend to be more valid and reliable than an equivalent number of true-false items (Adams, 1966).

It takes time, effort and creativity to write multiple-choice items that require students to think rather than to recall facts (Jacobs & Chase, 1992). A multiple-choice item consists of a stem and a number of possible responses, including the correct or best answer plus three or four distractors.

The expert system in Performance Instruction recommends a multiple-choice test be used for testing content at the application level, or level three. The expert system deems a matching test effective with level one content, level two content, and level three content. It is efficient with only

level three content.

Completion Test Items

Completion items, which call for a free response, are the simplest type of subjective test. They are often used to test vocabulary development and thus tend to be limited to lower-level testing (Vogler, 1991). Such items consist of a sentence with words or phrases omitted. Blank spaces are usually provided so students can complete the sentences. In completion tests, the student gets minimal clues and must construct the answer. Many test makers like to use completion items because they believe the demand on the student is greater due to the items' demand for recall of information, not recognition (Jacobs & Chase, 1992).

The expert system in Performance Instruction recommends a completion be used for testing content at the factual level (level one). The expert system deems a completion test effective with level one content; it is efficient with level one, level two, and level three content.

Short-Answer Test Items

Short-answer items are used to test facts and understanding. The responses required in a short-answer test are limited, but there is considerable variance as to the length. Students get minimal clues and must construct the answer. In this respect, short-answer test items are similar to completion items. Short-answer items are applicable to

almost all kinds of topics (Jacobs & Chase, 1992). Generally, three paragraphs and 250 words frame the outer limits of a short answer (Vogler, 1991).

The expert system in Performance Instruction recommends a short-answer test be used for testing content at the understanding level (level two). The expert system deems a short-answer test effective with level one content and level two content. It is efficient with level two and level three content.

Essay Test Items

Essay questions produce answers that require subjective scoring because each student's answer to a particular question will differ. Many faculty believe that essays convey thinking processes as essays provide no prompts like other objective tests (Jacobs & Chase, 1992). Facts, understanding and application can be tested through an essay. Stanley and Hopkins (1972) cited the works of Hoffman (1962) and La Fave (1966), who indicated that essay tests can measure more complex, higher levels of understanding. "Essay tests have many virtues, but efficiency, adequacy of sampling, and reliability of scoring are not among them" (Ebel, 1965, p. 14). Jacobs and Chase (1992) warned that essays are not practical to test factual matter and similar lower-level cognitive objectives.

The main advantage of essay tests, according to Jacobs

and Chase (1992), is that they provide a measure of the more complex learning outcomes. "Essay tests require students to recall information, to perceive the necessary relationships, and to construct the answer, unaided. They enable instructors to see how students select, organize, and evaluate ideas and apply them to answering the questions" (Jacobs & Chase, 1992, p. 103).

The expert system in Performance Instruction recommends an essay test be used to test content at the application level (level three). The expert system deems an essay test effective with level one content, level two content, and level three content. It is efficient with only level three content.

Summary

Teaching is central to the mission of the community college. The desired outcome of any teaching strategy ought to be effective and efficient instruction. This review presented instructional concepts deemed essential in facilitating efficiency and effectiveness. In addition, platform unity, based on the Principles of Performance Instruction, was viewed as a way to uphold curriculum alignment principles. Performance Instruction is the planning, delivery, and evaluation of learning and teaching. It is based upon the Vogler Curriculum-Pedagogy-Assessment (CPA) model (Vogler, 1991).

The integration of planning, delivering, and evaluating

is often neglected in traditional approaches to instruction. For example, content is often planned, but not evaluated. The inverse is also true. Performance Instruction holds that course content should be planned, delivered, and evaluated consistently to assure unity.

Over the years, scholarly contributions have focused on planning and evaluation. More recently, however, a focus uniting curriculum elements has surfaced. Figure 1 shows some authors who have dealt with each of these areas. All published works are mentioned within the body of this chapter. Individual citations appear in the reference section.

Planning-----Unity-----Testing

Bloom (1956)
 Krathwohl (1961)
 Simpson (1965)
 Gerhard (1971)
 Baker & Popham (1970)
 Popham (1973, 1978)
 Mager (1975)
 Gagne (1975)
 Gagne & Briggs (1979)
 Gronlund (1981, 1982, 1985)
 Diamond (1989)
 Stark (1991)
 Vogler (1991)

Tyler (1951)
 Cook (1951)
 Ebel (1965)
 Adams (1966)
 Stanley & Hopkins
 (1972)
 Copperud (1979)
 Airasian & Madaus
 (1983)
 Tuckman (1985)
 Vogler (1991)
 Jacobs & Chase
 (1992)

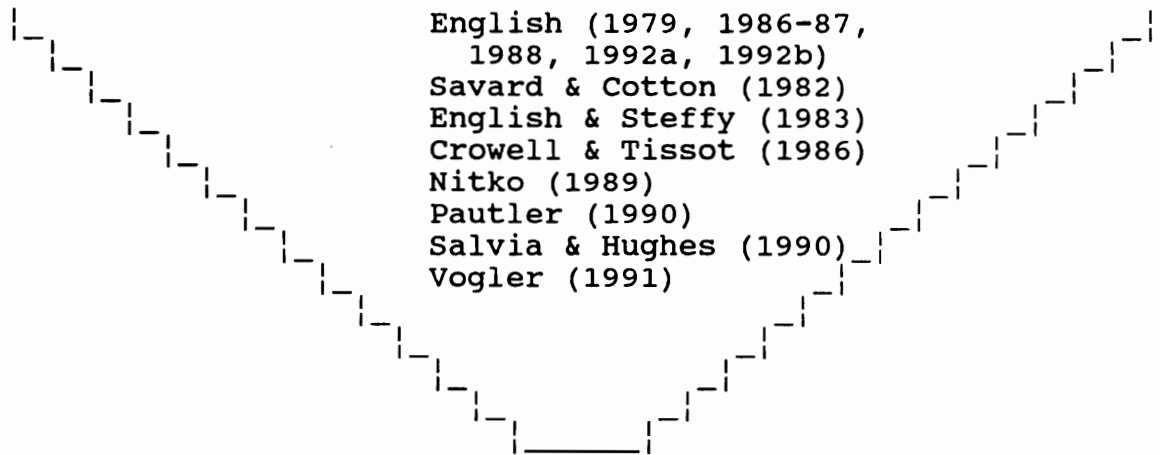


Figure One -- Matrix of Scholar Contributions

Faculty, when planning a course, plan outcomes of instruction. The key to successful planning is to create outcome-based objectives by selecting domain-referenced verbs. The same verbs are used in planning and testing to assure consistency. Testing is not a function separate from planning. When creating tests, faculty need to relate to content planning decisions. The domain-level at which content was planned becomes the basis for creating test items. According to Performance Instruction, content planning and testing occur at the same domain level to assure unity.

In the literature review, considerable attention was given to faculty's inability to plan and test content consistently. These are arguably important and necessary faculty tasks. If instruction is to be effective, then staff development efforts to explain effective planning and testing techniques should be undertaken. Unfortunately, even the literature on planning and testing treats them as separate, independent activities.

Tools to assist faculty in assessing criterion-referenced tests were deemed necessary based on the lack of such measurements in the extant literature. Five constructs were reviewed in this chapter. They were based on the concept of unity between planning and testing. These constructs included platform unity, overtesting, undertesting, effectiveness, and efficiency.

Faculty training in assessment ranges from very little (Coffman, 1983; Ward, 1982) to none (Stiggins & Bridgeford, 1985). Many faculty have had no formal coursework and teacher preparation programs do not require measurement training. In addition, faculty have had no in-service training in assessment (Stiggins, Conklin, & Bridgeford, 1986). From research on testing in the schools, the researcher arrived at the following points: (a) testing is conducted by teachers whose formal knowledge of assessment is minimal, (b) classroom tests are usually short and objective, (c) teacher-developed tests are of poor technical quality, and (d) tests generally call for memorization of facts.

Faculty in community colleges may have limited formal coursework in teacher preparation. In-service workshops may be one of the best ways to provide community college faculty with exposure to measurement techniques. Goslin (1967) reported that teachers with measurement training and experience generally tended to make the most use of tests. Kellaghan, Madaus, and Airasian (1982) argued it was not necessary for faculty to know more about interpreting test scores. Rather, they suggested offering training sessions that provide guidance in practical, quality control measures for teachers.

The lack of faculty skills to judge the worth of teacher-made, criterion-referenced tests is apparent. The gap affects

instructional practices as the use of classroom assessment techniques are increasing in postsecondary education.

There is little research in the field of tests and measurement regarding practical tools to help faculty evaluate criterion-referenced tests. Standard measurement tools used to evaluate norm-referenced tests, including reliability, validity, difficulty, and discrimination are generally not applied to criterion-referenced testing situations by faculty. The five analytical constructs reviewed in this study serve as specific and practical tools to judge unity, overtesting, undertesting, effectiveness, and efficiency of content planning and testing. These tools provide a measurement standard in criterion-referenced testing situations, similar to standards used in norm-referenced testing situations.

Concepts that undergird curriculum alignment were used to form the basis for exploring unity. The purpose of this study was to document the effects of applying the platform unity concept. The manner in which the unity concept was developed is presented in Chapter Three. The findings and implications are presented in Chapter Four and Chapter Five.

CHAPTER III

RESEARCH DESIGN AND METHODOLOGY

This chapter describes the research methodology and design, sampling plan, sample population, data collection procedures, instrumentation and field test, data analysis and recording, methodological assumptions, and analytical considerations. Developing an expert system Table of Test Type Specifications was a necessary step to align planning and testing prior to conducting analyses.

To fulfill the purpose of this study, which was to document the effects of applying the platform unity concept, the procedures included the following eleven steps:

- (a) developing formulae to test unity, overtesting, undertesting, effectiveness, and efficiency,
- (b) testing of formulae,
- (c) obtaining a list of outcome-based content goals,
- (d) designing survey instruments to analyze unity between content planning and testing based on Vogler's (1991) model using the content goals obtained in step C,
- (e) field-testing of survey instruments,
- (f) selecting institutions and survey participants,
- (g) mailing first phase survey instrument,
- (h) analyzing Phase I survey data,
- (i) verifying faculty planning and testing decisions obtained in Phase I through a second phase survey instrument,
- (j) analyzing Phase II data, and

(k) reporting findings from Phase I and Phase II.

This chapter presents the tasks that were accomplished to fulfill the above listed procedures. The research methodology involved two separate steps. First, computational formulae were developed and tested. Second, the formulae were applied to planning and testing decisions provided by survey respondents. This chapter concludes with an explanation of the theoretical constructs measured by the formulae.

Developing Formulae

In the first step, an extensive amount of research, development and experimentation occurred in defining the variables and using arithmetic logic to produce functional formulae capable of yielding accurate and meaningful information. The process required hours of testing and modification to enhance the ease of applicability while upholding sound pedagogical theory. The procedures in this step included the following:

- (a) reviewing the concept of curriculum alignment,
- (b) reviewing the theoretical constructs in course planning,
- (c) classifying planning decisions according to a domain-referenced taxonomy,
- (d) examining various test types used to test content,
- (e) creating an expert system table of test type specifications, and
- (f) amalgamating planning and testing concepts so that

they were capable of numerical representation through mathematical formulae.

Testing Formulae

Two mathematicians independently reviewed the formulae for mathematical soundness following extensive prototype testing. The three formulae used in this study included platform unity, effectiveness and efficiency. The expert panel deemed them functionally reliable and capable of discerning the indices that were later used for analytical decisions.

Based on successful trials, it was deemed possible to mathematically produce indices of the three constructs through the aid of these formulae. These indices were for platform unity between planning and testing, and effectiveness and efficiency of test type selection. Results of the platform unity calculation also determined the extent to which course content was overtested and undertested.

The formulae were based on alignment principles between content planning and testing. Essentially, the numerical input required domain levels of planned content to equate with corresponding test types. Three planning levels were matched with corresponding levels of test items. A review of the theoretical foundations of Vogler's (1991) domain classification provided the framework for formulae creation. The concepts of unity, effectiveness and efficiency are grounded in the principles of curriculum alignment. Alignment

is achieved if the written curriculum, the taught curriculum and the tested curriculum are congruent.

The concept of unity, along with the broader concept of alignment was detailed in Chapter Two. In this study, the theoretical construct of alignment contained practical dimensions based on the Vogler C-P-A model and software marketed by Instructional Performance Systems, Inc (Vogler, 1991). Platform unity was analyzed by examining faculty planning and testing decisions. Content plans were based on outcome-referenced content goals. Content planning and testing decisions needed for formulae calculations were obtained from faculty through survey research methodology. The rationale for the survey and steps used in constructing survey instruments are detailed in the next section.

Research Methodology

Dillman (1978) stated survey research methodology should be used when dealing with quantitative data, controlling costs of conducting research, and including a large portion of the population in the sample size. Survey research was used in this study to collect data and fulfill the purpose, which was to document the effects of applying the unity concept. Two phases of a survey served as the means of obtaining information from faculty regarding content planning and testing decisions.

Sampling Frame and Survey Sample

The sampling frame was a listing of all computer faculty in the Virginia Community College System (VCCS). From this listing, the researcher obtained observation units, consisting of faculty members teaching Introduction to Computers at one of 23 community colleges in Virginia. To be as accurate as possible in determining the population, faculty members names were obtained by placing telephone calls to each community college in Virginia. The division chairperson of computer technologies provided the names of faculty whose primary teaching responsibility included the course Introduction to Computers. This course is the introductory computer literacy course taught in the Virginia Community College System (VCCS). Fifteen community colleges were selected at random from the total list of community colleges. Two observation units were selected from each of the 15 campuses identified. The majority of community colleges had only two faculty who were responsible for Introduction to Computers. In cases where there were more than two per college, names were randomly selected. The first two drawn were selected to participate in the survey. A total of 30 faculty were thus identified as the survey population.

Steps in Constructing Survey Instruments

The steps listed below were used in the construction of the survey instruments.

1. The researcher obtained a list of instructor content goals used to plan Introduction to Computers. To accomplish this, the researcher contacted a faculty member who taught Introduction to Computers at Southwest Virginia Community College. This faculty member planned course content using Vogler's (1991) Principles of Performance Instruction. The researcher obtained six course content goals, two from each cognitive domain level, and included them on the survey instrument used in Phase I.
2. In addition to the content goals, eight demographic variables were included on the instrument used in Phase I. The demographic variables were considered control variables in the analysis of responses.
3. The second-phase instrument requested survey participants to verify Phase I information. The frequencies of test types selected by respondents in Phase I were listed on the Phase II instrument. The sampling unit for Phases I and II was identical.

Field Test of Instrument

The survey instruments used in this study are included in Appendices A and B. Field-test questionnaires are provided in Appendix F. Two separate field tests were conducted, one for each survey phase, using principles recommended by Babbie

(1990). The instrument used in Phase I was field tested among four community college computer faculty in Maryland and eight graduate students in the College of Education at Virginia Tech. Ten usable field-test instruments were obtained.

The instrument used in the first phase of data collection contained six content goals, two from each cognitive domain level identified by Vogler (1991). The content goals on the field-test instrument were actual content goals used in Introduction to Computers.

Field-test subjects were administered the instrument along with a cover letter and instructions. Each subject was ask to critique survey documents. The instructions to field-test participants are included in Appendix F. In addition, a check-off form requested field-test participants to review each item and suggest ways to improve it.

Appendix A presents the revised cover letter, instructions, and the final Phase I instrument. Based upon feedback from field-test subjects, revisions were made to the Phase I instrument. These revisions included, referring to the course as "Introduction to Computers" rather than by course number. Apparently, several community colleges in Virginia have different course titles for this course. CIS 100, used during the field test, was changed to "Introduction to Computers" to reflect a broader course and to avoid confusion. Changing to "Introduction to Computers" permitted

respondents to view the course consistently. Second, a content goal was changed to provide more clarity. Field-test respondents did not like the content goal "compare computer language usage." In its place, a new content goal, "compare microcomputer software applications" was added. Third, the format of the instrument was changed. A heading appearing on each page of the field-test instrument was eliminated. The result was a much cleaner-looking and condensed instrument. Fourth, two response stubs were changed to reflect more accuracy.

Appendix B presents the revised cover letter, instructions, and the final survey instrument used in Phase II. This instrument was field tested with four computer faculty in Maryland and four graduate students in the College of Education at Virginia Tech. Six usable field test instruments were obtained. The changes made to the Phase II instrument based upon field-test respondents dealt only with format. There were no substantive changes made to this instrument based on the field test.

Data Collection Procedures

Mode of Gathering Data

The mode of gathering data in this study was through a mailed survey. Timing was important to increase the response rate. Although the instruments were prepared long before they were mailed, the researcher purposely waited until faculty

returned to begin the Spring, 1993, semester. Requesting faculty to complete survey forms in December with all the other end-of-semester pressures was deemed inappropriate.

Respondents were assured confidentiality. The statement, "results will be tabulated so that individual institutions or respondents will not be identified" appeared on the cover letter to respondents in both survey phases. These cover letters appear in Appendices A and B.

Receipt Control

Receipt control, if used properly, can assist the researcher in obtaining a better overall rate of return (Babbie, 1990). In this study, the process of receipt control began when the survey participants were identified. Telephone calls were made to computer faculty identified as the sample at each community college. The telephone calls verified the participants and sought their agreement to participate in the survey. All 30 faculty contacted agreed to participate.

These telephone calls occurred during the second and third week of December, 1992, because the survey was to be distributed during the break between fall and spring semesters. Immediately following the telephone calls, post cards were mailed to faculty thanking them for agreeing to participate and reminding them that the instruments would be in their mailboxes when they returned from holiday break.

Phase I survey instruments were mailed on December 22,

1992 to all faculty. It was assumed the instrument would arrive in faculty mailboxes ready for them to complete when they returned to campus. Most faculty were scheduled to return from semester break on January 4, 1993. Phase I had a return deadline of January 12. The turn-around time was assumed reasonable as faculty obligations during the week of January 4 were minimal.

In Phase I, 19 out of 30 surveys (63.3%) were returned by the established due date. On January 13, 1993, the first follow-up was mailed to non-respondents. One response was obtained on this day. Therefore, the follow-up was mailed to ten individuals. The contents were identical to those in the first mailing: a cover letter, instruction sheet for completing the survey, a survey instrument, and a postage-paid return envelope. All 30 surveys were obtained on January 26, 1993.

On January 26, Phase II was mailed to all survey participants. Twenty-two of 30 surveys (73.3%) were returned by February 3, the established due date. On February 4, the first follow-up was mailed to non-respondents. One response was obtained on this day. Therefore, the follow-up was mailed to seven individuals. The contents were identical to those in the first mailing of Phase II: a cover letter, instruction sheet for completing the survey, a survey instrument, and a postage-paid return envelope. On February 18, 28 Phase II

surveys were obtained.

Surveys in both phases were closely monitored for receipt control. Return rate graphs were constructed to monitor returned surveys. A separate graph was developed for each survey phase. Figure 1 in Appendix C contains a receipt control chart used to monitor Phase I survey returns. Figure 2 in Appendix C contains a receipt control chart used to monitor Phase II survey returns. A daily and a cumulative rate of return was calculated and monitored by the researcher. As surveys were returned, the researcher recorded the date. Each survey instrument was examined to see if the data was usable. Usable data were entered into a computer database.

Callbacks

Callbacks were made to five participants in Phase II who exceeded the second due date. All respondents except two indicated they would mail back the survey instrument. After numerous attempts to reach non-respondents, messages were left with department secretaries. The messages requested faculty to complete and return the Phase II instrument if they had not already done so. As of February 18, 1993, two individuals remained uncontacted after numerous attempts. Data collection efforts were officially concluded after consultation with the researcher's committee chair, having obtained 28 usable Phase II responses.

Data Processing of Results

A coding scheme was developed to assist with data analysis. Two statistical computer software packages were used: Number Cruncher Statistical System (NCSS) and dBase III Plus. A coding sheet was used to guide the coding process. Coding for the variables was exhaustive and mutually exclusive, meaning every data element fit into only one category.

The Analysis Plan

According to classifications proposed by Isaac and Michael (1990), this study was descriptive. Descriptive research is "describes systematically the facts and characteristics of a given population or area of interest, factually and accurately" (Isaac & Michael, 1990, p. 46). To describe this situation, the method of analysis used was the application of unity, effectiveness, and efficiency formulae on faculty responses to survey questions. A total of five constructs were determined through the formulae. They included platform unity, effectiveness and efficiency of test types, and overtesting and undertesting of content. Formulae were based on the Vogler C-P-A model and the Principles of Performance Instruction.

In addition to determining numeric indices for unity, effectiveness, and efficiency, the hypothesis, "there is no difference between expert system and faculty test type

selection," was tested using the Chi-square Goodness of Fit Test. According to Hinkle, Wiersma, and Jurs (1979), the Chi-square statistic is frequently used to analyze nominal data. In this study, observed frequencies were compared to theoretical frequencies contained in the expert system. The Goodness of Fit Test indicates whether or not the observed frequencies are a "good fit" to the expected frequencies.

Table of Test Type Specifications

An ancillary purpose of this study was to create an expert system table of test type specifications. This table is necessary to equate test types with domain-referenced content plans. According to Vogler (1991), test types can be paired by cognitive domain level. For example, at the factual level, the lowest level of cognitive content, the only two test types having absolute platform unity are true-false and completion items. At the understanding level, the second level, the only two test types having absolute platform unity are matching and short answer. At the application level, the highest level of cognitive content, the only two test types having absolute platform unity are multiple choice and essay. Table 2 shows the cognitive domain levels in parity with expert system test types.

Table 2

Table of Test Type Specifications

<u>Domain Level</u>	<u>Test Items</u>
1 Fact	True-False Completion
2 Understanding	Matching Short Answer
3 Application	Multiple Choice Essay

Using Goodness of Fit To Test For Difference

To calculate Goodness of Fit using survey data, six separate Chi-square calculations were used, one for each content goal listed on the Phase I survey. The content goal verb used to plan instruction determined the cognitive domain level; the number of respondents (N) determined the expected frequencies in a two by two contingency table. The expected frequencies represent the expert system or absolute platform unity. The actual or observed frequencies were obtained from faculty survey responses.

Using Phase I survey data, faculty responses to test type choices were used for Goodness of Fit calculations. All faculty choices were counted. According to the Table of Test Type Specifications, these choices should be determined by the cognitive domain level. For calculations, test type choices

were considered dichotomous. Those test types matching the domain level were counted and placed in the upper left quadrant of the two by two contingency table. All other test type choices were placed in the upper right quadrant, as shown in the illustration below.

	Domain Level Test Type Frequency	All Other Test Type Frequencies
Observed	xx	xx
Expected	28.5	1.5

Since respondents should select domain-referenced test types, the correct choices according to the expert system, the hypothesized frequency assigned to the lower left column of the two by two contingency table was 95 percent. Using .95 allows for a five percent margin of error. The expert system expected frequencies were determined by the number of responses obtained, in this case 30. The Chi-square Goodness of Fit Test was then calculated for the six content goals in Phase I.

To obtain observed values for the domain level test type category, only the selected test types corresponding to the cognitive domain level were counted. For example, if the cognitive domain level was two, only "matching" and "short answer" survey responses were counted and recorded as observed, domain level frequencies. All other selected test

types were collapsed into the second cell in the observed row. The research hypothesis is that there is a difference between the expected (expert system) and the observed (faculty test type selection).

Control variables analyzed in the sample population include, earned academic degrees, academic rank, years teaching experience, years in present teaching position, having completed a recent course in curriculum, having completed a recent course in tests and measurement, how many times they taught Introduction to Computers in previous academic year, and how many years they taught Introduction to Computers. Responses to these variables were analyzed and are reported in Chapter Four.

Methodological Assumptions

According to Ary, Jacobs and Razavieh (1985), five assumptions must be met if resulting inferences from Chi-square analysis are to be considered valid:

1. Categories will be mutually exclusive; each response will appear in only one of the categories in a table.
2. The responses will be measured as frequencies.
3. The expected frequency in any cell will be at least five when the degrees of freedom equal one, and not less than five in 80 percent of the cells when the degrees of freedom are more than one.
4. The observed values with one degree of freedom will be

corrected for continuity to use the table of critical values of Chi-square.

5. The responses will be independent of one another.

Analytical Considerations

A major portion of the study involved analyzing platform unity. The theoretical underpinnings of unity serve as a methodological foundation. Before calculating unity, the principles that undergird Performance Instruction need to be established.

There are two elements needed to calculate platform unity--cognitive domain level codes and test type points. Simply stated, to achieve unity, the domain level code is matched with the identical level test type. However, the calculations were not always that easy especially when effectiveness and efficiency constructs are also desired. The next section explains the meaning behind the variables and how they interact in Performance Instruction.

Content Level Sorts in the Cognitive Domain

To help faculty clarify content intent in the planning stage, Vogler (1991) recommended sorting content at three levels ranging from simple to complex. The cognitive domain includes content goal statements intended to communicate some form of knowledge. The present tense action verb that begins each content goal is the key to sorting by domain in Performance Instruction. Although there might be thousands of

verbs available, Vogler (1991) specified that each content goal verb should lend itself to observation and measurement. As a consequence, the verb goal list is substantially reduced. Verbs such as "know," "gain," "learn," "understand," "appreciate," "foster," and "enhance" do not lend themselves to observation and measurement. A list of appropriate goal infinitives for the cognitive domain is provided in Table 3.

Table 3

Cognitive Domain Content Goal Verbs		
<u>Fact Domain Level One</u>	<u>Understanding Domain Level Two</u>	<u>Application Domain Level Three</u>
<u>define</u> indicate list <u>recall</u> reiterate specify state tabulate	select <u>classify</u> <u>compare</u> differentiate examine	formulate explain estimate develop <u>analyze</u> verify synthesize calculate validate apply critique modify organize <u>predict</u> prescribe solve

Note. Underlined verbs indicate those selected and included as content goal verbs in the Phase I survey instrument.

Source: Vogler, D. E. (1991). Performance instruction: Planning, delivering, evaluating. Eden Prairie, MN: Instructional Performance systems, Inc.

The cognitive domain in Vogler's model has three practical instructional decision levels coded level 1, 2, or 3. The levels include fact (level one or C1), understanding (level two or C2), and application (level three or C3). The letter "C" identifies the cognitive domain. The C1 level is a single concept and uses verbs like "define," "identify," and "list." The C2 level puts two or more concepts together. Typical verbs for this level; include "describe," "compare," and "contrast." The C3 level puts two or more concepts together to form something new. Typical verbs at this level include "explain," "apply," and "analyze." Verbs within the cognitive domain differ in levels of complexity. For example, the verb "define" is simpler than the verb "explain". Therefore, verbs used in writing content goals fall into certain categories where they can be matched to one of three decision levels. Table 4 shows these levels and the corresponding codes needed in the formulae calculations.

Table 4

Performance Instruction Domain Level Codes

<u>Cognitive Level</u>		<u>Cognitive Level Code</u>
Factual	(Level One)	1
Understanding	(Level Two)	2
Application	(Level Three)	3

The decision levels are important for course planning and testing. The levels selected by faculty when planning become important for selecting appropriate test types. Just as there are recommended verbs used at one of three cognitive planning decision levels, there are three corresponding levels used to test content. The expert system in Instructional Performance System's software recommends paired decision levels for planning and testing. In this study, using appropriate test types to test the planned content was termed platform unity, which in turn is directly related to the concept of curriculum alignment. By using Vogler's (1991) cognitive domain levels for content planning (C1, C2, C3), a numeric index can be constructed to determine appropriate test type selection. Here again, numeric values listed as points were needed for calculations, and Table 5 shows the test type points used.

The formulae were used to determine if planning and testing decisions were united. They were also used to determine effectiveness and efficiency of test type choices.

Table 5

Test Types and Associated Point Values

	<u>Test Types</u>	<u>Points</u>
Objective Tests:	True/False	1
	Matching	2
	Multiple Choice	3
Subjective Tests:	Completion	1
	Short Answer	2
	Essay	3

Planning and Testing Values

The point values associated with content planning and testing have to be determined before calculating unity. The values follow the same three decision levels presented earlier. Numeric values were based on difficulty and complexity. For example content taught, and tested, at the factual level is less difficult and complex than content taught at the application level. A true-false test, used to test facts, is less difficult than a multiple-choice test that requires students to select from one of four response items. Table 6 depicts the decision levels in the cognitive domain used to establish point values.

Table 6

Cognitive Test Item Types			
<u>Category of Test</u>	<u>Domain Level</u>	<u>Type of Test</u>	<u>Point Value</u>
Objective	1 (fact)	True-False	1
	2 (understanding)	Matching	2
	3 (application)	Multiple Choice	3

Subjective	1 (fact)	Completion	1
	2 (understanding)	Short Answer	2
	3 (application)	Essay	3

These point values, listed in Table 6, were used in the mathematical formula to determine platform unity. The larger the point value, the more complex the task; hence, the more complex the test type used to test content knowledge.

Calculating Unity

Formulae were designed to numerically analyze planning and testing decisions. The formulae used to establish the numeric indices for platform unity (PTU), effectiveness (EFT) and efficiency (EFF) are provided in Table 7.

Table 7

Unity, Effectiveness and Efficiency Formulae

Platform Unity or Planning Testing Unity (PTU)

$$PTU = \left(\sum \frac{\text{Test Type Points}}{\text{Number of Tests}} \right) - \text{Cognitive Domain Level Code}$$

Effectiveness (EFT)

$$EFT = \text{Cognitive Domain Level Code} - \left(\sum \frac{\text{Test Type Points}}{\text{Number of Tests}} \right) + |PTU|$$

Efficiency (EFF)

$$EFF = \text{Cognitive Domain Level Code} - \left(\sum \frac{\text{Test Type Points}}{\text{Number of Tests}} \right) - |PTU|$$

The PTU formula establishes various degrees of unity. Obtaining a zero through the formula equates to absolute platform unity, and means that the test types selected to test content match the level at which content was planned. Obtaining a unity index of +1.2 means a faculty member is testing closer to planned content than if the unity index was +2. The higher the value of the unity index, the less unity is exhibited. The sign of the index (positive or negative)

indicates that faculty were either overtesting content (positive) or undertesting content (negative). Overtesting occurred when the test type(s) used to test content represented a higher level of difficulty than the domain level established during content planning. Undertesting occurred when the test type(s) used represented a lower level of difficulty than the domain level established during content planning. The key to determining unity resided with the selection of the content goal verb used to plan content. The numerical value of the verb became the constant factor on which the determination of unity resided.

Effectiveness and Efficiency

Effectiveness and efficiency of test type selection represent two other constructs isolated through platform unity. Separate formulae, based on the unity index formula, were used to determine effectiveness and efficiency. Effectiveness and efficiency were tied to the domain level code established when content was planned, similar to the unity calculation. Faculty test effectively if the test type(s) selected to test content corresponded to the same or higher level used to plan content. Table 8 shows the effective test types corresponding to domain levels.

Table 8

Effectiveness Matrix

Domain Level Code	Point Values of Tests		
	1	2	3
C1	True-False Completion	Matching Short Ans.	Mult. Choice Essay
C2		Matching Short Ans.	Mult. Choice Essay
C3			Mult. Choice Essay

Faculty test efficiently if the test type(s) selected to test content corresponded to the same or lower levels used to plan content. Table 9 shows efficient test types corresponding to domain levels. It was possible to be effective and not efficient, or vice versa.

Table 9

Efficiency Matrix			
Domain Level Code	Point Values of Tests		
	1	2	3
C1	True-False Completion		
C2	True-False Completion	Matching Short Ans.	
C3	True-False Completion	Matching Short Ans.	Mult. Choice Essay

As with unity, a calculated value of zero assumed that faculty were effective and efficient in their test type selection. Degrees of effectiveness and efficiency were demonstrated numerically by the distance from zero. For example, an efficiency (EFF) index of -1 is more efficient than an EFF index of -1.5. The same is true for effectiveness indices, except that effectiveness indices were positive. Through the calculations, effectiveness indices were always positive values; efficiency indices were always negative values.

When examining unity, effectiveness and efficiency, if the result of the platform unity calculation were zero,

planning and testing decisions are united. A unity index of zero also signified that faculty test type selection was effective and efficient.

Further explanations of the indices and the information they yield are provided in Chapter Four. There were a possible combination of 63 test types per content goal. Appendix H contains a matrix of 189 indices calculated for test type combinations at fact, understanding, and application content levels. In Chapter Four, indices were calculated from actual content planning and testing decisions.

Summary

The research design addressed the procedural problem of developing and analyzing a numeric platform unity formula based on survey responses. This chapter illustrated how theoretical concepts of planning and testing were included in formulae to determine platform unity and related constructs. These constructs are overtesting and undertesting of course content, and effectiveness and efficiency of test type selection.

The sample identified in this study was computer faculty members. Data collection occurred through two phases of a mailed survey. The first phase instrument contained outcome-based content goal statements used to plan instruction. Faculty indicated if they planned and/or tested the content. If planned content were tested, faculty were asked to select

the test types they use to gauge students' content knowledge. The test types in this study included true-false, matching, multiple choice, completion, short answer and essay. One formula was used to determine the extent of platform unity between faculty planning and testing decisions. Two additional formulae examined the effectiveness and efficiency of faculty testing decisions.

The purpose of the second phase was to verify data obtained in the first phase. The second phase reported Phase I test type choices to respondents. Specifically, Phase II provided expert system recommendations along with Phase I responses. The expert system in this study represented absolute platform unity. Respondents were provided an explanation of how cognitive content planning decisions determined test type selection under conditions designed to preserve unity. After explaining this rationale, based on the expert system unity concept, Phase II respondents were asked to agree or disagree with the expert system recommendations. If there were certain situations where overtesting and undertesting prevailed, respondents were asked to indicate why they overtested or undertested content.

The instrument used in Phase I provided the researcher with an application of the platform unity concept, which was the purpose of the study. Phase II, a verification process, provided respondents an opportunity to state the purpose

behind their testing decisions if they disagreed with the expert system's recommendations. Data obtained in both survey phases were analyzed and are reported in Chapter Four.

CHAPTER IV

RESEARCH FINDINGS

The purpose of this study was to document the effects of applying the platform unity concept. This chapter addresses the principle research question--is there a difference between expert system and faculty test type selection? To address the purpose and research question, a group of faculty in the Virginia Community College System was surveyed. The data were collected in two phases via a mail questionnaire and tabulated using Number Cruncher Statistical System (NCSS) and dBase III Plus software systems. The Chi-square Goodness of Fit test was used to analyze the data to ascertain differences between faculty and expert system test-type selection in the first phase. Frequency of responses were noted in the second phase when faculty did not agree with the expert system recommendations.

This chapter also addresses the research questions--what test types do faculty use to test facts, understanding, and application? The research findings are presented in five sections. Section 1, Questionnaire Responses, discusses the participants and their demographics. Section 2, Phase I Findings, reports faculty planning and testing decisions. Results of the Chi-square Goodness of Fit test are reported in this section. Section 3, Analysis of Unity, Effectiveness, and Efficiency, provides the analysis of these constructs

based on formulae calculations. Decisions to overtest and undertest content are also detailed in this section. Section 4, Phase II Findings, reports on respondent verification of Phase I data through data collected in Phase II. The last section, Section 5, Summary, summarizes findings of this study.

Questionnaire Responses

Data were collected by using two survey questionnaires devised by the researcher. The instruments, referred to as Phase I and Phase II, were mailed to participants at two separate times. Appendix A contains the instrument used in Phase I. Appendix B contains the instrument used in Phase II.

Phase I Instrument

The Phase I survey was designed to solicit faculty planning and testing decisions. It contained six, actual content goals used in the course Introduction to Computers. Participants were asked to indicate whether or not they planned certain content goals on their course syllabus. Respondents then indicated if they tested this content. If content were tested, faculty listed the test types they used to evaluate content by selecting among six choices. Responses to these planning and testing questions were used to calculate indices for platform unity, effectiveness, and efficiency through formulae (explained in Chapter Three). Positive or negative values obtained through the calculations also

determined overtesting and undertesting situations.

Phase II Instrument

A second phase questionnaire verified the information obtained in Phase I. Faculty were provided the recommended test types according to the Performance Instruction model that contains the expert system recommendations. Respondents were asked if they agreed or disagreed with the model's decision. If they disagreed, they were asked to provide a reason.

Participants were limited to faculty members in the Virginia Community College System (VCCS) who teach the course Introduction to Computers. Fifteen of the 23 institutions were selected at random. Table 1 in Appendix G lists all 23 community colleges in Virginia. Table 2 in Appendix G lists the 15 randomly-selected community colleges that participated in this study.

Thirty full-time faculty, two from each of the 15 institutions, were randomly selected to participate. Telephone calls were placed to participating institutions to obtain the names of individual faculty teaching Introduction to Computers. Every institution selected had at least two faculty teaching Introduction to Computers. In institutions having only two computer faculty, both individuals were selected to participate in the survey.

A response rate of 100 percent was obtained from Phase I with one follow-up. All responses were deemed usable. In the

second phase, 28 out of 30 faculty returned a completed questionnaire, for a response rate of 93.3 percent. All responses in Phase II were deemed usable. A brief demographic section was included in Phase I. This information is reported in the following section.

Respondent Demographics

Demographic questions were limited to academic degrees, academic rank, the number of sections of Introduction to Computers taught during the 1991-92 academic year, the number of years respondents have taught the course, whether or not respondents had taken a curriculum course in the last three years, whether or not respondents had taken a tests and measurement course in the last three years, and years' teaching experience.

Ninety percent of the faculty surveyed hold master's and bachelor's degrees. Table 10 provides the proportion of degrees held. Two faculty members had attained only an associate's degree; one faculty member had a doctoral degree.

Table 10

Highest Academic Degree of Faculty	
Degree	Percent Holding Degree
Associates	6.7
Bachelors	30.0
Masters	60.0
Doctorate	3.3
N=30	

Table 11 reports respondent academic rank. Half of the respondents held academic rank at the associate professor or professor levels. In this study faculty academic rank did not contribute useful analytical data.

Table 11

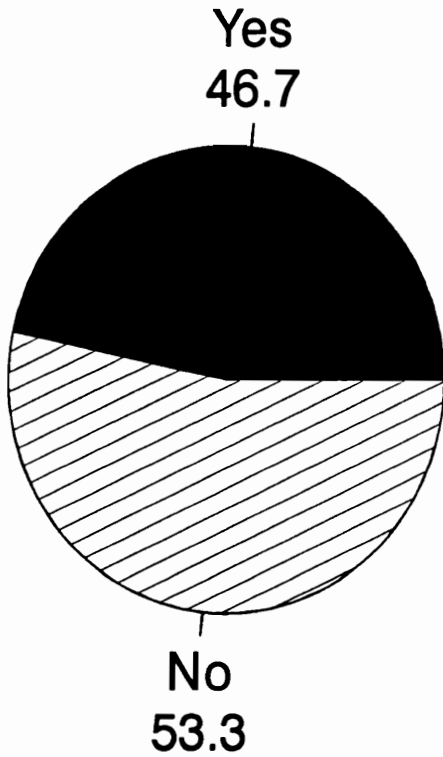
Academic Rank	
Academic Rank	Percent Holding Rank
Instructor	26.7
Assistant Professor	23.3
Associate Professor	40.0
Professor	10.0
N=30	

Respondents indicated their experience in teaching Introduction to Computers through responses to two questions.

Respondents taught a mean of 3.5 sections of Introduction to Computers in the 1991-92 academic year. The number of sections taught during this time ranged from 1 to 5. The mean number of years during which faculty taught Introduction to Computers was 5.3. This mean was negatively skewed as 70 percent of the respondents indicated they taught this course for more than six years. The category "more than six years" was the sixth and highest category provided for respondents on the survey instrument.

Curriculum and test and measurement issues were relevant in this study. Data indicate more respondents took a curriculum course than a course in tests and measurement in the past three years. The pie chart in Figure 2 graphically depicts these proportions. Approximately 47 percent of respondents took a curriculum course in the past three years; approximately 27 percent took a tests and measurement course in the past three years.

Taken Curriculum Course



Taken Tests and Measurement Course

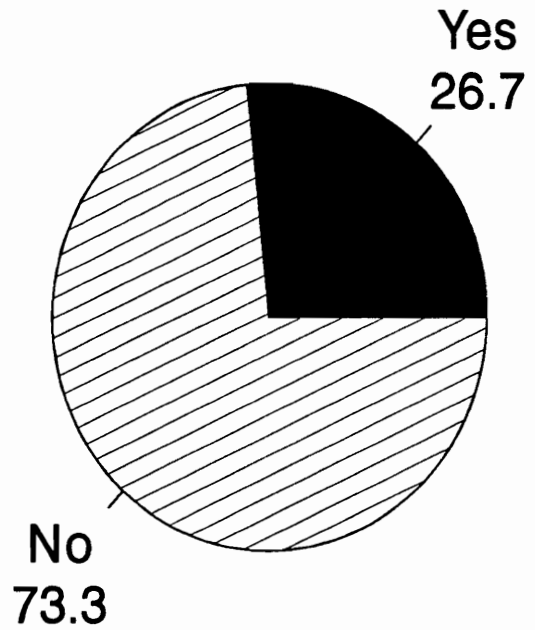


Figure 2

Enrollment In Courses In Past Three Years

Total teaching experience among the faculty cohort ranged from 1 year to 34 years with a mean of 14.23 years. The median was 13.5 years. Total teaching experience was determined by summing full-time years in college, graduate school, high school, and business and industry teaching positions. The largest category of respondent teaching experience was college with a mean of 11.8 years. No respondent had elementary school teaching experience. The number of years faculty reported holding their present position ranged from 1 to 28 years; the mean was 10.7.

Faculty characteristics were examined for relationships using correlation statistics. The characteristics analyzed included faculty rank, highest academic degrees earned, years teaching experience, and years in present teaching position. No significant relationships were found. In general, the typical respondent held a master's degree, was an associate professor, taught four sections of Introduction to Computers each academic year, had 14 years teaching experience, and held his/her present teaching position for 11 years.

Phase I Findings

The survey included six content goals, two from each cognitive domain. These content goals and the expert system domain level classifications are shown in Table 12.

Table 12

Content Goals and Domain Level

<u>Survey Item Number</u>	<u>Content Goal</u>	<u>Domain Level</u>
1	Define input device terms	One
2	Compare microcomputer software applications	Two
3	Analyze system design components	Three
4	Recall flowchart symbols	One
5	Classify computer operations	Two
6	Predict secondary storage requirements	Three

According to Vogler's (1991) expert system recommendations, factual content should be tested by true false and/or completion items; understanding content should be tested by matching and/or short answer items; application content should be tested by multiple-choice and/or essay items. Table 13 shows the domain levels and the test items that are considered congruent with cognitive content.

Table 13

Cognitive Test Item Types

<u>Domain Level</u>	<u>Test Items</u>
1 Fact	True-False Completion
2 Understanding	Matching Short Answer
3 Application	Multiple Choice Essay

Planning and Testing Decisions

Faculty indicated planning and testing decisions through the survey instrument. Each instrument contained six content goals. The population included 30 respondents. Therefore, a total of 180 content goals were provided. Out of 180, faculty did not plan and test 57 content goals. The remaining 123 content goals were analyzed.

Table 14 reports faculty rates of planning and testing decisions by content goal. The rates should be identical between the planned and tested categories if alignment principles and platform unity were followed.

Table 14

Planning and Testing Decisions by Content Goal

<u>Content Goal</u>	<u>Percent of Faculty Plan Content</u>	<u>Percent of Faculty Test Content</u>
1	80.0	96.7
2	76.7	80.0
3	66.7	63.3
4	40.0	40.0
5	76.7	76.7
6	40.0	43.3

N=30

Figure 3 graphically depicts the data from Table 14. Differences between planning and testing are apparent. Differences are attributed to misalignment. Misalignment or disunity occurred when planned content was not tested or when tested content was not planned.

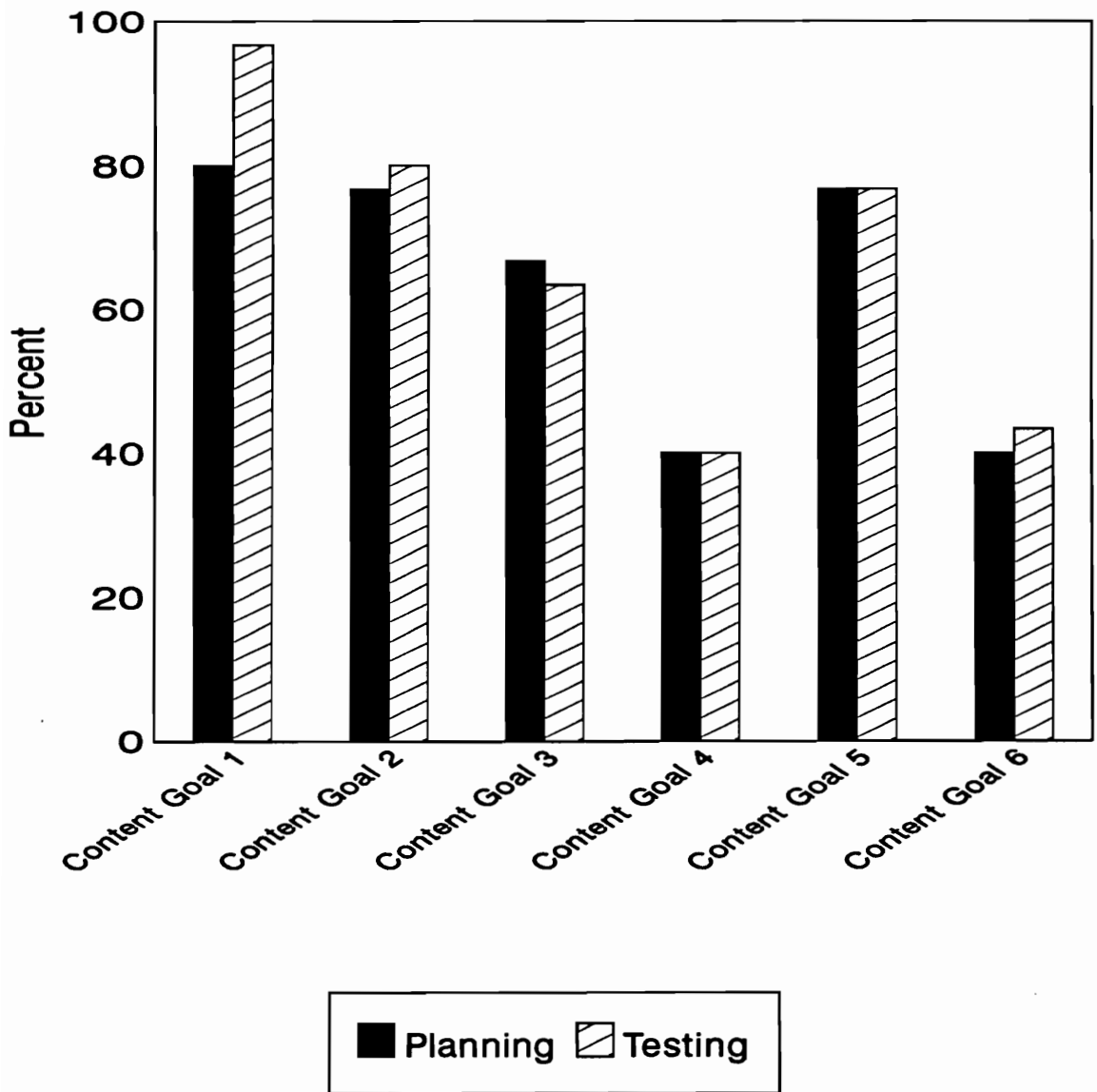


Figure 3

Faculty Planning and Testing Decisions

The rates of planned and tested content vary; they are all less than 100 percent. These rates indicate important differences. Differences result since not all faculty planned and tested the content provided as content goals on the survey. These data indicate that alignment between planning and testing occurred with content goal four and five. However, since decisions to plan but not test content cancel out decisions to test but not plan content, a closer examination was needed. This topic is discussed thoroughly in a later section dealing with an analysis of disunity.

Respondents were asked to make content planning and testing decisions based upon the content goal statements displayed in Table 12. Table 15 shows the six test item types and those that respondents selected per content goal.

Table 15

Faculty Test Type Selections

<u>Content Goal</u>	<u>Test Item Types</u>					
	<u>True False</u>	<u>Matching</u>	<u>Multiple Choice</u>	<u>Completion</u>	<u>Short Answer</u>	<u>Essay</u>
1	8	11	23	3	5	3
2	6	6	16	2	9	6
3	3	1	11	2	7	7
4	5	4	8	1	3	0
5	7	6	19	1	7	3
6	<u>3</u>	<u>4</u>	<u>8</u>	<u>0</u>	<u>6</u>	<u>2</u>
Total	32	32	85	9	37	21

N=30

Multiple-choice test item types surpass the others in terms of their use. Figure 4 graphically presents the test type choices across all six content goals.

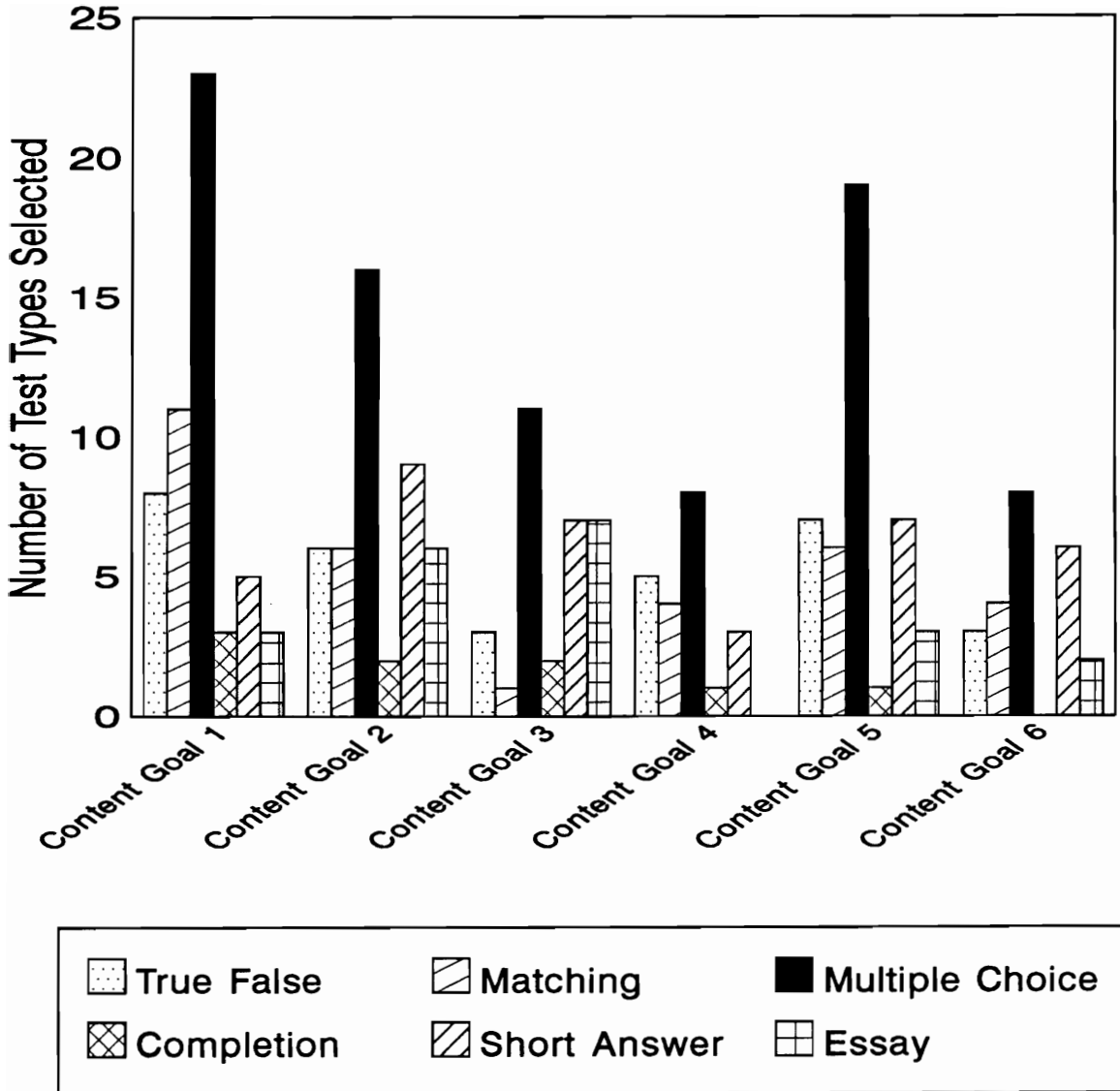


Figure 4

Phase I Test Type Choices

Table 16 reports the number of observations of each test type selected by respondents. The table also has the test types listed in order of their frequency of selection by faculty members. The mean is the number of observations divided by the number of respondents. The Kruskal-Wallis test was used to break ties.

Table 16

Ranked Preference of Test Types

<u>Test Type</u>	<u>Total Observations</u>	<u>Mean</u>	<u>Rank</u>
Multiple Choice	85	2.83	1
Short Answer	37	1.23	2
True False	32	1.07	3
Matching	32	1.07	4
Essay	21	0.70	5
Completion	9	0.30	6

N=30

Note. Kruskal-Wallis test used to adjust for ties.

On the survey, respondents could select as many of the six test item types as they desired. The six test item types are broken down into two groups. True-False, matching, and multiple-choice tests belong to a group of objective tests. Completion, short-answer and essay tests belong to a group of subjective test types.

According to the test type rankings, faculty preferred to

test content provided on the survey by using objective test types over subjective test types. The combined mean for objective tests was 4.97; the mean for subjective test types was 2.23.

Domain-referenced content was presented at three levels--fact, understanding, and application--on the survey. The test types that faculty selected to evaluate each content goal are presented in Tables 17, 18, and 19. A graphic representation of test type selection is provided in Figures 5, 6, and 7. At each domain level, multiple-choice test items are preferred.

What are faculty test type choices?

Data presented in Table 17 addressed the second research question: **What test types are used to test facts?** Although respondents chose all test types to test factual content, the majority selected multiple-choice tests.

Table 17

Faculty Selected Test Types To Test Facts

<u>Content Goal</u>	<u>Test Item Types</u>						
	<u>True False</u>	<u>Matching</u>	<u>Multiple Choice</u>	<u>Completion</u>	<u>Short Answer</u>	<u>Essay</u>	
1	8	11	23	3	5	3	
4	5	4	8	1	3	0	
N=30							

Data presented in Table 18 addressed the question: **What test types are used to test understanding?** Although respondents chose all test types to test understanding content, the majority selected multiple-choice tests.

Table 18

Faculty Selected Test Types To Test Understanding

<u>Content Goal</u>	<u>Test Item Types</u>					
	<u>True False</u>	<u>Matching</u>	<u>Multiple Choice</u>	<u>Completion</u>	<u>Short Answer</u>	<u>Essay</u>
2	6	6	16	2	9	6
5	7	6	19	1	7	3
N=30						

Data presented in Table 19 addressed the question: **What test types are used to test application?** Although respondents chose all test types to test application content, the majority selected multiple-choice tests.

Table 19

Faculty Selected Test Types To Test Application

<u>Content Goal</u>	<u>Test Item Types</u>					
	<u>True False</u>	<u>Matching</u>	<u>Multiple Choice</u>	<u>Completion</u>	<u>Short Answer</u>	<u>Essay</u>
3	3	1	11	2	7	7
6	3	4	8	0	6	2
N=30						

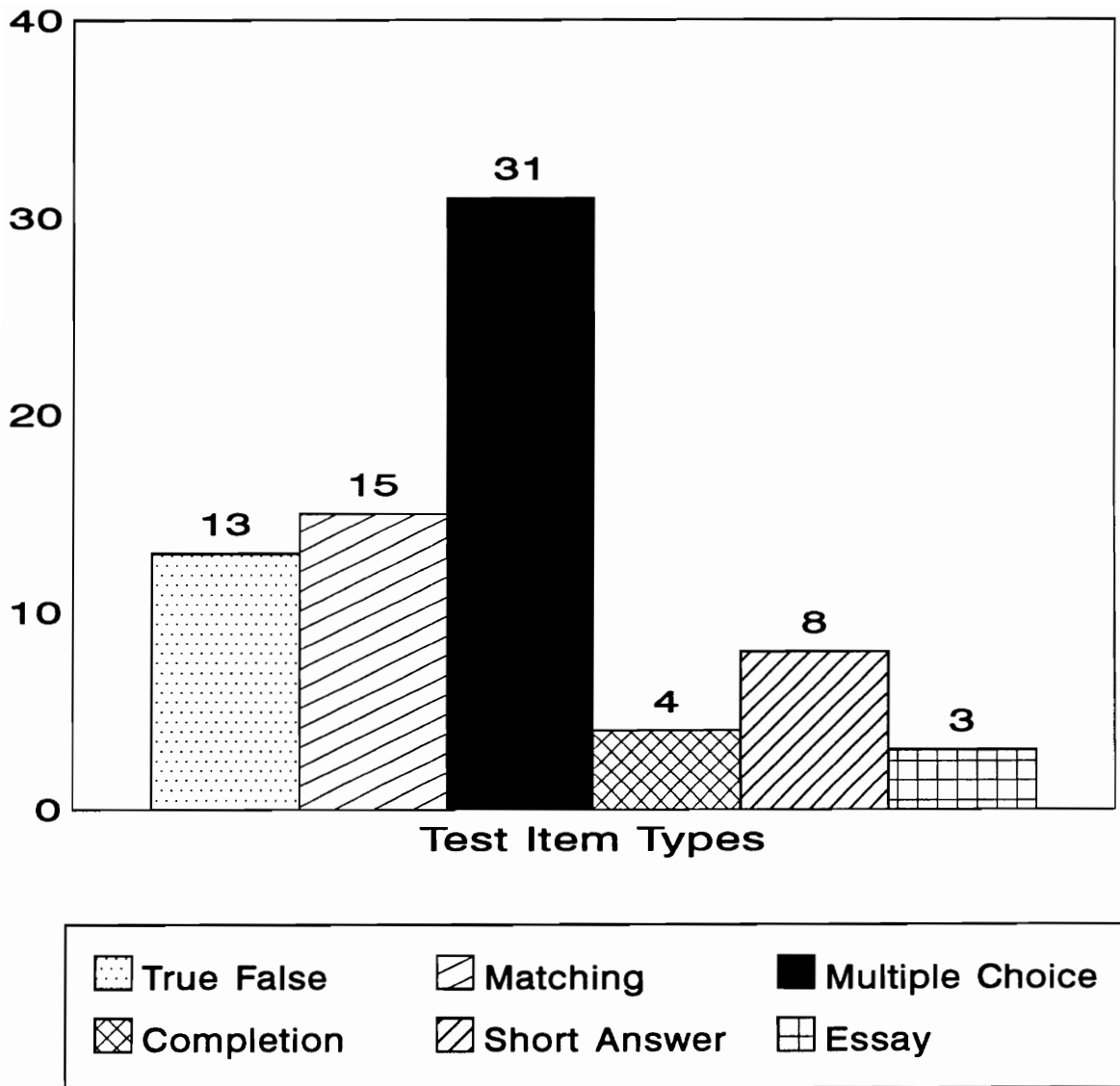


Figure 5

Selected Test Item Types To Test Facts

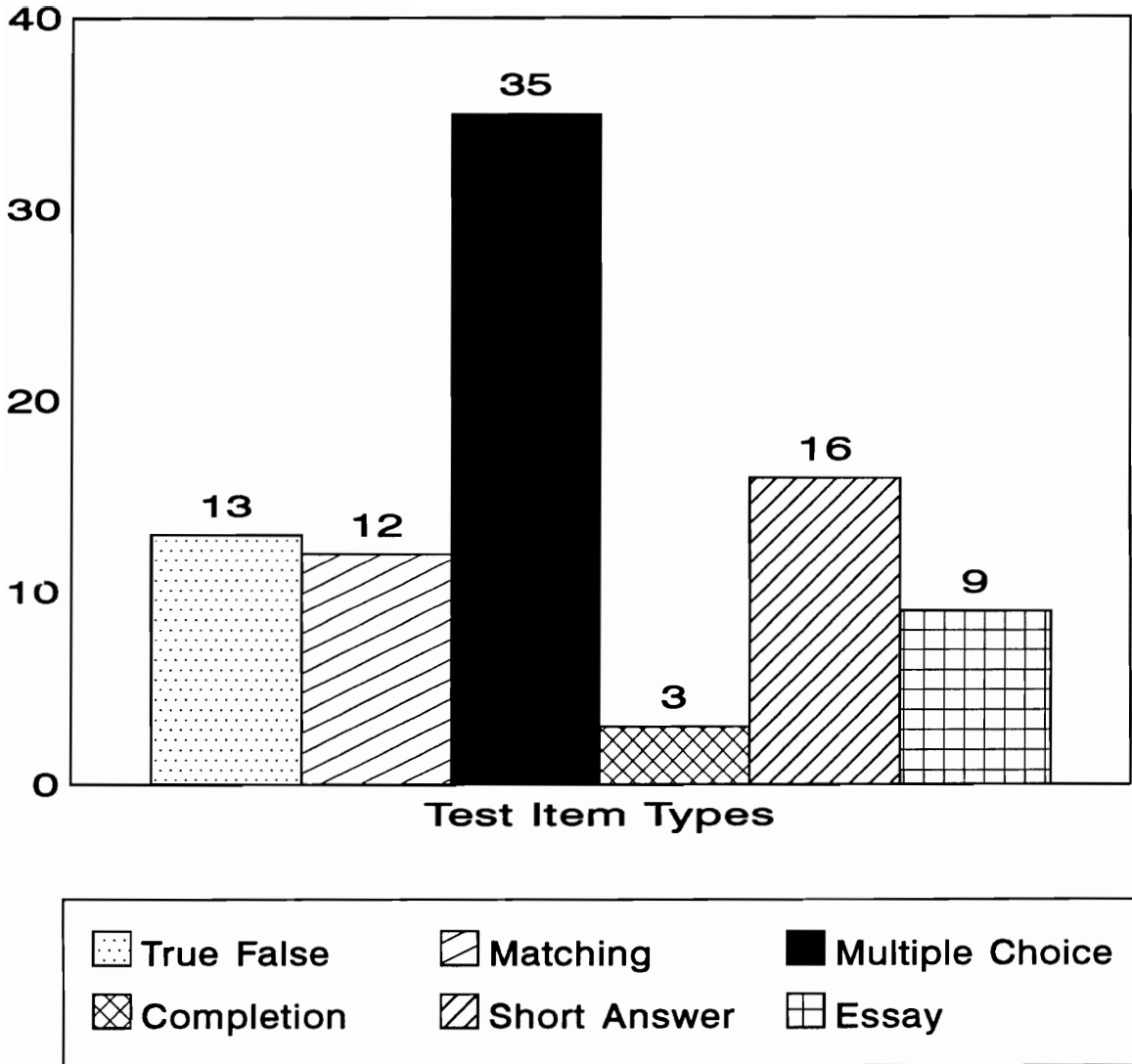


Figure 6
**Selected Test Item Types
 To Test Understanding**

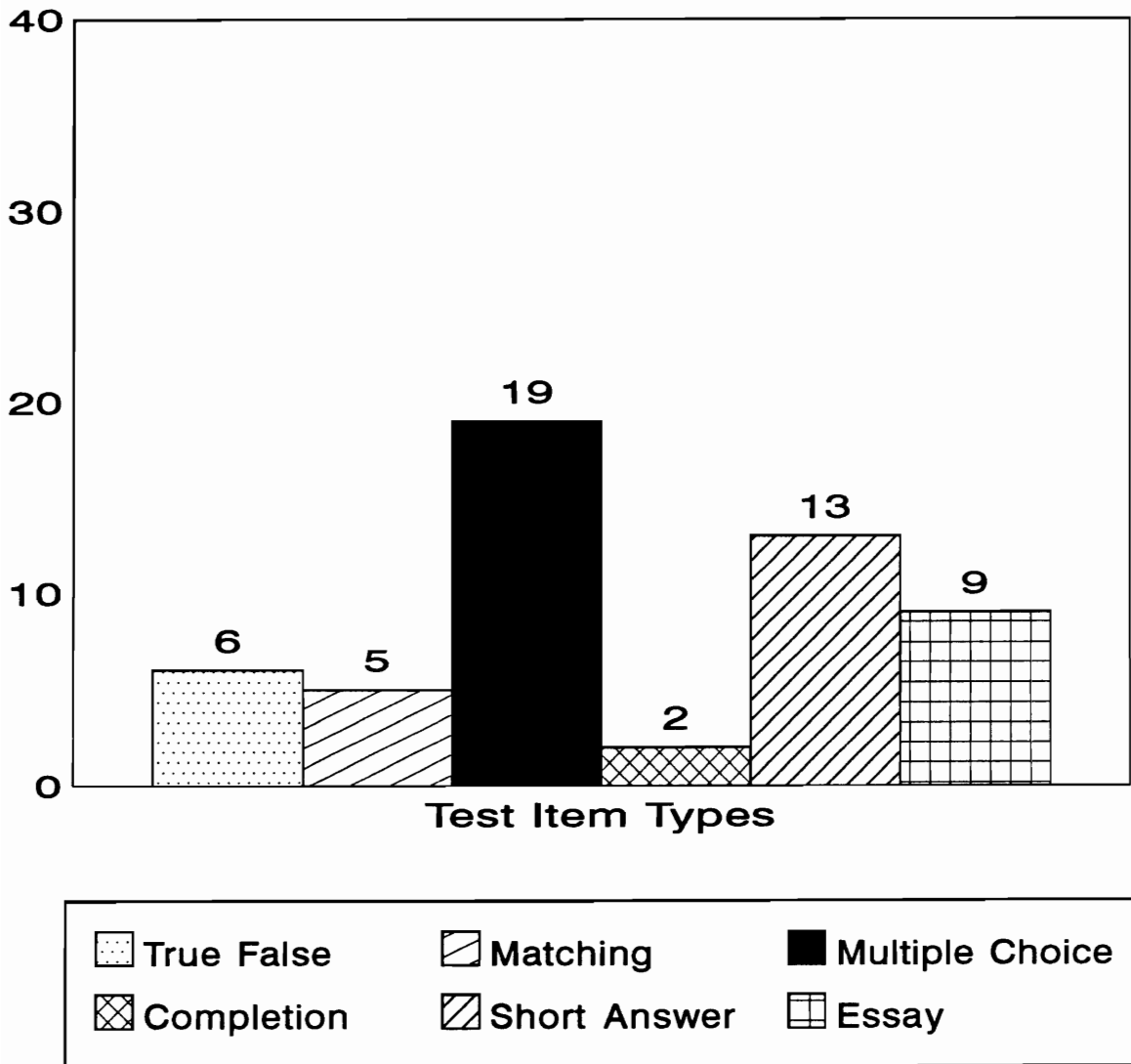


Figure 7

Selected Test Item Types To Test Application

Faculty and Expert System Test Type Differences

Chi-square statistical analysis was used to answer the research question: Is there a difference between expert system and faculty test type selection? Specifically, the Chi-square Goodness of Fit Test was used to analyze data for determining difference between faculty and expert system test type selection. Frequencies were significantly different if the probability of obtaining the computed Chi Square by chance were less than five percent.

To determine the Chi-square value for each content goal, observed frequencies of occurrence provided by faculty responses were compared to theoretical or expected frequencies provided by the expert system. Each content goal was analyzed by using the number of test item types checked by respondents. Expected values followed the expert system's recommendations. For example, if the content goal level were one, the expected test types corresponding to a level one content goal include true-false and completion. If faculty marked test item types other than the domain-level matched test types, the responses were collapsed into a category termed "other."

Expected frequencies were determined by counting the test item types checked that correspond to expert system recommendations. Observed frequencies were determined by all the other test item types checked. A margin of error of five percent was designed in the calculations.

Table 20 shows Chi-square calculations for each content goal.

Table 20

Differences Between Faculty and Expert System
Test Type Selections

<u>Content Goal</u>	<u>Test Types</u>		<u>Chi-Square</u>
1	Observed	TF and Comp. Other 11 42	615.06*
	Expected	50 3	
2	Observed	Mat & Sh/Ans. Other 15 30	360.26*
	Expected	43 2	
3	Observed	MC and Essay Other 18 13	89.03*
	Expected	29 2	
4	Observed	TF and Comp. Other 6 15	195.09*
	Expected	20 1	
5	Observed	Mat & Sh/Ans. Other 13 30	379.74*
	Expected	41 2	
6	Observed	MC and Essay Other 10 13	128.53*
	Expected	22 1	

* $p < .01$
df = 1

Note. TF = True-False; Mat = Matching; MC = Multiple Choice; Comp. = Completion; Sh/Ans. = Short Answer.

In Table 20, significant differences ($p < .01$) were obtained for all content goals. Statistically, there is less than a one percent chance that the different responses occurred by chance. Differences are apparent between expert system test type recommendations and the test types faculty use to test content. The large Chi-square values obtained do not indicate a "good fit" between faculty and expert system test type choices.

Analysis of Unity, Effectiveness, and Efficiency

The purpose of this study was to document the effects of applying the platform unity concept. A survey provided data from which calculations were made. The survey instrument contained six actual content goals that allowed faculty to indicate realistic content-planning and testing decisions. Five constructs were examined with each faculty response. Responses were coded and analyzed through specially-developed formulae. The formulae, introduced in Chapter Three, produced indices to enable the researcher to examine the constructs.

The range of index values was essential to interpreting obtained calculated indices produced by the formulae. Table 21 shows each construct and the range of values. Obtaining a value of zero for unity, effectiveness, and efficiency indices was desirable. Obtaining a zero meant that faculty plan and test congruently; they chose effective and efficient methods of testing planned content. The sign indicated an overtesting

and undertesting situation. When the platform unity value was positive, faculty overtested content. Similarly, when the platform unity value was negative, faculty undertested content.

Table 21

Range of Values Produced By Formulae

<u>Construct</u>	<u>Range of Values</u>
Platform Unity*	-2 to +2
Effectiveness	0 to +4
Efficiency	-4 to 0

*Positive values occur when faculty overtest content.
 Negative values occur when faculty undertest content.

Platform Unity

Platform unity was the specific alignment of planned, domain-referenced content with associated test types. To achieve unity, planned content must be tested with items that match the cognitive domain level of the planned content. When this occurred, the platform unity index was zero.

Platform unity was calculated for all usable responses (n=123). A response was deemed usable when faculty indicated they planned and tested content. A total of 123 content goals were deemed usable responses in the survey out of 180 total

content goals as 57 goals were neither planned nor tested. Unity between planning and testing decisions, or platform unity, was obtained in 32 content goals, for a platform unity rate of 26 percent. The mean platform unity index for all content goals was .64. Considering that absolute unity is zero, this rate indicated the tendency to overtest content. This value was positive which suggested that, on average, faculty overtest content.

Overtesting Content

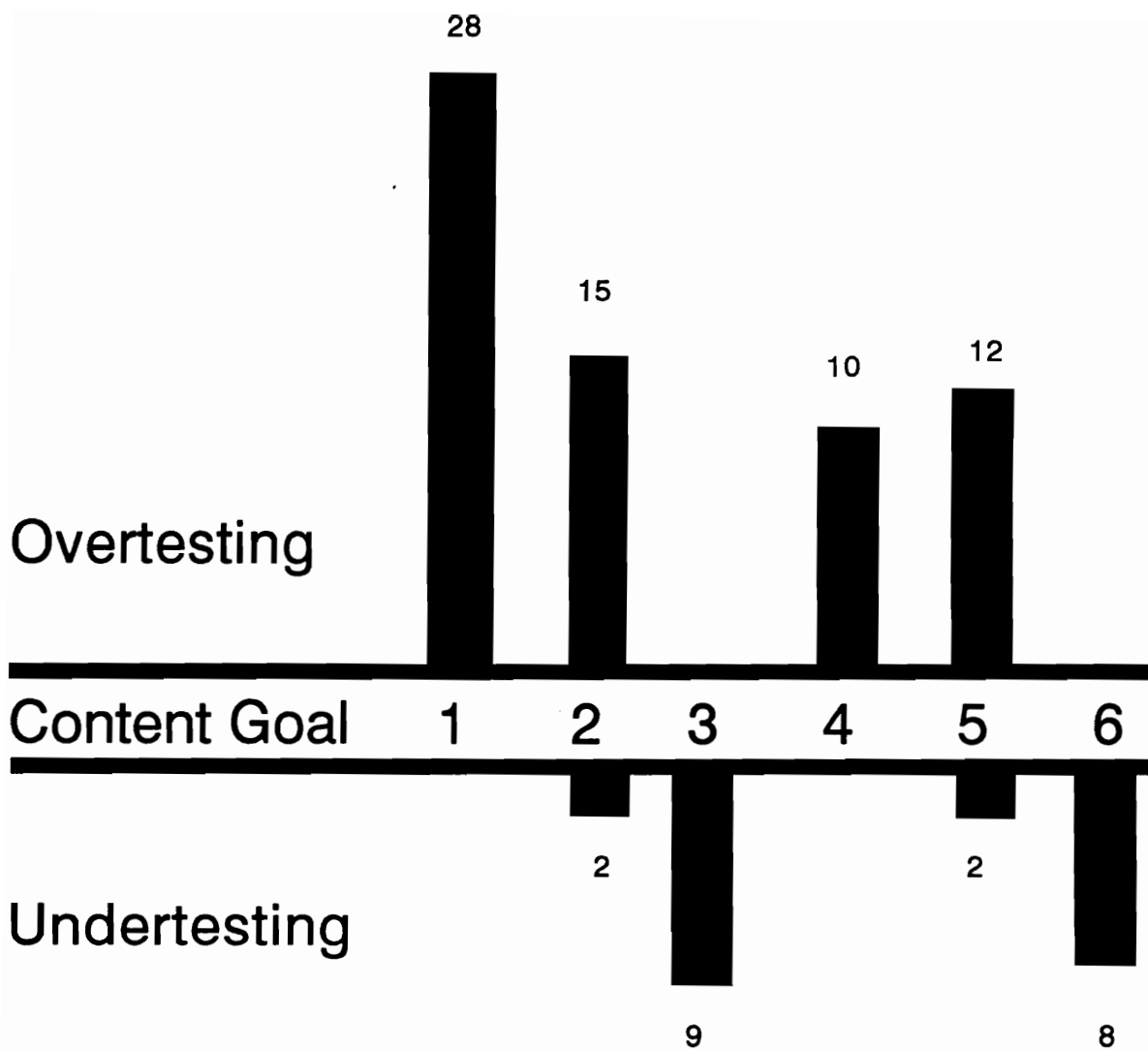
Overtesting was defined in this study as faculty selection of test types that test content at a higher level than the domain level on which content plans were established. If the calculated platform unity index resulted in a positive number, faculty overtested content. From their survey responses, 73.3 percent of faculty had positive platform unity indices. Out of 123 content goals analyzed, 68 or 55.3 percent were overtested.

Undertesting Content

Undertesting was defined in this study as faculty selection of test types that test content at a lower level than the domain level on which content plans were established. If the calculated platform unity index resulted in a negative number, faculty undertested content. From survey responses, 13.3 percent of faculty had negative platform unity indices. Out of 123 content goals analyzed, 23 or 18.7 percent were

undertested.

Faculty overtesting and undertesting decisions were examined for each content goal. Figure 8 reveals the number of faculty in Phase I who overtested and undertested content for each content goal. For content goal one, all but two faculty members selected test types that resulted in overtesting situations. Half of the faculty overtested content goal two; two faculty undertested it. The numbers reported in Figure 8 indicate the faculty's tendency to overtest content.



N = 30

Figure 8
 Overtested And Undertested Content
 Per Content Goal

Effectiveness of Test Type Selection

In this study effectiveness referred to selecting test item types that match or exceed the planned content level. These test items required more complex student responses. When the platform unity index was zero, the effectiveness index was also zero, and an effective means of testing planned content. Faculty selected effective test types in 100 or 81 percent of the content goals.

Efficiency of Test Type Selection

In this study efficiency referred to selecting test item types that match or fail to reach the planned content level. It represented the easiest way of soliciting a student response. When the platform unity index was zero, the efficiency index was also zero, meaning that the test type selected was an efficient means of testing the planned content. Faculty selected efficient test types in 55 or 48 percent of the content goals.

Using the formulae displayed and discussed in Chapter Three, indices for platform unity, effectiveness, and efficiency are shown in Table 22. The indices were derived by mean calculations from 123 usable content goals. The range for each index was also provided. The farther away the mean index was from zero, the less united, less effective, and less efficient it was. Theoretically, effectiveness values are always positive; efficiency values are always negative.

Table 22

Unity, Effectiveness, and Efficiency Means				
<u>Index</u>	<u>Range of Values</u>			<u>Calculated Mean</u>
Platform Unity	-2	to	+2	0.64
Effectiveness	0	to	+4	0.33
Efficiency	-4	to	0	-1.44
N=123 content goals				

Table 23 reports the construct indices for each specific level of content. The ranges provided in Table 22 are applicable with the indices displayed in Table 23.

Table 23

Unity, Effectiveness, and Efficiency Means For Levels of Cognitive Content				
<u>Index</u>	<u>Content Level</u>	<u>Calculated Mean</u>	<u>N</u>	
Platform Unity	1 - Fact	1.47	42	
	2 - Understanding	0.44	48	
	3 - Application	-0.39	33	
Effectiveness	1 - Fact	0.00	42	
	2 - Understanding	0.21	48	
	3 - Application	0.91	33	
Efficiency	1 - Fact	-2.76	42	
	2 - Understanding	-1.27	48	
	3 - Application	0.00	33	

Theoretically, faculty are always effective with level-one content regardless of the test type used. They are always efficient with level-three content regardless of the test type used. In theory, since one cannot undertest level one content, it was always effective. Similarly, in theory, since one cannot overtest level-three content, it was always efficient.

In Table 24, data were displayed per level of cognitive content per grouped content goals. The indices for unity and effectiveness were high with level-one content. This corresponds with a tendency to overtest level-one content. The indices were moderate for understanding content and low for application content.

Table 24

Unity, Effectiveness, and Efficiency Indices
Per Content Goal

Content Goal	Domain Level	N	Unity	Indices	
				EFT	EFF
1 and 4	Fact	42	1.47	.095*	-2.76
2 and 5	Understanding	48	0.44	.210	-1.27
3 and 6	Application	33	-0.39	.910	-0.12**

Note:

*effectiveness assigned due to disunity (otherwise, test type is always effective at level one)

**efficiency assigned due to disunity (otherwise, test type is always efficient at level three)

Considering the platform unity range goes from a minimum value of zero to a maximum value of +2 on the positive side, the unity index was rather large for factual content. Out of the three levels, factual content was aligned to a lesser degree than understanding or application content. The zero values were replaced with small indices for effectiveness (EFT) at the factual level and for efficiency (EFF) at the application level. As noted at the bottom in Table 24, these two values were assigned, not calculated. Respondent disunity prompted the researcher to assign unity, effectiveness, and efficiency values for individual content goals.

Disunity

A situation termed, "absolute disunity," occurred if faculty planned but did not test content or if faculty tested but did not plan content. Out of 123 content goals, disunity occurred in 12 or 9.8 percent of the content goals. Nine faculty or 30 percent of total respondents presented decisions that represent absolute disunity.

Six out of the nine faculty, or 66.7 percent, reported testing content without planning it, resulting in an overtesting situation. Three, or 33.3 percent, planned content without testing it, resulting in an undertesting situation. Table 25 shows respondents having disunited planning and testing decisions on each content goal.

Table 25

Disunited Planning and Testing Decisions

<u>Content Goal</u>	<u>Respondent</u>	<u>Decision</u>	<u>Platform Unity Index*</u>
1	#9	Test but not Plan	+3
	#18	Test but not Plan	+3
	#21	Test but not Plan	+3
	#26	Test but not Plan	+3
	#27	Test but not Plan	+3
2	#18	Test but not Plan	+3
3	#6	Plan but not Test	-3
4	#18	Test but not Plan	+3
	#28	Plan but not Test	-3
5	#5	Plan but not Test	-3
	#14	Test but not Plan	+3
6	#21	Test but not Plan	+3

*Platform unity index assigned.

If faculty tested content that was not planned (overtesting), a value of +3 was assigned to the decision. Conversely, if faculty planned content but did not test it (undertesting), a value of -3 was assigned. When platform unity was +3, EFT and EFF indices were assigned as zero and -4 respectively. When platform unity was -3, EFT and EFF indices were assigned as +4 and zero respectively.

Phase I Summary

Although the mean unity index for all respondents was .64, responses to individual survey questions varied widely. Some faculty planned content but did not test it; some tested content that was not planned. Many faculty testing decisions favored overtesting of course content. Therefore, indices measuring effectiveness tended to be close to zero while indices measuring efficiency tended to be the most distant value from zero.

Total years teaching experience and years teaching Introduction to Computers had no relationship to the constructs examined in the study. Although most faculty held master's degrees, academic degree had no correlational relationship to the constructs examined.

Despite fourteen respondents having a recent curriculum course and despite eight respondents having taken a recent course in tests and measurement, the unity indices did not differ significantly from all other respondents. However, faculty with high school teaching experience had higher than average unity indices.

Most faculty supported planning and testing content. Of 180 content goals, faculty neither planned nor tested 57 of them. From 123 content goals, only 12 had assigned unity, effectiveness, and efficiency values. This occurred through a situation termed absolute disunity, a condition where

testing occurred without planning or planning occurred without testing. As presented in Table 25, testing unplanned content occurred in only 9 out of 123 content goals.

Table 14 contains a summary of planning and testing decisions by content goal. When there was not unity between planning and testing decisions, overtesting or undertesting occurred. Theoretically, level-one content is always united or overtested and level-three content is always united or undertested. Level one is the lowest level content; it cannot be undertested. Level-three content is the highest; it cannot be overtested. Level-two content, on the other hand, can go either way. There is an equal chance of overtesting and undertesting content at level two. Participating faculty's propensity to overtest content is evident when examining their level-two content testing decisions. Twenty seven overtested this level while four undertested content. Overall, 73.3 percent of respondents had positive platform unity indices. This illustrates the extent of overtesting that occurred in this study. By far, multiple-choice test items exceeded other types in their frequency of use. Faculty selected multiple choice as their first choice of tests at all domain levels of content.

The platform unity index with factual content was most distant from zero, but closest to zero with application. This means faculty test types were more closely aligned with the

content plans at the application level than the factual level. The effectiveness indices were all relatively close to zero meaning that faculty are effective in their test type selection. Effectiveness in this case was a function of overtesting content. The efficiency index for factual content was distant from zero. Faculty are woefully inefficient in their selection of test types, meaning that content testing exceeded what would be considered an appropriate test type matching the planned level of content. An overtesting situation contributed to inefficient testing. In this study, Phase I data supported that faculty sacrificed efficiency in favor of effectiveness.

The main statistical test in this study compared faculty responses to the expert system by using Chi-square Goodness of Fit. The research hypothesis was that faculty choices to test item types will differ from expert system test type recommendations. The result yielded statistically significant differences across all content goals.

Phase II Findings

The survey's second phase was designed to verify Phase I information by obtaining reactions to the expert system recommendations. This phase reported responses from the first phase with a brief description of Performance Instruction principles. An illustration served to reinforce the concept. After providing participants with test type selection data,

they were asked to agree or disagree with the expert system's recommendations. If faculty disagreed, they were asked to select one of six options to explain why they did not agree.

A copy of the survey instrument used for Phase II is provided in Appendix B. If the options provided on the instrument did not accurately frame faculty reasons for disagreeing, space was provided to list other reasons for disagreeing with the expert system. The six reasons provided on the instrument included the following:

- a. Want to purposely undertest content.
- b. Want to purposely overtest content.
- c. Want to guarantee employers that student have marketable skills when they graduate.
- d. Want to keep students in program.
- e. Personal dislike for above listed test types.
- f. Difficulty with terminology.

There were 28 Phase II respondents. Each respondent examined test type data from six content goal statements. Therefore, 168 content goals were examined. Of 168 content goals, faculty agreed with the expert system 107 times or 63.7 percent. Table 26 shows the content goal and the number of faculty agreeing and disagreeing with the expert system recommendations. The content goal statements examined in Phase II were identical to the statements in Phase I.

Table 26

Faculty Responses To Expert System Recommendations				
<u>Content Goal</u>	<u>Number of Faculty Who Agree</u>	<u>%</u>	<u>Number of Faculty Who Disagree</u>	<u>%</u>
1	9	32.1	19	67.9
2	18	64.3	10	35.7
3	25	89.3	3	10.7
4	11	39.3	17	60.7
5	22	78.6	6	21.4
6	22	78.6	6	21.4
N=28				

Agreement With Expert System

As in Phase I, the content goal statements were developed from the three levels in the cognitive domain. Content goal statements one and four were factual content; statements two and five were understanding content; and, statements three and six were application content. The three levels of cognitive content are shown in Table 27. The percentage rates appearing in the third column show the rates at which respondents agreed with the expert system test type recommendations for each cognitive domain level.

Table 27

Respondents Agreeing With Expert System
For Cognitive Domain Level

<u>Level</u>	<u>Content Goals</u>	<u>Percent Agreeing</u>
1	1 and 4	35.7
2	2 and 5	71.4
3	3 and 6	83.9

N=28

Disagreement With Expert System

Respondents agreed with the expert system recommendations on 63.7 percent of the content goals. They disagreed with expert system recommendations on 36.3 percent of the content goals. The information presented below deals with those respondents who disagreed with the expert system test type recommendations.

Table 28 reports the six reasons listed on the survey along with the number and percent of respondents who selected the reason. The information reported in Table 28 counted all reasons checked by respondents.

Table 28

Reasons For Disagreeing With Expert System Test Types

<u>Reason</u>	<u>Number Disagreeing</u>	<u>Percent Disagreeing</u>
a. Want to purposely undertest content	3	1.8
b. Want to purposely overtest content	3	1.8
c. Want to guarantee employers that students have marketable skills when they graduate	20	11.9
d. Want to keep students in program	0	0.0
e. Personal dislike for above listed test types	29	17.3
f. Difficulty with terminology	9	5.4

N=168 content goals

Table 29 presents respondent reasons for disagreeing with the expert system test type recommendations for each cognitive level. The levels included fact, understanding, and application. The reasons listed in Table 28 were abbreviated in Table 29 by only listing the letter of the reason. Only those reasons provided as options on the instrument were listed in the table. There were no written, "other" reasons summarized here. Reason D, "want to keep students in program," was not checked by respondents in any

domain level.

Table 29

**Reasons For Disagreeing With Expert System Test Types
For Factual, Understanding, and Application Content**

R E A S O N	Fact		Understanding		Application	
	<u>Number</u> <u>Disagree</u>	<u>Percent</u> <u>Disagree</u>	<u>Number</u> <u>Disagree</u>	<u>Percent</u> <u>Disagree</u>	<u>Number</u> <u>Disagree</u>	<u>Percent</u> <u>Disagree</u>
A	1	1.6	0	0	2	3.3
B	0	0	2	3.3	1	1.6
C	12	18.8	6	9.4	2	3.3
D	0	0	0	0	0	0
E	20	31.3	9	14.1	0	0
F	2	3.3	4	6.3	3	4.7

N=64 Reasons for Disagreeing

Respondents had the opportunity to pencil in "other" reasons for disagreeing with expert system test type recommendations. Respondent written comments were limited. Often, respondents reinforced a checked reason. When this occurred, the researcher elected to record the prepared response instead of the respondent's written response. Three respondents included written reasons that were considered

substantive. For factual content, respondent #15 said "this content has limited value; it can be undertested." Respondent #24 disagreed with the expert system stating the desire for students to incorporate writing skills. On another factual item, respondent #24 preferred to have students complete a sample problem, such as a case study.

There were no "other" reasons given for level two, understanding content. For level three, application content, respondent #13 preferred to test the concept with matching or short-answer items. The respondent wanted students to briefly describe the situation without having to write an essay.

Phase II Summary

In Phase II, respondents agreed with the expert system test type recommendations in 64 percent of the content goals. Table 27 reported respondent agreement with the expert system per level of cognitive content. The lowest percentage of agreement occurred with factual content. A larger proportion of respondents agreed with the expert system at the understanding and application levels.

Disagreeing with the expert system's recommendations for lower level content supports faculty's earlier decisions to overtest content. At lower cognitive levels, if respondents disagreed with the expert system, they favor testing with more effective test types, creating a more accurate overtesting situation. The lower rate of respondents who disagreed with

expert system test type recommendations at higher levels of cognitive content supported the expert system recommendations at higher levels of cognitive content.

Table 28 reported reasons why respondents disagreed with the expert system. Most often reported was "personal dislike for test types," followed by "want to guarantee employers that students have marketable skills when they graduate". "Difficulty with terminology" was the third highest reason. Here, terminology appeared to be related to the content goal statement. Several respondents indicated that they disagreed with the content goal verb used to plan instruction. Two other reasons, "want to purposely undertest content" and want to purposely overtest content" were ranked fourth with only 1.8 percent of the respondents selecting these reasons. The reason "want to keep students in program" was not selected by any respondent.

Summary of Results

The purpose of this study was to document the effects of applying the platform unity concept. Research findings and data analysis from respondents were presented in this chapter. Responses to survey questions were numerically analyzed through formulae.

Phase I data indicate statistically significant differences between expert system and faculty testing decisions for planned content at all domain levels. Faculty

were effective in their selection of test types, but not efficient. They tended to overtest planned content. This finding was reinforced by the numeric indices obtained for unity and effectiveness.

Phase II verified Phase I responses. When given an explanation of the expert system test type recommendations, approximately 64 percent of the respondents agreed with the expert system. Phase II data reinforce the tendency to overtest content. When reasons were provided, they appeared as legitimate reasons for wanting to be effective. The second highest reason for disagreeing was to guarantee employers that students have marketable skills when they graduate. When disagreeing, most of the respondents expressed a personal like for the expert system test-type choices. Their alternative test-type choices were most likely effective measures of testing rather than efficient ones.

Phase I data support respondents' decision not to select "want to purposely undertest content," as respondents tended not to undertest content. In fact, only 19 percent of the content goals were undertested in Phase I; 55 percent were overtested. On the other hand, it is unusual that the response, "want to purposely overtest content" was not selected by a larger proportion of Phase II respondents since 55 percent of content goals were overtested. These data suggest that faculty do not realize that they are overtesting

content as 73.3 percent overtested content in Phase I. Further, data suggest that faculty do not understand curriculum alignment principles.

CHAPTER V

CONCLUSION AND RECOMMENDATIONS

Chapter five includes an overview of the study. The overview describes the purpose of the study, procedures used to gather data, and the statistical design used to analyze the data. The conclusions address the research questions based upon the data analyses. Recommendations are proffered based on these conclusions and areas for possible further research. This chapter is presented in five sections: study overview, summary of findings, conclusions, recommendations, and researcher's commentary.

Study Overview

The purpose of this descriptive study was to document the effects of applying the platform unity concept in criterion referenced situations. Specifically, faculty planning and testing practices were examined as a way to fulfill the purpose.

This study was predicated on English's (1986-1987, 1988, 1989, 1992a, 1992b) curriculum alignment concept, where the written curriculum ought to be the delivered and the evaluated curriculum. Curriculum alignment principles were analyzed through Vogler's (1991) C-P-A model and Performance Instruction containing the expert system. Based on Gronlund's (1981) table of specifications, a new table was developed to equate domain-level content plans with associated test types.

Various principles on planning and testing were researched before developing formulae that provided quantitative indices of platform unity between content planning and testing, and effectiveness and efficiency of test type selection. Numerically, the platform unity index was greater than zero when content was overtested, less than zero when the content was undertested. Absolute unity or perfect alignment occurred when the unity index was zero. Faculty were effective and efficient in test item type selection when unity occurred.

Misalignment was also examined. Numerical values exceeding the normal calculated index range were assigned. The most severe cases of misalignment occurred when planned content was not tested and when tested content was not planned. In the former condition, an index of -3 was assigned, indicating undertesting. In the latter condition, an index of +3 was assigned, indicating overtesting.

The principle research question in this study was the following: Is there a difference between expert system and faculty test type selection? The null hypothesis was that there is no difference between expert system and faculty test-item type choices. In addition, the following research questions were addressed:

1. How can a unity index be used?
2. What test types do faculty use to test facts,

understanding, and application?

Chapter Two consisted of a literature review. From the review, a series of questions were addressed: (a) How is alignment related to unity? (b) What is overtesting and undertesting? (c) What is effective and efficient testing? (d) What is the rationale to support domain-level planning decisions as predictors of test item types? and (e) What is the connection among taxonomy classifications?

Two phases of a mailed survey were used to collect data and test the null hypothesis. Calculations yielded numeric indices that were compared to values representing theoretical constructs. A sample (n=30) of full-time faculty teaching Introduction to Computers in Virginia community colleges responded to the survey. Each instrument contained six content goals. For Phase I, a response rate of 100 percent was obtained; for Phase II, a 93.3 percent response rate (n=28) was obtained.

Responses from Phase I were used to calculate indices of platform unity and effectiveness and efficiency of test-type selection. Indices were obtained for 123 of 180 content goals. Based on the platform unity index, faculty overtesting and undertesting course content were determined. The Chi-square Goodness of Fit test was used to analyze differences between the expert system recommendations and faculty test-type choices of planned, cognitive, domain referenced content.

In Phase II respondents verified the collective responses from Phase I. Respondents were provided the opportunity to agree or disagree with expert system domain-referenced test-item type recommendations across 168 content goals. Respondents selected reasons for disagreeing with the expert system from a prepared list of alternatives.

Summary of Findings

Abbreviated findings from Phase I and Phase II surveys are reported in this section. A complete summary of survey findings was presented at the conclusion of Chapter Four.

Demographics

Respondents were computer faculty who had, on average, 14 years' teaching experience. They taught four sections of Introduction to Computers in the past academic year. Seventy percent of the faculty taught this course for more than six years.

Phase I Findings

Faculty planning and testing decisions were analyzed in the first survey phase using formulae to compute numerical indices for platform unity, effectiveness and efficiency of test-type selection. The overall mean platform unity index was +0.64; the effectiveness index was +0.33; the efficiency index was -1.44. These values indicated faculty were effective but not efficient in their test-type selection. The tendency to overtest content is apparent, particularly at the

fact and understanding levels.

Summary of Calculated Indices

Indices of five constructs calculated in this study readily point to four conclusions about respondents.

1. Faculty did not plan and test congruently.
2. Faculty tended to overtest content.
3. Faculty were effective in selecting test types.
4. Faculty were inefficient in selecting test types.

Differences Between Expert System and Faculty Test Types

The research hypothesis in this study was that faculty do not plan and test congruently. Using the expert system to represent congruence, or unity, the Chi-square Goodness of Fit Test compared faculty responses with the expert system. The test results supported the research hypothesis and yielded significant differences ($p < .01$) between the expert system test type recommendations and faculty selection for all six content goals presented on the survey. The differences indicated that faculty test types were not congruent with the domain level used to plan content.

Phase II Findings

Phase II reported a summary of selected test types from Phase I. Faculty indicated agreement or disagreement with the expert system test type recommendations for each content goal. Respondents agreed with expert system recommendations on 63.7 percent of the goals. Faculty agreed with expert system

recommendations at the higher domain levels, i. e., multiple-choice and essay tests.

Respondents disagreed with the expert system on 36.3 percent of the content goals. A personal dislike for test types was the dominant reason for disagreement. The second highest reason was that faculty wanted to guarantee employers that students had marketable skills when they graduate.

Purposes of Testing

The purpose of testing may determine differences with platform unity and expert system test-type recommendations. Phase II attempted to identify testing purposes, with two out of six reasons on the instrument related to purposes of testing.

One purpose of testing that legitimized overtesting was to guarantee employers that students have marketable skills when they graduate. In this case, faculty need to be effective in testing. On the other hand, a purpose of testing that legitimized undertesting was the desire to keep students in the program. In this case, faculty need to be efficient in testing.

Of the faculty disagreeing with the expert system, only 12 percent of Phase II content goals were selected for the purpose of guaranteeing employers with students having marketable skills. As shown in Table 28, there were no content goals selected to keep students in the program.

Differences Between Phase I and Phase II Data

Phase II data were used to verify Phase I respondent data. An interesting phenomena occurred when Phase I platform unity rates per content goal were compared with Phase II agreement rates per content goal. This comparison assumes that agreeing with expert system recommendations in Phase II was an affirmation of unity.

The expert system represents absolute unity. To achieve unity, planned content and tested content occur at the same domain levels, according to a Table of Test Type Specifications. A Table of Specifications showing parity between planned and tested content was presented in Chapter Three.

A nonparametric, Two-Sample Proportion Test was used to test for differences between Phase I and Phase II responses. This test specifically compares two population proportions to see if they are different. The hypothesis was that there is no difference between Phase I and Phase II rates of unity. The number of content goals that were united in Phase I was 32 out of 123. In Phase II the number of respondents agreeing with the expert system test-type recommendations was 107 out of 168. A Z-value of -6.36 ($p < .01$) was obtained. Consequently, the hypothesis was rejected, thus indicating a difference between rates of unity from Phase I to Phase II.

These differences suggest that when faculty were informed

of the platform unity/alignment concept, it made sense. The reason why the rates of unity differ can only be attributed to the explanation of the platform unity concept provided as instructions to the Phase II survey. A copy of this instruction sheet appears in Appendix B. Respondents learned about planning and testing unity, thought it made sense, and consequently agreed with the expert system.

Seeing how dramatically faculty accepted curriculum alignment principles indicates their fundamental need for curriculum information that directly affects what they do in the classroom. Faculty responses overwhelmingly indicated that they supported curriculum alignment. Realizing that faculty are unaware of curriculum alignment principles was a conclusion of this study. Getting the information to the faculty in unobtrusive ways represents a recommendation and a challenge for academic administrators.

Conclusions

The following conclusions are based on the findings of this study. The conclusions include a discussion of the implications of the study. The order of the conclusions follows the order of topics presented in the Review of the Literature. Each conclusion, linked to the literature, appears in bold type in the narrative.

1. Savard and Cotton (1982) stated that implementing alignment as an educational practice was not easy nor

widespread. The findings in this study support this view.

In the literature, information on planning and testing was widely developed and sophisticated. While it is widely assumed that faculty know how to plan and evaluate course content, evidence presented in this study suggests they do not. Unfortunately, literature on the activities of planning, delivery and evaluation as a unit is scant. They are treated separately in the literature and in practice. Techniques for evaluating tests is notably deficient for criterion-referenced tests. **Therefore, it was concluded that relevant information dealing with the alignment of planned and tested content is not widely developed. It does not receive adequate attention in practice. Information must become part of curriculum courses and included in the professional development agenda for faculty.**

2. English and Steffy (1983) stated that a problem in curriculum design occurs when what is planned is not tested. Pautler (1990) advocated instructional design principles to develop courses of study in community colleges. He believed if community college faculty were given more training in instructional design principles, improved institutional effectiveness would result. The findings in this study support design deficiencies on the part of faculty. **Therefore it was concluded that faculty need structured, instructional design principles to assist them in developing aligned content**

plans and tests.

3. Pautler (1990) believed the written documentation of curriculum design may be one of the least understood activities within community colleges. Findings in the study indicated that six faculty tested content without having planned it. Faculty violated English's (1992b) frontloading alignment principle. **Therefore, it was concluded that faculty in this study did not understand the importance of written curriculum documentation in the form of courses of study.**

4. An understanding of platform unity is essential. Nitko (1989) stated that tests created by classroom teachers often inadequately measure the teachers' espoused instructional goals. He cautioned that certain, undesirable consequences might result when tests are inadequately linked to instructional plans. These consequences include students' motivation for learning could be reduced, decisions about students might be made unfairly, the effectiveness of instruction may be evaluated incorrectly, and teachers and students might be inappropriately informed about learning progress and learning difficulties.

Although this study did not examine actual course planning documents, there is evidence to suggest that teacher-made tests did not match planned, curriculum goals (represented in this study as stated content goals). Phase I findings indicated a high rate of disunity because faculty

overtested content. While their testing decisions were effective, they were inefficient. Faculty did not relate the purpose of testing to support their decisions to overtest content. In some cases, disunity was absolute, meaning tested content was not planned and planned content was not tested. **Therefore it was concluded that teacher-made tests inadequately measure stated instructional goals.**

5. Baker and Popham (1970) stated that objectives and evaluation should be identical; test items should be drawn from the class of behaviors specified in the objectives. According to Vogler (1991), the verb used to plan instruction should be the same verb used when delivering and testing instruction. Using the same verb to plan and test assures consistency. While this principle undergirds curriculum alignment, faculty in the study did not consider the planning verb provided in a content goal statement when they tested content.

Faculty in the study did not plan and test content consistently. Their test items came from a different domain than the outcomes specified in the content plans. A platform unity index provided evidence of this incongruence. Overall, this index was +0.64. The data suggest that faculty did not know that the levels of their content plans and tests differ. **Therefore, it was concluded that respondents in the study show no distinct knowledge of curriculum alignment principles.**

6. Popham (1973) believed a great deal of sophisticated work needs to be done to devise measures that satisfactorily serve to sample the behavior and content domains delimited by a well-stated objective. The formulae developed in the study to ascertain platform unity and effectiveness and efficiency of test type selection represent examples of measures to assess criterion-referenced instruction. A sophisticated tool that combines planning, delivery and evaluation is Performance Instruction, which serves as the model on which planning and testing decisions were based. Indices based on the system's expert system provided sophisticated ways to determine platform unity, effectiveness and efficiency. **Therefore it was concluded that unity, effectiveness, and efficiency indices developed through this study represent sound, sophisticated measures to assess criterion-referenced examinations.**

7. Ebel (1965) believed classroom tests prepared by the faculty member are likely to fit the content and objectives of a particular course better than a test prepared by anyone else. Although outside standardized testing documents were not analyzed, the findings in this study challenge Ebel's belief. Teacher-selected test items were not found to be better than test items prepared by outsiders. Consequently, alignment, as determined through formulae, was weak. This study did not show a strong match between planned and tested

content through a platform unity analyses. Therefore it was concluded that in this study, teacher-made, criterion referenced tests did not fit the content better than a standardized test prepared by an external source.

8. Ball, Doss and Dewalt (1986) and Stanley and Hopkins (1972) suggested that teachers are weak in test construction. Testing often occurs at the lowest level, and teacher-constructed tests lack relevant difficult items. Consequently, students are not encouraged to use their minds. Since this study did not examine testing documents, the findings were inconclusive on test construction. The findings of this study do refute the conclusion that testing occurred at the lowest level.

In this study, indices of platform unity tended to be positive, emphasizing content overtesting. Faculty preferred to test all levels of content through multiple-choice items. This was an effective means of testing, albeit not efficient. It was possible that multiple-choice questions were just four true-false questions. This was recognized as a limitation since the actual testing documents were not examined. Therefore, it was concluded that faculty are effective, but not efficient in their test type selections. The dominance of multiple-choice tests was the principle cause for high levels of overtesting.

9. Fleming and Chambers (1983) found that teachers

(a) used short-answer questions most often; (b) avoided essay items; (c) used more matching items than multiple-choice; (d) devised more test questions to sample knowledge of facts; and (e) developed few questions to test application behaviors.

The findings of this study appear to refute the conclusions of Fleming and Chambers. Faculty tested content with 216 total test item types, and their use of the six test item types varied. However, the majority of test items were multiple-choice. Based on faculty responses, short-answer items were the second choice. Essay items were selected for testing content on 21 out of 216 test questions. Faculty did not use more matching items. Multiple-choice items were used to test content 85 times. Matching test items were used to test content 32 times.

Fleming and Chambers observed that more test questions were written to sample knowledge of fact. The opposite was found in this study. While there were equal proportions of content from three cognitive domains, there were more test items devised to test application. Across the three domain levels, 19 percent were used to test facts, 32 percent were used to test understanding, and 49 percent were used to test application.

Based on these findings, previous findings by Fleming and Chambers were not replicated in this study. **Therefore, it was concluded that although faculty in this study tested content**

with all six test types, they preferred to use higher level test types. Of these test types, faculty preferred multiple choice tests.

10. Jacobs and Chase (1992) reported that college instructors feel poorly prepared to construct tests in their classes because they have never received any kind of formal training in this area. Judgement cannot be made on the quality of test questions as test documents were not studied. Based on indices of platform unity, it could be inferred that faculty in this study had no formal training on test construction.

Specifically, this study asked faculty to report if they had a course in tests and measurement in the last three years. Twenty seven percent of respondents reported taking a course in tests and measurement. Although this study did not measure faculty opinions, data refute the conclusion that faculty never received formal training in test construction. Evidently, knowledge learned in formal courses was not applied to testing decisions in this study. Even after having taken a recent course in curriculum theory or in tests and measurement, faculty had no better indices of unity than the rest who did not take curriculum coursework. **Therefore it was concluded that more formal training needs to occur to instruct faculty on principles of test construction. Practical examples should be provided so theoretical constructs can be**

applied to practical situations.

11. According to Anderson (1987), faculty need to gain a perspective on their purposes of testing. Findings in this study support this conclusion. Data from Phase II found faculty not able to state a reason for disagreeing with the expert system. Testing purposes are either not known or not explicit. In Phase I, 55 percent of the content goals were overtested; in Phase II, reasons to support overtesting decisions were provided by 13 percent of faculty. **Therefore it was concluded that the gap between the rate of content goals overtested and the rate of reasons selected to legitimize overtesting was attributed to a lack of the purposes of testing.**

12. Jacobs and Chase (1992) deemed it essential that test item content relates back to a plan. The principle of curriculum alignment used in this study support this conclusion from the literature. In actual testing decisions, this study did not find test items relating to a plan. In Phase I, 20 percent of the respondents tested a content goal for which there were no plans, resulting in overtesting. This severe case of misalignment required the researcher to assign a unity value. **Therefore, it was concluded that testing decisions did not relate to a plan.**

13. Gronlund (1985) advocated using of a table of specifications to identify evaluation instruments that provide

the most direct evidence of stated learning outcomes. The findings in this study pointed to a need for using a table of specifications. Twenty-six percent of content goals were planned and tested in unity, suggesting that faculty testing practices do not match their content plans 74 percent of the time. Therefore it was concluded that a systematic tool to examine planning and testing could benefit faculty. The table of specification represents the kind of tool that would assist faculty.

14. Test items are usually divided into two general categories, objective and subjective. Powell and Gillespie (1990) used the term, "constructed response item," to refer to subjective tests. They indicated that, despite inherent complexities in scoring, the use of constructed-response tests is increasing.

Findings in this study refute this statement. In the survey, faculty preferred to test content with objective tests rather than subjective tests. An unpaired t-test was used to assess statistical differences between the test types, assuming unequal variance between the two groups. The null hypothesis was that there was no difference between objective and subjective test items used to test content. A significant t-value of 3.32 ($p < .01$) was obtained and the null hypothesis rejected.

Figures 5, 6, and 7 in Chapter Four show the dominance of

objective tests for testing fact, understanding, and application content. **Therefore, it was concluded that faculty tested content by using objective means rather than subjective.**

15. Kellaghan, Madaus, and Airasian (1982) believed it was not necessary for faculty to know more about interpreting test scores. Rather, they argued it was essential to provide guidance in practical, quality control measures so faculty could evaluate their classroom tests.

The formulae developed in this study were especially designed to work with teacher-constructed, criterion referenced evaluation instruments and produced indices whereby planning and testing decisions could be analyzed. The calculated data yielded numeric indices that, at even a quick glance, determined faculty planning and testing was not united. The indices also indicated that faculty were effective in testing, but not efficient. **Therefore, it was concluded that the formulae provided useful indications of planning and testing practices that have practical, quality control measures for faculty to use.**

16. Ebel (1979) observed that direct institutional support of teaching in higher education is weak. As part of an overall plan for higher education to become more effective, Diamond (1989) recommended matching objectives to assessment instruments at all instructional levels.

The issue of curriculum alignment and platform unity is directly related to institutional effectiveness. However, in compiling a taxonomy of the literature on effectiveness from 1970 to 1990, Welker and Morgan (1991) found references to community college curriculum issues in only five percent of the literature. Only two of these were research items. **Therefore, it was concluded that curriculum development issues and curriculum alignment need to become a topic with which community college faculty become familiar.**

17. At the outset of Phase II, survey participants were given an explanation of Performance Instruction and expert system test-type recommendations. Based on their Phase II responses, faculty agreed with expert system recommendations in 63.7 percent of the content goal statements, suggesting that once given an explanation of curriculum alignment principles, they agree with them.

To substantiate this notion, a two-sample proportion test yielded a statistically significant difference between two proportions of Phase I unity indices and Phase II unity, measured by respondent agreement with expert system recommendations. The significant difference indicates that faculty recognize that the expert system test-type recommendations equate to platform unity. Stated differently, Phase II responses indicate if faculty were to test planned content after exposure to the expert system, they would select

test-item types that align with domain-referenced content plans. **Therefore, it was concluded that faculty will incorporate curriculum alignment theories in their practice if they are informed and provided examples of how to do it.**

18. The conditions of disunity between planning and testing revealed in this study suggest a need for broad-based discourse on the subjects of unity, effectiveness, and efficiency. Cochran (1989) viewed such curriculum development activities as an important part of the instructional effectiveness continuum. **Therefore it was concluded that information on curriculum alignment needs to reach the faculty.** This could be accomplished through faculty development. Providing faculty with relevant information on effective instructional principles is a strategy that could contribute to improvements in institutional effectiveness.

Recommendations

This section offers recommendations based on the findings and conclusions of this study. It also offers recommendations for further research. Each recommendation appears in bold type in the narrative.

1. This study examined alignment between planned and tested content. Findings revealed faculty shortcomings with fundamental curriculum issues. More broadly, it was concluded that faculty are unfamiliar with the concept of curriculum alignment. However, evidence indicated that faculty were

receptive to the concept, and many of the conclusions point to professional development as a way to inform current teaching faculty.

The platform unity concept should be used as an information source to promote the principles of curriculum alignment. Specifically, the indices produced through formulae should be introduced to faculty to use as a self-analysis tool in criterion-referenced testing situations to ensure that their content plans and tests are united, effective, and efficient. Therefore, it is recommended that the concept of curriculum alignment and the platform unity analyses become an explicit part of the faculty development process. It is further recommended that faculty development include essential information about instructional design principles. Further study might examine platform unity, effectiveness, and efficiency indices and make refinements to the formal constructs to ensure ease of application and interpretation. In addition, further study might reconceptualize the numeric intervals so that they become open-ended.

2. The topic of "alignment" is virtually absent in the literature, the language and practice of higher education. Courses in tests and measurement focus on standardized, norm-referenced testing situations. Therefore, it is recommended that more emphasis on the topic is needed in the professional

literature.

3. In this study, we do not really know how students performed on the job following formal schooling. Criterion-referenced testing supposedly indicates a level of preparedness for individual students. Though this study was not a commentary on teacher testing practices, teacher-made tests were shown to inadequately measure stated instructional goals. **Therefore, it is recommended that faculty be provided opportunities to specifically explore criterion-referenced testing concepts through a tests and measurement course.** As a recommendation for further study, teacher testing practices could be examined along with student success as a concomitant variable. In such a study, various faculty platform unity mean indices could be tested to see if absolute unity serves as the best predictor of student success.

4. It is not known if faculty testing practices are practical or applicable to real world situations. This study offered no explanation on actual test questions, and individual test questions were not part of this study. However, the findings suggested that teacher-made tests did not fit the content any better than standardized tests.

In this study, only 26 percent of the content goals were planned and tested in unity. It is possible that faculty were testing in the way they were taught. They may be selecting test items from a standardized, publisher-prepared test item

bank unrelated to content plans. If this is true, unity calculations would need to consider examining actual planning and testing documents. Further studies may want to investigate these documents to determine unity or match between the planned and tested curriculum. Examining the actual planning and testing documents serves as a validation process that would substantiate data used in numeric calculations. **Therefore, it is recommended that faculty become familiar with constructing a Table of Test Type Specifications.** This table shows content domain used to plan content on one dimension and appropriate test types on the other. The test types match the domain level of the planned content.

5. The two-sample proportion test yielded significant differences between Phase I and Phase II responses. This difference suggests that changes in testing practices may occur by providing faculty with theoretical information regarding the alignment of planned and tested content. The changes are more in line with the expert system recommendations. While this finding may be serendipitous, it represents a topic for further study. Similar information could be provided to larger groups of faculty to see if they agree with expert system recommendations. **Therefore, it is recommended that further study investigate platform unity indices among faculty familiar with Performance Instruction to**

see if content planning and testing decisions come closer to absolute unity.

6. It was concluded that faculty knew little about the purposes of testing, and their testing decisions did not relate to a plan. When the purposes of testing are made explicit, variations in the calculated indices would be permitted. For example, to guarantee employers receive graduates with marketable skills represents a situation where overtesting would be appropriate. Faculty in this study demonstrated a propensity to overtest content (though faculty did not know why they overtested to the extent they did). They preferred to use objective rather than subjective test types.

This study did not have faculty state their testing purpose when they provided testing data in Phase I. Knowing the purpose of the test is an integral part of the overall training in educational measurement. **Therefore, it is recommended that faculty professional development include information on the purposes of testing.** Further studies into testing purposes could examine appropriate variations in the formulae indices.

7. Faculty in this study violated the principles contained in Performance Instruction (Vogler, 1991). They did not understand the need for documented curriculum information, and they showed no distinct knowledge of curriculum alignment

principles. Their testing decisions did not relate to a plan. However, when informed, faculty responded positively to curriculum alignment information.

The formulae were found to be useful indicators of planning and testing practices. Due to their utility, it was concluded that practical, formal training needs to occur. A mechanism to bring about change, inform faculty about principles of alignment, and encourage effective practices in their teaching duties is available through the software suite containing the expert system. The expert system in the software tutors the user in correct design, lesson planning, and creating test items and criterion-referenced examinations. Therefore, it is recommended that an awareness of curriculum alignment become a part of the professional development activities in which faculty and administrators participate. Further, it is recommended that Principles of Performance Instruction become the vehicle for reaching the faculty with sound principles of instructional design.

The formulae created for this study were based on Vogler's model. Software users make planning and testing decisions. As a tool for users, the formulae could be heuristically reproduced and provided to evaluate decisions for platform unity, effectiveness and efficiency. Certain warnings or prompts could be added to alert users when an overtesting or undertesting situation is created. Indices

could be calculated per content goal or aggregated for an entire course or program.

Summary of Recommendations

Many of the conclusions and recommendations called for the need to formalize curriculum alignment information and make it available to faculty through professional development. However, it may be necessary to determine who sets the agenda for faculty development. Do administrators supervising the instructional programs know about curriculum alignment? Beyond that, do administrators know faculty plans and tests are not united? The fact that faculty do not know about curriculum alignment may stem from administrators' unfamiliarity with the concept. **Therefore it is recommended that further study investigate administrator awareness of curriculum alignment principles.**

Researcher's Commentary

Researcher observations beyond the findings of this study conclude Chapter Five.

Community colleges are institutions regarded for their emphasis on teaching. Yet, faculty awareness of a fundamental concept such as curriculum alignment appears woefully inadequate. Institutions claiming to support teaching have a job to do. The task of providing fundamental information on basic teaching concepts rests with administrators and their commitment to faculty development. Faculty employed in

community college occupational programs tend to have experience in their field of specialty rather than in instructional methodology. Computer faculty, the population surveyed in this study, represented a group who typically learned computing skills in business and industrial settings instead at academic settings. Regardless of where faculty develop their expertise, community college administrators need to realize faculty deficiencies exist and provide professional development experiences for them to learn fundamental teaching and learning concepts, such as curriculum alignment.

Cochran (1989) stated that postsecondary faculty want to improve their professional competence. How institutions respond may not be effective. It is no secret that faculty relish autonomy. They prefer to design courses, plan instructional activities, experiment, and innovate on their own. Most approaches to faculty development limit autonomy by putting faculty into relatively passive roles (Menges, 1991). According to Menges (1991),

"when teaching improvement is the goal, resources should be provided that facilitate natural feedback. Information imposed by others tends to increase rather than reduce uncertainty. Information over which the teacher perceives no control tends to raise anxiety rather than contain it" (p. 34).

Savard and Cotton (1982) said that while the idea of curriculum alignment is simple, its implementation as an

educational practice is not easy; yet, the potential it holds for improving the quality and effectiveness of education demands that the topic be addressed.

Curriculum alignment is a practice that can make the educational process effective. Much more funding needs to be earmarked toward faculty improvement of the planning, delivery, and evaluation of instruction. Pautler (1992) recommended instructional design principles for community college faculty. The computerized model of curriculum development advocated by Vogler (1991) makes sense. It follows sound principles of instructional design and upholds the principles of curriculum alignment. Faculty need to be introduced to such tools.

The tools recommended in this study represent a self-analysis method whereby faculty can perform individual analyses on content planning and testing. The formulae represent a simple way to analyze a simple principle. Making them available heuristically might be the key to acceptance. If the analytical methods could be incorporated into a curriculum development system, faculty would have immediate and practical feedback via indices on platform unity, overtesting and undertesting of course content, and effectiveness and efficiency of test-type selection. With such information, the quality and effectiveness of education is bound to improve.

REFERENCES

- Adams, G. S. (1966). Measurement and evaluation in education, psychology, and guidance. New York: Holt, Reinhart and Winston.
- Airasian, P. W., & Madaus, G. F. (1983). Linking testing and instruction: Policy issues. Journal of Educational Measurement, 20, 103-118.
- American Association of School Administrators. (1989). Testing: Where we stand. Arlington, VA: Author. (ERIC Document Reproduction Service No. ED 314 854).
- Anderson, S. B. (1987). The role of teacher-made tests in higher education. In D. Bray and M. Belcher (Eds.), Issues in Student Assessment. New Directions for Community Colleges, No. 59 (pp. 39-44). San Francisco: Jossey-Bass.
- Ammons, M. (1969). Objectives and outcomes. In R. L. Ebel (Ed.), Encyclopedia of educational research (4th ed.) (pp. 908-914). New York: Macmillan.
- Ary, D., Jacobs, L. C., & Razavieh, A. (1985). Introduction to research in education. Philadelphia: Holt, Reinhart and Winston.
- Babbie, E. (1990). Survey research methods. (2nd ed.). Belmont, CA: Wadsworth.
- Babbin, P. (1987). Instructional objectives. (ERIC Document Reproduction Service, No. ED 299 236).
- Baker, E. L., & Popham, W. J. (1970). Systematic instruction. Englewood Cliffs, NJ: Prentice-Hall.
- Baker, E. L. & Herman, J. L. (1983). Task structure design: Beyond linkage. Journal of Educational Measurement, 20, 149-163.
- Ball, D. W., Doss, A. R., & Dewalt, M. W. (1986). Level of teacher objectives and their classroom tests: Match or mismatch. Journal of Social Studies Research, 2(10), 27-30.
- Bergquist, W. H., & Armstrong, J. L. (1986). Planning effectively for educational quality. San Francisco: Jossey-Bass.

- Bloom, B. S. (1961). Quality control in education. Tomorrow's teaching. Oklahoma City: Frontiers of Science Foundation.
- Bloom, B. S., Englehart, M. D., Furst, G. J., Hill, W. H., & Krathwohl, D. L. (1956). Taxonomy of educational objectives: Handbook I, Cognitive domain. New York: David McKay Company.
- Bloom, B. S., Madaus, G. F., & Hastings, J. T. (1981). Evaluation to improve learning. New York: McGraw-Hill.
- Boersma, W. C. (1967). The effectiveness of the evaluative criteria as a stimulus for school improvement in eleven Michigan high schools. Unpublished doctoral dissertation, University of Michigan, Ann Arbor, MI.
- Brickell, H. M. (1976). Needed: Instruments as good as our eyes. Evaluation Center Occasional Paper Series No. 7. Kalamazoo, MI: Western Michigan University.
- Cochran, L. H. (1989). Administrative commitment to teaching. Cape Girardeau, MO: Step Up, Inc.
- Coffman, W. E. (1983, July). Testing in the schools: A historical perspective. Paper presented at the Center for the Study of Evaluation Annual Invitational Conference, University of California at Los Angeles, Los Angeles, CA. (ERIC Document Reproduction Service No. ED 338 670)
- Cook, J. M. (1978). Developing learning outcomes: Module 2 of a handbook on clarifying college learning outcomes. Columbia, MD: Council for the Advancement of Experiential Learning.
- Cook, W. W. (1951). The functions of measurement in the facilitation of learning. In E. F. Lindquist (Ed.), Educational measurement. Washington, DC: American Council on Education.
- Copperud, C. (1979). The test design handbook. Englewood Cliffs, NJ: Educational Technology Publications.
- Cross, K. P. (1987). Teaching for learning. AAHE Bulletin, 39(8), 3-7.

- Crowell, R., & Tissot, P. (1986). Curriculum alignment. Washington, DC: U. S. Department of Education. (ERIC Document Reproduction Service No. ED 280 874)
- Cunningham, G. (1986). Educational and psychological measurement. New York: Macmillan.
- Cureton, E. E. (1969). Measurement theory. In R. L. Ebel (Ed.), Encyclopedia of educational research (4th ed.), (pp. 785-804). New York: Macmillan.
- DeBrunner, L. S., DeBrunner, V. E., & Vogler, D. E. (1991). Performance-based instruction in electrical engineering. In ASEE 1991 Annual Conference Proceedings.
- Deutsch, G. (1991). Principles and purposes of outcomes assessment. In M. E. Knight, D. L. Lumsden & D. Gallaro (Eds.), Outcomes assessment at Kean College of New Jersey (pp. 5-14). Lanham, MD: University Press of America.
- Diamond, R. M. (1989). Designing and improving courses and curricula in higher education: A systematic approach. San Francisco: Jossey-Bass.
- Dillman, D. A. (1978). Mail and telephone surveys: The total design method. New York: John Wiley & Sons.
- Dyer, H. S. (1970). Basic elements of the measurement process. In Norman E. Gronlund (Ed.), Readings in measurement and evaluation (pp. 3-10). New York: Macmillan.
- Ebel, R. L. (1965). Measuring educational achievement. Englewood Cliffs, NJ: Prentice-Hall.
- Ebel, R. L. (1979). Essentials of educational measurement. (3rd ed). Englewood Cliffs, NJ: Prentice-Hall.
- Ebel, R. L., & Frisbie, D. (1986). Essentials of educational measurement (4th ed.). Englewood Cliffs, NJ: Prentice-Hall.
- English, F. W. (1979). Management practice as key to curriculum leadership. Educational Leadership, 36, 408-413.
- English, F. W. (1986-1987). It's time to abolish conventional curriculum guides. Educational Leadership, 44(4), 50-53.

- English, F. W. (1988). Curriculum auditing. Lancaster, PA: Technomic Publishing Co.
- English, F. W. (Ed.). (1992a, January). Curriculum audit training handbook. (Available from AASA, 1801 N. Moore St., Arlington, VA 22209).
- English, F. W. (1992b). Deciding what to teach and test: Developing, aligning, and auditing the curriculum. Newberry Park, CA: Corwin Press.
- English, F. W., & Steffy, B. E. (1983). Differentiating between design and delivery problems in achieving quality control in school curriculum management. Educational Technology, 24(2) 29-32.
- Fleming, M., & Chambers, B. (1983). Teacher-made tests: Windows on the classroom. In W. E. Hathaway (Ed.), Testing in the schools. New Directions for Testing and Measurement, No. 19 (pp. 29-38). San Francisco: Jossey Bass.
- Fortune, J. C., and Associates (1985). Understanding testing in occupational licensing. San Francisco: Jossey-Bass.
- Feldhausen, J. F. (1980). The three-stage model of course design. Englewood Cliffs, NJ: Educational Technology Publications.
- Gaff, J. G. (1975). Toward faculty renewal. San Francisco: Jossey-Bass.
- Gagne, R. M. (1975). Essentials of learning for instruction. New York: Holt, Reinhart and Winston.
- Gagne, R. M., & Briggs, L. J. (1979). Principles of instructional design (2nd ed.). New York: Holt, Reinhart and Winston.
- Gentry, D. L. (1989). Teacher-made test construction. (ERIC Document Reproduction Service No. ED 313 444).
- Gerhard, M. (1971). Effective teaching strategies with the behaviorial outcomes approach. West Nyack, NY: Parker Publishing Company.
- Goslin, D. A. (1967). Teachers and testing. New York: Russell Sage Foundation.

- Green, J. A. (1975). Teacher-made tests (2nd ed.). New York: Harper & Row.
- Greeno, J. G. (1976). Cognitive objectives of instruction: Theory of knowledge for solving problems and answering questions. In David Klahr (Ed.), Cognition and instruction (pp. 3-11). Hillsdale, NJ: Lawrence Erlbaum Associates.
- Gronlund, N. E. (1981). Measurement and evaluation in teaching. (4th ed.). New York: Macmillan.
- Gronlund, N. E. (1982). Constructing achievement tests. (3rd ed.). Englewood Cliffs, NJ: Prentice-Hall.
- Gronlund, N. E. (1985). Stating objectives for classroom instruction (3rd ed.). New York: Macmillan.
- Gunter, M. A., Estes, T. H., & Schwab, J. H. (1990). Instruction: A models approach. Boston: Allyn and Bacon.
- Hartzell, M. S. (1984). Checking for curriculum/test overlap: Two methods discussed. Portland, OR: Northwest Regional Educational Laboratory. (ERIC Document Reproduction Service No. ED 246 081).
- Heywood, J. (1989). Assessment in higher education (2nd ed.). New York: John Wiley and Sons.
- Hinkle, D. E., Wiersma, W. & Jurs, S. G. (1979). Applied statistics for the behavioral sciences. Chicago: Rand McNally.
- Hoffman, B. (1962). The tyranny of testing. New York: Macmillan.
- Isaac, S., & Michael, W. B. (1990). Handbook in research and evaluation. (2nd ed.). San Diego: EdITS.
- Jacobs, L. C., & Chase, C. I. (1992). Developing and using tests effectively: A guide for faculty. San Francisco: Jossey-Bass.
- Kellaghan, T., Madaus, G. F., & Airasian, P. W. (1982). The effects of standardized testing. Boston: Kluwer-Nijhoff.

- Kibler, R. J., Cegala, D. J., Watson, K. W., Barker, L. L., & Miles, D. T. (1981). Objectives for instruction and evaluation (2nd ed.). Boston: Allyn and Bacon.
- Kogan M. (1986). Education accountability. London: Hutchinson.
- Krathwohl, D. L., Bloom, B. S., & Masia, B. B. (1964). Taxonomy of educational objectives: Handbook II, Affective domain. New York: David McKay Company.
- Kropp, R. P., Stoker, H. W., & Bashaw, W. L. (1966). The construction and validation of tests of the cognitive processes as described in the taxonomy of educational objectives. Cooperative Research Project No. 2117. Washington, DC: Office of Education, U. S. Department of Health, Education, and Welfare.
- La Fave, L. (1964). Essay versus multiple choice: Which test is preferable? Psychology in the Schools, 3, 65-69.
- Lazar-Morris, C., Polin, L., May, R., & Barry, L. (1980). A review of the literature on test use. Los Angeles: University of California, Center for the Study of Education. (ERIC Document Reproduction Service No. ED 204 411).
- Leinhardt, G. (1983). Overlap: Testing whether it's taught. In G. F. Madaus (Ed.), The courts, validity, and minimum competency testing. Boston: Kluwer-Nijhoff
- Leinhardt, G., & Seewald, A. M. (1981). Overlap: What's tested, what's taught? Journal of Educational Measurement, 18, 85-95.
- Linn, R. L. (1983). Testing and Instruction: Links and distinctions. Journal of Educational Measurement, 20, 179-189.
- Linn, R. L. (1989). Current perspectives and future directions. In Robert L. Linn (Ed.) Educational measurement. (3rd ed.) (pp. 1-10). New York: Macmillan.
- Lozak, J. (1987). Assessment and improvement in education. In D. Bray and M. Belcher (Eds.), Issues in Student Assessment. New Directions for Community Colleges, No. 59 (pp. 25-30). San Francisco: Jossey-Bass.

- MacCuish, D. A. (1986, February). A quality assurance model for higher education: A pilot study. paper presented at the Testing and Quality Assurance in Higher Education Conference. (ERIC Document Reproduction Service No. ED 273 168)
- Mager, R. F. (1962). Preparing instructional objectives. San Francisco: Fearon.
- Mager, R. F. (1973). Measuring instructional intent. Belmont, CA: Fearon-Pittman.
- Mehrens, W. A., & Lehmann, I. J. (1984). Measurement and evaluation in educational psychology. (3rd ed.). New York: Holt, Reinhart & Winston.
- Menges, R. J. (1991). The real world of teaching improvement: A faculty perspective. In M. Theall & J. Franklin (Eds.) Effective practices for improving teaching (pp. 21-37). New Directions for Teaching and Learning, No. 48. San Francisco: Jossey-Bass.
- Menges, R. J., & Mathis, C. B. (1988). Key resources on teaching, learning, curriculum, and faculty development. San Francisco: Jossey-Bass.
- Metfessel, N. S., Michael, W. B., & Kirsner, D. A. (1969). Instrumentation of Bloom's and Krathwohl's taxonomies for the writing of educational objectives. Psychology in the Schools, 6, pp. 227-231.
- Milton, O., and Associates (1978). On college teaching: A guide to contemporary practices. San Francisco: Jossey-Bass.
- Nitko, A. J. (1989). Designing tests that are integrated with instruction. In Robert L. Linn (Ed.), Educational measurement. (3rd ed.) (pp. 447-474). New York: Macmillan.
- Pautler, A. J. (1989). Curriculum alignment as a means for improving institution effectiveness. Community/Junior College Quarterly of Research and Practice, 13, 173-180.
- Pautler, A. J. (1990). Designing functional courses of study for community colleges. Journal of Studies in Technical Careers, 12, 173-181.

- Pautler, A. J. (1992). Curriculum leadership, innovation, and change. Community/Junior College Quarterly of Research and Practice, 16, 133-140.
- Payne, D. A. (1974). The assessment of learning. Lexington, MA: D. C. Heath.
- Payne, D. A. (1992). Measuring and evaluating educational outcomes. New York: Macmillan.
- Popham, W. J. (1973). The uses of instructional objectives: A personal perspective. Belmont, CA: Fearon.
- Popham, W. J. (1978). Criterion-referenced measurement. Englewood Cliffs, NJ: Prentice-Hall.
- Powell, J. L., & Gillespie, C. (1990, December). Assessment: all tests are not created equally. Paper presented at the Annual Meeting of the American Reading Forum, Sarasota, FL. (ERIC Document Reproduction Service No. ED 328 908)
- Roid, G., & Haladyna, T. (1982). A technology for test item writing. New York: Academic.
- Salvia, J., & Hughes, C. (1990). Curriculum-based assessment: Testing what is taught. New York: Macmillan.
- Savard, W. G., & Cotton, K. (1982). Curriculum alignment: Research on school effectiveness project. (ERIC Document Reproduction Service No. ED 265 631)
- Schiro, M. (1978). Curriculum for better schools: The great ideological debate. Englewood Cliffs, NJ: Educational Technology Publications.
- Scott, R. (1983). Curriculum alignment as a model for school improvement. (ERIC Document Reproduction Service No. ED 252 508).
- Simpson, E. J. (1966). The classification of educational objectives: Psychomotor domain. Urbana: University of Illinois Press.

- Stahl, R. J., & Murphy, G. T. (1981, April). The domain of cognition: An alternative to Bloom's cognitive domain within the framework of an information processing model. Paper presented at the annual meeting of the American Educational Research Association, Los Angeles. (ERIC Document Reproduction Service No. ED 314 854).
- Stanley, J. C., & Bolton, D. L. (1957). A review of Bloom's taxonomy of educational objectives. Educational and Psychological Measurement, 17, 631-634.
- Stanley, J. C., & Hopkins, K. D. (1972) Educational and psychological measurement and evaluation. Englewood Cliffs, NJ: Prentice-Hall.
- Stark, J. S., Lowther, M. A., Sossen, P. L. & Shaw, K. M. (1991). Course planning exploration for program self-study. Ann Arbor, MI: National Center for Research to Improve Postsecondary Teaching and Learning.
- Stiggins, R. J., & Bridgeford, N. J. (1985). The ecology of classroom assessment. Journal of Educational Measurement, 22, 271-286.
- Stiggins, R. J., Conklin, N. F., & Bridgeford, N. J. (1986). Classroom assessment: A key to effective instruction. Educational measurement: Issues and Practice, 5(2) 5-17.
- Stufflebeam, D. L., Foley, W. J., Gephart, W. J., Guba, E. G., Hammond, R. L., Merriman, H. O., & Provus, M. M. (1971). Educational evaluation and decision making. Bloomington, IN: Phi Delta Kappa.
- Swezey, R. (1981). Individual performance assessment: An approach to criterion-referenced test development. Reston, VA: Reston.
- Thorndike, R., & Hagen, E. (1969). Measurement and evaluation in psychology and education (3rd ed.). New York: John Wiley and Sons.
- Tuckman, B. W. (1985). Evaluating instructional programs. (2nd ed). Boston: Allyn and Bacon.
- Tuckman, B. W. (1975). Measuring educational outcomes: Fundamentals of testing. New York: Harcourt Brace Jovanovich.

- Tyler, R. W. (1951). The functions of measurement in improving instruction. In E. F. Lindquist (Ed.), Educational measurement. Washington, DC: American Council on Education.
- Vogler, D. E. (1991). Performance instruction: Planning, delivering, evaluating. Eden Prairie, MN: Instructional Performance Systems, Inc.
- Ward, J. G. (1982). An overview of the AFT's teaching and testing. In S. B. Anderson and L. B. Colburn (Eds.) Academic testing and the consumer (pp. 47-52). San Francisco: Jossey-Bass.
- Welker, W. F., & Morgan, S. D. (1991). One step beyond what the literature says on institutional effectiveness for community, junior and technical colleges. Community College Review, 19,(2), 25-31.
- Wesman, A. (1971). Writing the test item. In R. Thorndike (Ed.) Educational Measurement (2nd ed.) (pp. 81-130). Washington, DC: American Council on Education.
- Yeh, J. P. (1978). Test use in schools. Washington, DC: U. S. Department of Health, Education and Welfare and National Institute of Education.

APPENDIX A

**PHASE I SURVEY:
COVER LETTER, INSTRUCTIONS, AND INSTRUMENT**

750 Hethwood Blvd.
Apartment 300-L
Blacksburg, VA 24060
January 4, 1993

Mr.
Wytheville Community College
1000 East Main Street
Wytheville, VA 24382

Dear :

Thank you for agreeing to participate in this survey. There are two phases to my study. Phase I is enclosed. It examines the relationship between course plans and testing decisions. Phase II, which will be mailed to you after Phase I results are tabulated, will ask you to verify the decisions you made in Phase I.

Your responses will be grouped and tabulated with other respondents so that individual institutions and individual respondents can not be identified. Please be sure to provide your name and phone number in the spaces provided on the last page in case I need to discuss your comments with you.

My study is based on a curriculum system called Performance Instruction. For participating, I plan to mail you a book at the conclusion of Phase II. This book contains the principles which undergird the curriculum model used in this study.

A postage-paid return envelope is provided. You will assist me greatly by completing and returning the instrument by January 12, 1993.

Sincerely,

Thomas Leitzel

Enclosures

Instructor Planning and Testing Selection Inventory

BACKGROUND:

The nature of this survey is designed to explore the relationship between course plans, stated in the form of content goals, and the types of tests faculty select to evaluate student learning of content. A goal statement is the starting point in planning a course. Content goal statements, used below, are effective in communicating expectations to learners.

Six, actual course content goals were selected at random from an introductory computer course. These are listed below beside each item number. As an instructor of this course, you may or may not agree with the content. In fact, you may or may not plan this content in your course. You may or may not test this content in your course. That's fine.

The purpose of this survey is to obtain an idea of the types of decisions you as an instructor of this course make about planning and testing course content. Therefore, please base all your responses on the specific decisions related to content planning and testing you make for your course.

DIRECTIONS:

Next to each content goal statement, indicate (by choosing either "Yes" or "No") whether you explicitly list this content in your course syllabus. Then indicate whether you test this content in the course you teach. If you said "Yes" (meaning you test students on the content), then indicate (by marking an "X" in space provided) the primary type or types of test(s) you use to test the particular content. The test types listed on the survey are limited to true-false, matching, multiple choice, completion, short answer, and essay. Select as many testing types as you actually use in Introduction to Computer Science.

An example is provided on the next page.

A WORKED EXAMPLE:

xx. **Specify document printer speeds.**

Is this outcome content expressed in your CIS course syllabus?

No

Yes

Do you test this content in your course?

No (If "No", Go to next statement)

Yes (If yes, select the primary
type or types of test(s) you use)

True/False

Matching

Multiple Choice

Completion

Short Answer

Essay

Turn the page to begin Part I of the survey.

Instructor Planning and Testing Selection Inventory

Part I.

Page 1 of 5

1. Define input device terms.

Is this outcome content expressed in your CIS course syllabus?

No

Yes

Do you test this content in your course?

No (If "No", Go to next statement)

Yes (If yes, select the primary
type or types of test(s) you use)

True/False

Matching

Multiple Choice

Completion

Short Answer

Essay

2. Compare microcomputer software applications.

Is this outcome content expressed in your CIS course syllabus?

No

Yes

Do you test this content in your course?

No (If "No", Go to next statement)

Yes (If yes, select the primary
type or types of test(s) you use)

True/False

Matching

Multiple Choice

Completion

Short Answer

Essay

3. Analyze system design components.

Is this outcome content expressed in your CIS course syllabus?

___ No

___ Yes

Do you test this content in your course?

___ No (If "No", Go to next statement)

___ Yes (If yes, select the primary
type or types of test(s) you use)

- ___ True/False
- ___ Matching
- ___ Multiple Choice
- ___ Completion
- ___ Short Answer
- ___ Essay

4. Recall flowchart symbols.

Is this outcome content expressed in your CIS course syllabus?

___ No

___ Yes

Do you test this content in your course?

___ No (If "No", Go to next statement)

___ Yes (If yes, select the primary
type or types of test(s) you use)

- ___ True/False
- ___ Matching
- ___ Multiple Choice
- ___ Completion
- ___ Short Answer
- ___ Essay

5. Classify computer operations.

Is this outcome content expressed in your CIS course syllabus?

- No
- Yes

Do you test this content in your course?

- No (If "No", Go to next statement)
- Yes (If yes, select the primary type or types of test(s) you use)

- True/False
- Matching
- Multiple Choice
- Completion
- Short Answer
- Essay

6. Predict secondary storage requirements.

Is this outcome content expressed in your CIS course syllabus?

- No
- Yes

Do you test this content in your course?

- No (If "No", Go to Part II on Page 4)
- Yes (If yes, select the primary type or types of test(s) you use)

- True/False
- Matching
- Multiple Choice
- Completion
- Short Answer
- Essay

Please respond to the following statements by circling the appropriate number.

Your highest academic degree is:

1. associate
2. bachelor's
3. master's
4. doctoral degree

Your current academic rank is:

1. not applicable on this campus
2. lecturer, adjunct, assistant instructor
3. instructor
4. assistant professor
5. associate professor
6. professor

How many sections of Introduction to Computers have you taught in the previous academic year (1991-92 academic year) ?

1. none
2. this is the first time I'm teaching Intro to Computers
3. one to three sections
4. four to six sections
5. seven to nine sections
6. more than nine sections

How many years have you taught Introduction to Computers as part of your regular teaching load?

1. none
2. this is the first year I taught Introduction to Computers
3. one to two years
4. three to four years
5. five to six years
6. more than six years

Please record your response in the space provided to the following set of questions.

Have you taken a curriculum course in the past three years?

_____ No _____ Yes

Have you taken a test and measurement course in the past three years?

_____ No _____ Yes

How many years have you worked in each of the following positions? (indicate full years; complete all that apply)

- _____ Full-time college teacher
- _____ Graduate teaching assistant in college
- _____ High school teacher
- _____ Elementary school teacher
- _____ Teacher/trainer in business or industry

How many years have you held your present teaching position?

_____ years

Please provide the following information:

Your Name _____

Title _____

School _____

Address _____

Phone Number _____

**This concludes the survey. Thank you for responding.
Place this document in the postage-paid envelope provided.**

APPENDIX B

**PHASE II SURVEY:
COVER LETTER, INSTRUCTIONS, AND INSTRUMENT**

750 Hethwood Blvd.
Apartment 300-L
Blacksburg, VA 24060
January 26, 1993

Mr.
Wytheville Community College
1000 East Main Street
Wytheville, VA 24382

Dear :

Thank you for responding to the first phase of this study. The instrument for completing the second and final phase is enclosed. The second phase simply reports the collective responses obtained in the first phase and asks if you agree or disagree with them. It should take you approximately five minutes to complete.

The purpose and directions for completing Phase Two are provided on the next page. The curriculum model used in this study recommends certain types of tests be used with certain types of planned content. You may or may not agree with the curriculum model. That's fine. If you do not agree, I'd like to know why.

As with Phase One, your responses will be grouped and tabulated with other respondents so that individual institutions and individual respondents can not be identified. Please be sure to provide your name and institution in the spaces provided on the last page in case I need to discuss your comments with you.

As I indicated to you previously, this study is based on a curriculum system called Performance Instruction. For participating, I will mail you a book containing the principles which undergird the curriculum model used in this study. Please allow three to four weeks to receive this book.

I am grateful for your participation. A postage-paid return envelope is provided. Please complete and return this instrument by **February 3, 1993**.

Sincerely,

Thomas Leitzel

Enclosures

**Instructor Planning and Testing Selection Inventory
Phase II**

Purpose: The purpose of this survey is to validate the results obtained in Phase I by having you offer explanations confirming variations in faculty instructional testing practices.

The basis for making correct planning and testing decisions in this study is tied to an outcome-based curriculum model which is grounded in pedagogy theory. In essence, this model posits that testing of content should occur at the same domain level as the content plans.

The "expert system" in this computer-based curriculum model establishes three distinct levels in the cognitive domain. These levels are (1) fact, (2) understanding, and (3) application. Facts represent the simplest level and involve the teaching of a single concept. Similarly, the second level, understanding, puts two or more concepts together; application puts two or more concepts together to form something new.

For example, if the content you planned were to have students specify document printer speeds, you would test their knowledge of this content through either a true-false or completion test. The content goal verb "specify" which was used to plan instruction is classified as level one. To be consistent with the plans, the type of test selected also has to be a level one. True-false and completion tests are level one tests. The following table shows the test types to be used with content plans in order to have consistency between planning and testing.

<u>Domain Level</u>	<u>Type of Test</u>
1 Fact	True-False Completion
2 Understanding	Matching Short Answer
3 Application	Multiple Choice Essay

On the next page, the content goals used in Phase I are listed along with test types consistent with the curriculum model. Then, Phase I faculty responses are reported showing the proportion of test types computer faculty indicated they would use to test the planned content.

Directions: Examine the expert system recommendations and the faculty responses. Indicate, by placing an "X" in the appropriate place, if you agree or disagree with the curriculum model. If you disagree, please tell me why.

Turn The Page To Begin The Survey

Instructor Planning and Testing Selection Inventory

1. Content Goal: Define input device terms.

Primary test types selected from Phase I:

<u> 27%</u> True/False	<u> 10%</u> Completion
<u> 37%</u> Matching	<u> 17%</u> Short Answer
<u> 77%</u> Multiple Choice	<u> 10%</u> Essay

Curriculum Model:

<u>Cognitive Level</u>	<u>Expert System Test Type Recommendations</u>
One	True-False, Completion

Question: Do you agree or disagree with expert system test types?

 Agree (Go to goal statement 2)

 Disagree (If you disagree, please indicate one or more of the following reasons)

A. Want to purposely undertest content

B. Want to purposely overtest content

C. Want to guarantee employers that students have marketable skills when they graduate

D. Want to keep students in program

E. Personal dislike for above listed test types

F. Difficulty with terminology

Other reasons: (please state):

2. Content Goal: Compare microcomputer software applications.

Primary test types selected from Phase I:

<u> 20%</u> True/False	<u> 7%</u> Completion
<u> 20%</u> Matching	<u> 30%</u> Short Answer
<u> 53%</u> Multiple Choice	<u> 20%</u> Essay

Curriculum Model:

<u>Cognitive Level</u>	<u>Expert System Test Type Recommendations</u>
Two	Matching, Short Answer

Question: Do you agree or disagree with expert system test types?

- Agree (Go to goal statement 3)
- Disagree (If you disagree, please indicate one or more of the following reasons)
 - A. Want to purposely undertest content
 - B. Want to purposely overtest content
 - C. Want to guarantee employers that students have marketable skills when they graduate
 - D. Want to keep students in program
 - E. Personal dislike for above listed test types
 - F. Difficulty with terminology

Other reasons: (please state):

3. Content Goal: Analyze system design components.

Primary test types selected from Phase I:

<u>10%</u> True/False	<u>7%</u> Completion
<u>3%</u> Matching	<u>23%</u> Short Answer
<u>37%</u> Multiple Choice	<u>23%</u> Essay

Curriculum Model:

<u>Cognitive Level</u>	<u>Expert System Test Type Recommendations</u>
Three	Multiple Choice, Essay

Question: Do you agree or disagree with expert system test types?

- Agree (Go to goal statement 4)
- Disagree (If you disagree, please indicate one or more of the following reasons)
 - A. Want to purposely undertest content
 - B. Want to purposely overtest content
 - C. Want to guarantee employers that students have marketable skills when they graduate
 - D. Want to keep students in program
 - E. Personal dislike for above listed test types
 - F. Difficulty with terminology

Other reasons: (please state):

4. Content Goal: Recall flowchart symbols.

Primary test types selected from Phase I:

<u> 17%</u> True/False	<u> 3%</u> Completion
<u> 13%</u> Matching	<u> 10%</u> Short Answer
<u> 27%</u> Multiple Choice	<u> 0%</u> Essay

Curriculum Model:

Cognitive Level Expert System Test Type Recommendations

One

True-False, Completion

Question: Do you agree or disagree with expert system test types?

 Agree (Go to goal statement 5)

 Disagree (If you disagree, please indicate one or more of the following reasons)

A. Want to purposely undertest content

B. Want to purposely overtest content

C. Want to guarantee employers that students have marketable skills when they graduate

D. Want to keep students in program

E. Personal dislike for above listed test types

F. Difficulty with terminology

Other reasons: (please state):

5. Content Goal: Classify computer operations.

Primary test types selected from Phase I:

<u> 23%</u> True/False	<u> 3%</u> Completion
<u> 20%</u> Matching	<u> 23%</u> Short Answer
<u> 63%</u> Multiple Choice	<u> 10%</u> Essay

Curriculum Model:

<u>Cognitive Level</u>	<u>Expert System Test Type Recommendations</u>
Two	Matching, Short Answer

Question: Do you agree or disagree with expert system test types?

 Agree (Go to goal statement 6)

 Disagree (If you disagree, please indicate one or more of the following reasons)

A. Want to purposely undertest content

B. Want to purposely overtest content

C. Want to guarantee employers that students have marketable skills when they graduate

D. Want to keep students in program

E. Personal dislike for above listed test types

F. Difficulty with terminology

Other reasons: (please state):

6. Content Goal: Predict secondary storage requirements.

Primary test types selected from Phase I:

<u> 10%</u> True/False	<u> 0%</u> Completion
<u> 13%</u> Matching	<u> 20%</u> Short Answer
<u> 27%</u> Multiple Choice	<u> 7%</u> Essay

Curriculum Model:

<u>Cognitive Level</u>	<u>Expert System Test Type Recommendations</u>
Three	Multiple Choice, Essay

Question: Do you agree or disagree with expert system test types?

- Agree (Go to Part II)
- Disagree (If you disagree, please indicate one or more of the following reasons)
- A. Want to purposely undertest content
 - B. Want to purposely overtest content
 - C. Want to guarantee employers that students have marketable skills when they graduate
 - D. Want to keep students in program
 - E. Personal dislike for above listed test types
 - F. Difficulty with terminology

Other reasons: (please state):

Part II. - Demographics: Please provide name.

Name _____

**This Concludes The Survey. Thank Your For Participating.
Place Instrument In Postage Paid Envelope Provided.**

APPENDIX C

**SURVEY RATE OF RETURN GRAPHS
(PHASE I AND II)**

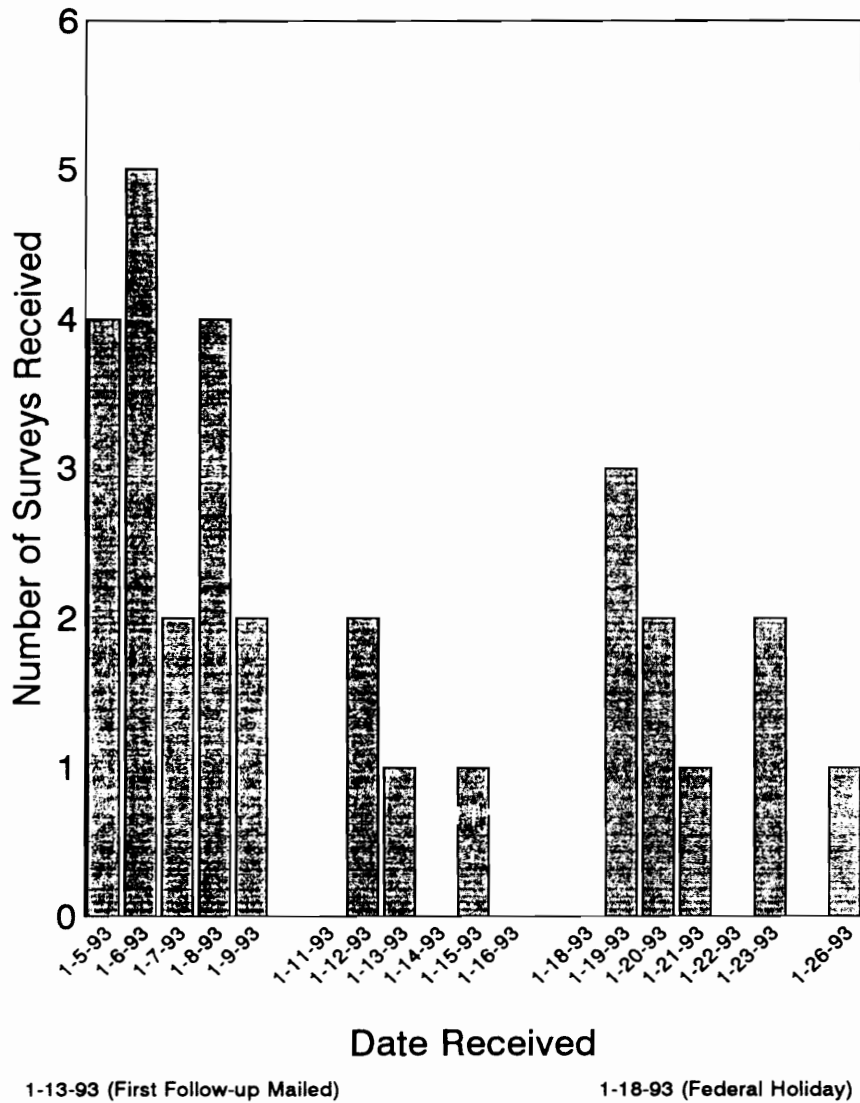


Figure 1
Phase I Receipt Control Graph

APPENDIX D
SURVEY FOLLOWUP LETTERS

750 Hethwood Boulevard
Apartment 300 - L
Blacksburg, VA 24060
January 13, 1993

Mr.
Wytheville Community College
1000 East Main Street
Wytheville, VA 24382

Dear :

If your responses to the enclosed questionnaire are in the mail, please accept my sincere thanks and disregard this mailing. If you did not return the questionnaire, I hope that you will let me know your opinions concerning content planning and testing.

Since you were selected among 29 other participants at 14 institutions in the VCCS, my concern is that a non-response may bias the results. I would like to include your responses in my analysis.

A copy of the instrument is enclosed. Please take a few minutes required to mark your responses and return the questionnaire in the envelope provided by January 21. Please be assured that your responses will be grouped and tabulated along with all other respondents to preserve individual identity.

Sincerely,

Thomas Leitzel

Enclosures: Survey
Return Envelope

750 Hethwood Boulevard
Apartment 300 - L
Blacksburg, VA 24060
February 4, 1993

Mr.
Wytheville Community College
1000 East Main Street
Wytheville, VA 24382

Dear :

If your responses to the enclosed questionnaire are in the mail, please accept my sincere thanks and disregard this mailing. If you did not return the questionnaire, I hope that you will let me know your opinions concerning content planning and testing.

Phase I returns yielded a 100 percent response rate. I hope to obtain an identical response rate from the second phase. Your cooperation would greatly help. With a strong response from Phase I, my concern is that a non-response may bias Phase II results.

A copy of the instrument is enclosed. Please take a few minutes required to mark your responses and return the questionnaire in the envelope provided by **February 12**. Please be assured that your responses will be grouped and tabulated along with all other respondents to preserve individual identity.

Sincerely,

Thomas Leitzel

Enclosures: Survey
Return Envelope

APPENDIX E
PARTICIPANT THANK YOU LETTER

750 Hethwood Blvd.
Apartment 300-L
Blacksburg, VA 24060
February 17, 1993

Mr.
Wytheville Community College
1000 East Main Street
Wytheville, VA 24382

Dear :

Thank you for participating in the survey of faculty planning and testing decisions. Your responses proved to be very valuable to my study. Essentially, what I was testing was a concept called platform unity. By definition, this is the specific alignment of planned, domain-referenced content with associated test types. It is closely related to a concept called curriculum alignment.

As promised, I'm enclosing a book called Performance Instruction. Dr. Vogler's model follows the concept of curriculum alignment. This concept states that planned content ought to be the delivered content as well as the tested content. A course, Special Problems in Education (EDAE 6914), will be taught this summer in Roanoke and this fall in Blacksburg. This concept is included in the three credit hour course. Dr. Vogler, a professor in the Community College Program Area at Virginia Tech, is teaching the course. For more information, you may contact him at 703-231-9702.

I'd be delighted to discuss any aspect of the research. My telephone number is 703-951-5343. Again, thank you for responding to my questionnaires. Enjoy the book.

Sincerely,

Thomas Leitzel
Graduate Student
Virginia Tech

Enclosure

APPENDIX F
QUESTIONNAIRES FOR FIELD TEST
(PHASE I AND II)

Instructor Planning and Testing Selection Inventory

Field Test of Instrument

This instrument is designed to be completed by faculty who teach an introductory-level computer literacy course at the postsecondary level. For this field test, assume you are a computer faculty member teaching a basic, computer literacy course. Further, assume the course number for this course is CIS 101.

In addition to the instrument (numbered pages 1-5), there are additional questions to examine. Since this instrument is being field tested, please respond to these additional questions along with completing the actual instrument. Provide any comments that would strengthen or clarify this instrument. I intend to send this instrument on-line to computer faculty in the Virginia Community College System (VCCS) through a users network called PROFS.

Please be sure to provide your name and phone number in the spaces provided on the last page (page 12) in case I need to discuss your comments with you. Thank you for assisting me with this project. The next eight pages contain the actual instrument plus directions.

A postage-paid return envelope is provided. You will assist me greatly by completing and returning the instrument by November 20.

**Field Test
Individual Analysis of Questions**

During this field test phase, please answer the following questions regarding each item presented in the instrument.

Directions

Were the directions clear?

_____No _____Yes

Were you able to complete the instrument without difficulty?

_____No _____Yes

Content Goal 1: Define input device terms.

Is this question clear?

_____No _____Yes

Do you understand the question?

_____No _____Yes

Do you personally possess enough information to answer it?

_____No _____Yes

Were the response stubs adequately framed?

_____No _____Yes

Is there any reason a person like you wouldn't answer the question?

_____No _____Yes

If you answered "YES", what is a better way to ask the question?

(restate question here) _____

Content Goal 2: Compare computer language usage.

Is this question clear?

No Yes

Do you understand the question?

No Yes

Do you personally possess enough information to answer it?

No Yes

Were the response stubs adequately framed?

No Yes

Is there any reason a person like you wouldn't answer the question?

No Yes

If you answered "YES", what is a better way to ask the question? (restate question here)

Content Goal 3: Analyze system design components.

Is this question clear?

No Yes

Do you understand the question?

No Yes

Do you personally possess enough information to answer it?

No Yes

Were the response stubs adequately framed?

No Yes

Is there any reason a person like you wouldn't answer the question?

No Yes

If you answered "YES", what is a better way to ask the question? (restate question here)

Content Goal 4: Recall flowchart symbols.

Is this question clear?

No Yes

Do you understand the question?

No Yes

Do you personally possess enough information to answer it?

No Yes

Were the response stubs adequately framed?

No Yes

Is there any reason a person like you wouldn't answer the question?

No Yes

If you answered "YES", what is a better way to ask the question? (restate question here)

Content Goal 5: Classify computer operations.

Is this question clear?

No Yes

Do you understand the question?

No Yes

Do you personally possess enough information to answer it?

No Yes

Were the response stubs adequately framed?

No Yes

Is there any reason a person like you wouldn't answer the question?

No Yes

If you answered "YES", what is a better way to ask the question? (restate question here)

Content Goal 6: Predict secondary storage requirements.

Is this question clear?

No Yes

Do you understand the question?

No Yes

Do you personally possess enough information to answer it?

No Yes

Were the response stubs adequately framed?

No Yes

Is there any reason a person like you wouldn't answer the question?

No Yes

If you answered "YES", what is a better way to ask the question? (restate question here)

Part II. Demographics Section

Question: Your highest academic degree. (this response asked you to circle your highest academic degree)

Is this question clear?

No Yes

Do you understand the question?

No Yes

Do you personally possess enough information to answer it?

No Yes

Were the response stubs adequately framed?

No Yes

Question: Your current academic rank.

Is this question clear?

No Yes

Do you understand the question?

No Yes

Do you personally possess enough information to answer it?

No Yes

Were the response stubs adequately framed?

No Yes

Question: Number of times you taught CIS 100 in previous academic year?

Is this question clear?

No Yes

Do you understand the question?

No Yes

Do you personally possess enough information to answer it?

No Yes

Were the response stubs adequately framed?

No Yes

Question: Number of years you taught CIS 100 as part of regular teaching load?

Is this question clear?

No Yes

Do you understand the question?

No Yes

Do you personally possess enough information to answer it?

No Yes

Were the response stubs adequately framed?

No Yes

Question: Have you had a curriculum course in the past three years?

Is this question clear?

No Yes

Do you understand the question?

No Yes

Do you personally possess enough information to answer it?

No Yes

Were the response stubs adequately framed?

No Yes

Question: Have you had a test and measurement course in the past three years?

Is this question clear?

No Yes

Do you understand the question?

No Yes

Do you personally possess enough information to answer it?

No Yes

Were the response stubs adequately framed?

No Yes

Question: How many years have you worked in each of the positions listed?

Is this question clear?

No Yes

Do you understand the question?

No Yes

Do you personally possess enough information to answer it?

No Yes

Were the response stubs adequately framed?

No Yes

Question: How many years have you held your present teaching position?

Is this question clear?

No Yes

Do you understand the question?

No Yes

Do you personally possess enough information to answer it?

No Yes

Were the response stubs adequately framed?

No Yes

Overall Instrument

Is the organization of the overall instrument free from distractions?

No Yes

Please note any distracting features of the instrument.

What would you recommend I do to the instrument to improve it?

Field Test Respondents: Please provide the following information

Your Name _____

Title _____

School _____

Address _____

Phone Number _____

Instructor Planning and Testing Selection Inventory

Field Test of Instrument Phase II

Note to field-test subjects:

This instrument, which reports findings obtained in Phase I, is designed to be completed by faculty who teach an introductory-level computer literacy course at the postsecondary level. For this field test, assume you are a computer faculty member teaching a basic, computer literacy course.

In addition to the instrument (numbered pages 1-6), there are additional questions to examine. Since this instrument is being field tested, please respond to these additional questions along with completing the actual instrument. Provide any comments that would strengthen or clarify this instrument. I intend to mail this survey to participants.

Please be sure to provide your name and phone number in the spaces provided in case I need to discuss your comments with you. Thank you for assisting me with this project.

A postage-paid return envelope is provided. You will assist me greatly by completing and returning the instrument by December 20.

Field Test
Individual Analysis of Questions

During this field test phase, please answer the following questions regarding each item presented in the instrument.

Purpose:

Is the purpose of the survey clear?

No Yes

Explanation:

Whether you agree or disagree with the curriculum model:

Is the explanation of domain level planning clear?

No Yes

Is the explanation of appropriate test types clear?

No Yes

Directions:

Are the directions for completing the instrument clear?

No Yes

Were you able to complete the instrument without difficulty?

No Yes

Expert System:

Do you understand what is meant by "expert system"?

No Yes

Response Stems:

Are the response stems (items A-F) complete and exhaustive?

No Yes

Should there be additional response stems listed?

No Yes (If yes, please suggest additional response stems below)

In the space below, please identify any areas you believe respondents would have difficulty completing. Feel free to make any suggestions.

Thank you for your assistance with this instrument.

Place the instrument in the envelope provided.

APPENDIX G
VIRGINIA COMMUNITY COLLEGES

Table 1

Virginia Community Colleges

<u>College</u>	<u>Location</u>
1. Blue Ridge CC	Weyers Cave
2. Central VA CC	Lynchburg
3. Dabney S. Lancaster CC	Clifton Forge
4. Danville CC	Danville
5. Eastern Shore CC	Melfa
6. Germanna CC	Locust Grove
7. J. Sargeant Reynolds CC	Richmond
8. John Tyler CC	Chester
9. Lord Fairfax CC	Middletown
10. Mountain Empire CC	Big Stone Gap
11. New River CC	Dublin
12. Northern VA CC	Annandale
13. Patrick Henry CC	Martinsville
14. Paul D. Camp CC	Franklin
15. Piedmont VA CC	Charlottesville
16. Rappahannock CC	Glenns
17. Southside VA CC	Alberta
18. Southwest VA CC	Richlands
19. Thomas Nelson CC	Hampton
20. Tidewater CC	Virginia Beach
21. Virginia Highlands CC	Abingdon
22. Virginia Western CC	Roanoke
23. Wytheville CC	Wytheville

Table 2

Randomly Selected
Institutions Participating in the Study

<u>College</u>	<u>Location</u>
Blue Ridge CC	Weyers Cave
Central VA CC	Lynchburg
Dabney S. Lancaster CC	Clifton Forge
Danville CC	Danville
J. Sargeant Reynolds CC	Richmond
Lord Fairfax CC	Middletown
Mountain Empire CC	Big Stone Gap
New River CC	Dublin
Patrick Henry CC	Martinsville
Piedmont VA CC	Charlottesville
Southwest VA CC	Richlands
Thomas Nelson CC	Hampton
Virginia Highlands CC	Abingdon
Virginia Western CC	Roanoke
Wytheville CC	Wytheville

APPENDIX H
FORMULAE CALCULATION SPREADSHEETS

Domain Level Code: 1

True-False =1 Completion =1
 Matching =2 Short Answer=2
 Multiple Choice=3 Essay =3

Test Types	Sum of Test Points	Number of Test Types	Unity Index	EFT Index	EFF Index
True-False	1	1	0	0	0
Matching	2	1	1	0	-2
Multiple Choice	3	1	2	0	-4
Completion	1	1	0	0	0
Short Answer	2	1	1	0	-2
Essay	3	1	2	0	-4
True-False, Matching	3	2	0.5	0	-1
True-False, Multiple Choice	4	2	1	0	-2
True-False, Completion	2	2	0	0	0
True-False, Short Answer	3	2	0.5	0	-1
True-False, Essay	4	2	1	0	-2
Matching, Multiple Choice	5	2	1.5	0	-1
Matching, Completion	3	2	0.5	0	-2
Matching, Short Answer	4	2	1	0	-2
Matching, Essay	5	2	1.5	0	-2
Multiple Choice, Completion	4	2	1	0	-2
Multiple Choice, Short Answer	5	2	1.5	0	-3
Multiple Choice, Essay	6	2	2	0	-4
Completion, Short Answer	3	2	0.5	0	-1
Completion, Essay	4	2	1	0	-3
Short Answer, Essay	5	2	1.5	0	-3

Test Types	Sum of Test Points	Number of Test Types	Unity Index	EFT Index	EFF Index
True-False, Matching, Multiple Choice	6	3	1	0	-2
True-False, Matching, Completion	4	3	0.33	0	-0.67
True-False, Matching, Short Answer	5	3	0.67	0	-1.33
True-False, Matching, Essay	6	3	1	0	-2
True-False, Multiple Choice, Completion	5	3	0.67	0	-1.33
True-False, Multiple Choice, Short Answer	6	3	1	0	-2
True-False, Multiple Choice, Essay	7	3	1.33	0	-2.67
True-False, Completion, Short Answer	4	3	0.33	0	-0.67
True-False, Completion, Essay	5	3	0.67	0	-1.33
True-False, Short Answer, Essay	6	3	1	0	-2
Matching, Multiple Choice, Completion	6	3	1	0	-2
Matching, Multiple Choice, Short Answer	7	3	1.33	0	-2.67
Matching, Multiple Choice, Essay	8	3	1.67	0	-3.33
Matching, Completion, Short Answer	5	3	0.67	0	-1.33
Matching, Completion, Essay	6	3	1	0	-2
Matching, Short Answer, Essay	7	3	1.33	0	-2.67
Multiple Choice, Completion, Short Answer	6	3	1	0	-2

Test Types	Sum of Test Points	Number of Test Types	Unity Index	EFT Index	EFF Index
Multiple Choice, Completion, Essay	7	3	1.33	0	-2.67
Multiple Choice, Short Answer, Essay	8	3	1.67	0	-3.33
Completion, Short Answer, Essay	6	3	1	0	-2
True-False, Matching, Multiple Choice, Completion	7	4	0.75	0	-1.5
True-False, Matching, Multiple Choice, Short Answer	8	4	1	0	-2
True-False, Matching, Multiple Choice, Essay	9	4	1.25	0	-2.5
True-False, Matching, Completion, Short Answer	6	4	0.5	0	-1
True-False, Matching, Completion, Essay	7	4	0.75	0	-1.5
True-False, Matching, Short Answer, Essay	8	4	1	0	-2
True-False, Multiple Choice, Completion, Short Answer	7	4	0.75	0	-1.5
True-False, Multiple Choice, Completion, Essay	8	4	1	0	-2
True-False, Multiple Choice, Short Answer, Essay	9	4	1.25	0	-2.5
True-False, Completion, Short Answer, Essay	7	4	0.75	0	-1.5
Matching, Multiple Choice, Completion, Short Answer	8	4	1	0	-2

Test Types	Sum of Test Points	Number of Test Types	Unity Index	EFT Index	EFF Index
Matching, Multiple Choice, Completion, Essay	9	4	1.25	0	-2.5
Matching, Multiple Choice, Short Answer, Essay	10	4	1.5	0	-3
Matching, Completion, Short Answer, Essay	8	4	1	0	-2
Multiple Choice, Completion, Short Answer, Essay	9	4	1.25	0	-2.5
True-False, Matching, Multiple Choice, Completion, Short Answer	9	5	0.8	0	-1.6
True-False, Matching, Multiple Choice, Completion, Essay	10	5	1	0	-2
True-False, Matching, Multiple Choice, Short Answer, Essay	11	5	1.2	0	-2.4
True-False, Matching, Completion, Short Answer, Essay	9	5	0.8	0	-1.6
True-False, Multiple Choice, Completion, Short Answer, Essay	10	5	1	0	-2
Matching, Multiple Choice, Completion, Short Answer, Essay	11	5	1.2	0	-2.4
True-False, Matching, Multiple Choice, Completion, Short Answer, Essay	12	6	1	0	-2

Domain Level Code: 2

True-False =1 Completion =1
 Matching =2 Short Answer =2
 Multiple Choice =3 Essay =3

Test Types	Sum of Test Points	Number of Test Types	Unity Index	EPT Index	EPF Index
True-False	1	1	-1	2	0
Matching	2	1	0	0	0
Multiple Choice	3	1	1	0	-2
Completion	1	1	-1	2	0
Short Answer	2	1	0	0	0
Essay	3	1	1	0	-2
True-False, Matching	3	2	-0.5	1	0
True-False, Multiple Choice	4	2	0	0	0
True-False, Completion	2	2	-1	2	0
True-False, Short Answer	3	2	-0.5	1	0
True-False, Essay	4	2	0	0	0
Matching, Multiple Choice	5	2	0.5	0	-1
Matching, Completion	3	2	-0.5	1	0
Matching, Short Answer	4	2	0	0	0
Matching, Essay	5	2	0.5	0	-1
Multiple Choice, Completion	4	2	0	0	0
Multiple Choice, Short Answer	5	2	0.5	0	-1
Multiple Choice, Essay	6	2	1	0	-2
Completion, Short Answer	3	2	-0.5	1	0
Completion, Essay	4	2	0	0	0
Short Answer, Essay	5	2	0.5	0	-1

Test Types	Sum of Test Points	Number of Test Types	Unity Index	EFT Index	EFF Index
True-False, Matching, Multiple Choice	6	3	0	0	0
True-False, Matching, Completion	4	3	-0.67	1.33	0
True-False, Matching, Short Answer	5	3	-0.33	0.67	0
True-False, Matching, Essay	6	3	0	0	0
True-False, Multiple Choice, Completion	5	3	-0.33	0.67	0
True-False, Multiple Choice, Short Answer	6	3	0	0	0
True-False, Multiple Choice, Essay	7	3	0.33	0	-0.67
True-False, Completion, Short Answer	4	3	-0.67	1.33	0
True-False, Completion, Essay	5	3	-0.33	0.67	0
True-False, Short Answer, Essay	6	3	0	0	0
Matching, Multiple Choice, Completion	6	3	0	0	0
Matching, Multiple Choice, Short Answer	7	3	0.33	0	-0.67
Matching, Multiple Choice, Essay	8	3	0.67	0	-1.33
Matching, Completion, Short Answer	5	3	-0.33	0.67	0
Matching, Completion, Essay	6	3	0	0	0
Matching, Short Answer, Essay	7	3	0.33	0	-0.67
Multiple Choice, Completion, Short Answer	6	3	0	0	0

Test Types	Sum of Test Points	Number of Test Types	Unity Index	EFT Index	EFF Index
Multiple Choice, Completion, Essay	7	3	0.33	0	-0.67
Multiple Choice, Short Answer, Essay	8	3	0.67	0	-1.33
Completion, Short Answer, Essay	6	3	0	0	0
True-False, Matching, Multiple Choice, Completion	7	4	-0.25	0.5	0
True-False, Matching, Multiple Choice, Short Answer	8	4	0	0	0
True-False, Matching, Multiple Choice, Essay	9	4	0.25	0	-0.5
True-False, Matching, Completion, Short Answer	6	4	-0.5	1	0
True-False, Matching, Completion, Essay	7	4	-0.25	0.5	0
True-False, Matching, Short Answer, Essay	8	4	0	0	0
True-False, Multiple Choice, Completion, Short Answer	7	4	-0.25	0.5	0
True-False, Multiple Choice, Completion, Essay	8	4	0	0	0
True-False, Multiple Choice, Short Answer, Essay	9	4	0.25	0	-0.5
True-False, Completion, Short Answer, Essay	7	4	-0.25	0.5	0
Matching, Multiple Choice, Completion, Short Answer	8	4	0	0	0

Test Types	Sum of Test Points	Number of Test Types	Unity Index	EFT Index	EFF Index
Matching, Multiple Choice, Completion, Essay	9	4	0.25	0	-0.5
Matching, Multiple Choice, Short Answer, Essay	10	4	0.5	0	-1
Matching, Completion, Short Answer, Essay	8	4	0	0	0
Multiple Choice, Completion, Short Answer, Essay	9	4	0.25	0	-0.5
True-False, Matching, Multiple Choice, Completion, Short Answer	9	5	-0.2	0.4	0
True-False, Matching, Multiple Choice, Completion, Essay	10	5	0	0	0
True-False, Matching, Multiple Choice, Short Answer, Essay	11	5	0.2	0	-0.4
True-False, Matching, Completion, Short Answer, Essay	9	5	-0.2	0.4	0
True-False, Multiple Choice, Completion, Short Answer, Essay	10	5	0	0	0
Matching, Multiple Choice, Completion, Short Answer, Essay	11	5	0.2	0	-0.4
True-False, Matching, Multiple Choice, Completion, Short Answer, Essay	12	6	0	0	0

Domain Level Code: 3

True-False =1 Completion =1
 Matching =2 Short Answer =2
 Multiple Choice =3 Essay =3

Test Types	Sum of Test Points	Number of Test Types	Unity Index	EFT Index	EFF Index
True-False	1	1	-2	4	0
Matching	2	1	-1	2	0
Multiple Choice	3	1	0	0	0
Completion	1	1	-2	4	0
Short Answer	2	1	-1	2	0
Essay	3	1	0	0	0
True-False, Matching	3	2	-1.5	3	0
True-False, Multiple Choice	4	2	-1	2	0
True-False, Completion	2	2	-2	4	0
True-False, Short Answer	3	2	-1.5	3	0
True-False, Essay	4	2	-1	2	0
Matching, Multiple Choice	5	2	-0.5	1	0
Matching, Completion	3	2	-1.5	3	0
Matching, Short Answer	4	2	-1	2	0
Matching, Essay	5	2	-0.5	1	0
Multiple Choice, Completion	4	2	-1	2	0
Multiple Choice, Short Answer	5	2	-0.5	1	0
Multiple Choice, Essay	6	2	0	0	0
Completion, Short Answer	3	2	-1.5	3	0
Completion, Essay	4	2	-1	2	0
Short Answer, Essay	5	2	-0.5	1	0

Test Types	Sum of Test Points	Number of Test Types	Unity Index	EFT Index	EPF Index
True-False, Matching, Multiple Choice	6	3	-1	2	0
True-False, Matching, Completion	4	3	-1.67	3.33	0
True-False, Matching, Short Answer	5	3	-1.33	2.67	0
True-False, Matching, Essay	6	3	-1	2	0
True-False, Multiple Choice, Completion	5	3	-1.33	2.67	0
True-False, Multiple Choice, Short Answer	6	3	-1	2	0
True-False, Multiple Choice, Essay	7	3	-0.67	1.33	0
True-False, Completion, Short Answer	4	3	-1.67	3.33	0
True-False, Completion, Essay	5	3	-1.33	2.67	0
True-False, Short Answer, Essay	6	3	-1	2	0
Matching, Multiple Choice, Completion	6	3	-1	2	0
Matching, Multiple Choice, Short Answer	7	3	-0.67	1.33	0
Matching, Multiple Choice, Essay	8	3	-0.33	0.67	0
Matching, Completion, Short Answer	5	3	-1.33	2.67	0
Matching, Completion, Essay	6	3	-1	2	0
Matching, Short Answer, Essay	7	3	-0.67	1.33	0
Multiple Choice, Completion, Short Answer	6	3	-1	2	0

Test Types	Sum of Test Points	Number of Test Types	Unity Index	EFT Index	EFF Index
Multiple Choice, Completion, Essay	7	3	-0.67	1.33	0
Multiple Choice, Short Answer, Essay	8	3	-0.33	0.67	0
Completion, Short Answer, Essay	6	3	-1	2	0
True-False, Matching, Multiple Choice, Completion	7	4	-1.25	2.5	0
True-False, Matching, Multiple Choice, Short Answer	8	4	-1	2	0
True-False, Matching, Multiple Choice, Essay	9	4	-0.75	1.5	0
True-False, Matching, Completion, Short Answer	6	4	-1.5	3	0
True-False, Matching, Completion, Essay	7	4	-1.25	2.5	0
True-False, Matching, Short Answer, Essay	8	4	-1	2	0
True-False, Multiple Choice, Completion, Short Answer	7	4	-1.25	2.5	0
True-False, Multiple Choice, Completion, Essay	8	4	-1	2	0
True-False, Multiple Choice, Short Answer, Essay	9	4	-0.75	1.5	0
True-False, Completion, Short Answer, Essay	7	4	-1.25	2.5	0
Matching, Multiple Choice, Completion, Short Answer	8	4	-1	2	0

Test Types	Sum of Test Points	Number of Test Types	Unity Index	EFT Index	EPF Index
Matching, Multiple Choice, Completion, Essay	9	4	-0.75	1.5	0
Matching, Multiple Choice, Short Answer, Essay	10	4	-0.5	1	0
Matching, Completion, Short Answer, Essay	8	4	-1	2	0
Multiple Choice, Completion, Short Answer, Essay	9	4	-0.75	1.5	0
True-False, Matching, Multiple Choice, Completion, Short Answer	9	5	-1.2	2.4	0
True-False, Matching, Multiple Choice, Completion, Essay	10	5	-1	2	0
True-False, Matching, Multiple Choice, Short Answer, Essay	11	5	-0.8	1.6	0
True-False, Matching, Completion, Short Answer, Essay	9	5	-1.2	2.4	0
True-False, Multiple Choice, Completion, Short Answer, Essay	10	5	-1	2	0
Matching, Multiple Choice, Completion, Short Answer, Essay	11	5	-0.8	1.6	0
True-False, Matching, Multiple Choice, Completion, Short Answer, Essay	12	6	-1	2	0

VITA

THOMAS CHARLES LEITZEL

EDUCATION

- 1993 Doctor of Philosophy (Community College Administration), Virginia Polytechnic Institute and State University, Blacksburg, VA
- 1982 Master of Science (Education), Temple University, Philadelphia, PA
- 1978 Bachelor in Arts (Business Administration), Lycoming College, Williamsport, PA
- 1976 Associate in Arts (Business Administration), Pennsylvania College of Technology, Williamsport, PA

PROFESSIONAL EXPERIENCE

- 1991-1992 Graduate Assistant, Virginia Polytechnic Institute and State University, Blacksburg, VA
- 1985-1991 Associate Dean of Business, Science, and Technology, Chesapeake College, Wye Mills, MD
- 1982-1985 Assistant Director, Business and Computer Technology Division, Williamsport Area Community College, Williamsport, PA
- 1981-1982 Evening and Weekend College Coordinator, Business and Computer Technology Division, Williamsport Area Community College, Williamsport, PA
- 1978-1981 Instructor in Retail Management, Williamsport Area Community College, Williamsport, PA

PERSONAL DATA

Married
Birthdate: May 6, 1956


Thomas Charles Leitzel