HumRRO

The Development of a New Research Paradigm for Studying Aptitude-Treatment Interactions

Mary Ann Lightfoot Statman
Monica A. Gribben
Jennifer A. Naughton
Rodney A. McCloy

19981020 005

Human Resources Research Organization
66 Canal Center Plaza, Suite 400 • Alexandria, VA 22314-1591

DTIC QUALITY INSPECTED 4

October 13, 1998

Defense Technical Information Center
Attn: Crystal Riley/Selection Group
8725 John J. Kingman Road
Suite 0944
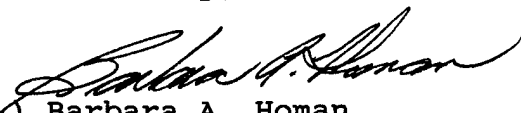Ft. Belvoir, VA 22060-6218

Dear Sirs:

Enclosed please find one unbound copy of "The Development of a New Research Paradigm for Studying Aptitude-Treatment Interactions," by Mary Ann Lightfoot Statman, Monica A. Gribben, Jennifer A. Naughton, Rodney A. McCloy.

We would like to have this report entered into the DTIC system. If possible, could someone return one copy of this letter to us with the AD number noted on it? I have enclosed a self-addressed envelope for this purpose.

Thanks so much for your assistance.

Sincerely,

(Mrs.) Barbara A. Homan
         Manager, Personnel &
         Administration

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

Public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden to Washington Headquarters Services, Directorate for information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302, and to the Office of Management and Budget, Paperwork Reduction Project (0704-0188), Washington, DC 20503

1. AGENCY USE ONLY *(Leave Blank)*	2. REPORT DATE July, 1998	3. REPORT TYPE AND DATES COVERED Final Report, September 1995 - July 1998

4. TITLE AND SUBTITLE
The Development of a New Research Paradigm for Studying Aptitude-Treatment Interactions

5. FUNDING NUMBERS
F41624-95-C-5027
Project 1121

6. AUTHOR(S)
Mary Ann Lightfoot Statman, Monica A. Gribben, Jennifer A. Naughton, Rodney A. McCloy

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
Human Resources Research Organization (HumRRO)
66 Canal Center Plaza, Suite 400
Alexandria, Virginia 22314

8. PERFORMING ORGANIZATION REPORT NUMBER
FR-WATSD-97-25

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)
Department of the Air Force
Air Force Materiel Command, AL/HR (HSC/PKRH)
7909 Lindbergh Drive
Brooks AFB, TX 78235-5352

10. SPONSORING/MONITORING AGENCY REPORT NUMBER

11. SUPPLEMENTARY NOTES

Contracting Officer's Technical Representative, Dr. Robert Yadrick

12a. DISTRIBUTION/AVAILABILITY STATEMENT

Approved for public release, distribution unlimited.

12b. DISTRIBUTION CODE

13. ABSTRACT *(Maximum 200 words)*

The purpose of this study was to adapt the personnel classification research paradigm to investigate aptitude treatment interactions (ATI) in training settings. This report is divided into Introduction and Method sections. We present in the Introduction a description of personnel classification theory, method, and research findings, and a review of the training literature. We describe in the Method the classification-ATI research design and the Training Characteristics Survey (TCS), both of which were developed in this project. The classification-ATI research design employs a person-treatment matching procedure to measure ATIs. The TCS is an instrument for measuring the variation in training settings due to course content, method of instruction, course difficulty, and occupation type. We believe that the classification-ATI research paradigm will improve ATI research, by providing a means for comparing several ATIs within a single study and producing a quantitative index of the practical effects of ATIs on training performance.

14. SUBJECT TERMS Aptitude treatment interaction, Student course matching, Person-job matching, Personnel classification, Training Evaluation

15. NUMBER OF PAGES
74

16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
Unclassified	Unclassified	Unclassified	UL

NSN 7540-01-280-5500

Standard Form 298 (Rev 2-89)
Prescribed by ANSI Std 239-18 298-102

THE DEVELOPMENT OF A NEW RESEARCH PARADIGM FOR STUDYING APTITUDE-TREATMENT INTERACTIONS

EXECUTIVE SUMMARY

Requirement:

The purpose of this project was to develop a new research method for studying *aptitude-treatment interactions* (ATIs). Our approach was to adapt to training evaluation research the personnel classification paradigm used in employment testing. We present in the Introduction of this report detailed descriptions of personnel classification theory, method, and research findings, and a review of the training literature. We describe in the Method the classification-ATI research paradigm we designed.

Procedure:

Our approach consisted of three phases. First, we conducted a review of the training literature to identify characteristics of training settings that showed potential for forming ATIs with learner characteristics. Our review, which spanned the literature of technical training, industrial and organizational psychology, educational psychology, and instructional design, is presented in the Introduction of this report.

Second, we used the training variables we identified to construct a *Training Characteristics Survey* (TCS) that quantifies variation in specific training characteristics across instructional settings. The TCS is designed to provide course-specific data that is used to compute prediction equations for a set of courses under investigation. Thus, it forms an integral component of the classification-ATI research method.

Third, we adapted the differential classification paradigm to ATI research by modifying the Johnson and Zeidner (1994) person-job matching simulation method. The classification-ATI design employs the cross-validation procedure recommended by Johnson and Zeidner, and a person-treatment matching process, which is implemented by linear programming software. In the event that sample sizes are small, we discuss the pros and cons of using *Monte Carlo synthetic data generation techniques* to increase sample size to support the cross-validation procedure.

The major modifications we made to the personnel classification method were the addition of multilevel regression and the TCS, which are used in combination to compute course-specific prediction equations. Multilevel regression is ideal for classification and ATI research because it uses treatment-specific variables to model variation in differential prediction equations, includes statistical tests of ATI terms, and can accommodate small samples with minimal instability to predictor weights.

The classification paradigm is well suited to ATI research, because it is based on a psychometric theorem that provides a novel way to conceptualize ATIs, and provides a quantitative measure of the practical impact of ATIs on training performance. Further, the classification-ATI research method can be expanded to estimate the dollar benefits to training budgets of student-course matching strategies that capitalize on ATIs.

Findings:

Since this was a methodological study to develop the classification-ATI research paradigm, we did not conduct analyses. However, the method is described in detail for immediate application in training evaluation research. The TCS is presented in the Appendix.

Utilization of Findings:

We conceived of this report as the basis for designing and conducting ATI research that employs a person-treatment matching procedure. The method also is appropriate for other types of training evaluation studies that involve comparative analyses of courses, because it provides a well-researched and psychometrically sound basis for quantifying training outcomes. The TCS was designed to measure the variation in specific variables common to most training settings. It can be used for a variety of purposes other than ATI research, such as revising or evaluating training courses.

THE DEVELOPMENT OF A NEW RESEARCH PARADIGM FOR STUDYING APTITUDE-TREATMENT INTERACTIONS

Table of Contents

Table of Contents (Cont).

List of Tables

List of Figures

THE DEVELOPMENT OF A NEW RESEARCH PARADIGM FOR STUDYING APTITUDE-TREATMENT INTERACTIONS

Chapter 1: Introduction

Definition of Aptitude-Treatment Interaction

One of the most important questions in designing and evaluating training is the effect of *aptitude treatment interactions* (*ATIs*) on performance (Cronbach & Snow, 1977; Goldstein, 1993). The study of ATIs examines the relationships between characteristics of the learner and the training environment. The ATI hypothesis states that no single learning environment is best for all students, but that individual differences in aptitudes, motivation, and other variables (e.g., learning styles), interact with situational variables associated with different learning settings to enhance or diminish training performance (Cronbach & Snow, 1977). An ATI is present when the slope of the regression line predicting the outcome measure for Treatment A differs statistically from that of Treatment B, using the same predictor information.

Cronbach and Snow (1977) defined aptitude in the context of ATI as "any characteristic of a person that forecasts his probability of success under a given treatment" (p. 6). This definition makes clear that Cronbach and Snow do not restrict the concept of aptitude in learning situations solely to cognitive abilities, and that a more appropriate term may be person-treatment interaction, because it encompasses all individual difference variables related to learning. Treatment has been defined as "any instructional strategy or combination of instructional strategies that structures information for the purpose of having students learn that information" (Parkhust, 1975, p. 42, cited in Thompson, Simonson, & Hargrave, 1992).

Savage, Williges, and Williges (1982) recognized that there is a fundamental problem in understanding and measuring ATIs, because training evaluation usually focuses on *group mean performance* in a single course (i.e., treatment) or set of courses, rather than on the *differential performance of individuals in alternative courses or training environments*. They stated that:

> Skill training is usually an *individual* rather than a *group* experience, [however,] research to evaluate training procedures usually employs *group statistics* in which a fixed population of students is assumed and the training alternative producing the highest *mean performance* is sought. Unfortunately, in many cases the training approach selected does not provide optimal training for each of the *individual students* (p. 417, [italics added]).

The purpose of this report is to present a new paradigm for studying ATIs. The research method we describe is a modification of the differential classification research paradigm, which we transported from the personnel testing literature and adapted to training settings. Differential personnel classification refers to the assessment of job applicants for many different jobs or occupations at the same level within an organization, and the matching of each person to the job for which he or she is predicted to be most successful. The term differential personnel

classification has been used by the military services for many years to refer to their recruit-job-assignment procedure. However, a more general term for this process is person-job (or occupation) matching.

Organization of the Report

This report is divided into two sections, Introduction and Method. Our objective in the Introduction is to establish a firm foundation for the adaptation of the classification paradigm to training evaluation. We begin the Introduction with a discussion of the similarities in the concepts of ATI in training and person-job matching in employment testing. We follow this with descriptions of differential classification theory, methodology and research findings. We conclude the Introduction with a review of the training literature that we conducted as the basis for designing the *Training Characteristics Survey (TCS)*. The TCS is an instrument we propose be used in ATI research to identify and measure specific training variables that may interact with learner characteristics. We believe that the TCS will improve ATI research by providing insight into those aspects of learning environments that do and do not interact with student characteristics.

The Method presents a complete description of the classification-ATI research paradigm we developed in the project as the foundation of this report. Our objective is to show that the personnel classification paradigm, as we have adapted it for training evaluation research, provides a more sensitive, accurate and informative basis for detecting and explaining ATIs than the traditional ATI research method of comparing simple regression equations across multiple treatments. We begin the Method Chapter with descriptions of the TCS development process and *multilevel regression (MLR)*. These are followed by suggestions for a course sampling procedure, criterion variables, and predictors. The final section of the Method Chapter is a detailed explanation of the student-course matching simulation procedure, which forms the basis of the classification-ATI paradigm.

ATIs and Differential Personnel Classification

It is interesting to note that the distinction between group mean performance within treatment and differential individual performance across treatments raised by Savage, Williges, and Williges (1982) is also found in the person-job matching context (Statman, 1992, 1993). Typically, employment testing relies on a simple selection model for predicting performance in a single job or set of jobs. The selection model uses "group statistics" (i.e., multiple regression and correlation) to rank and choose candidates from the top down for a job. However, Brogden (1951) and Horst (1954) observed that most organizations would be better served by assessing applicants for multiple jobs and optimizing the match of each individual's pattern of abilities and interests to the occupation with the most congruent pattern of qualifications.

This *optimal person-job matching (OPJM)* model of employment testing suggested by Brogden (1959) and Horst (1954) is based upon differential classification theory, which

addresses the measurement of both intra-individual and inter-individual differences in performance in multiple occupations (Johnson & Zeidner, 1991). The measurement of intra-individual differences is accomplished by predicting each job candidate's success in a variety of occupations. Inter-individual differences are measured by rank-ordering all members of the applicant pool according to their potential success within each occupation. Practical applications of OPJM models (e.g., the military Services' recruit assignment systems) are implemented by an optimization algorithm that places each applicant in his or her best-fitting occupation, subject to practical constraints like adequate vacancies.

Cronbach, Snow and others (e.g., Cronbach & Gleser, 1965; Cronbach & Snow 1977; Snow & Lohman, 1984; Ward, 1983) recognized that personnel classification, or OPJM, in the employment testing context is analogous to the problem of matching students to appropriate training settings. Personnel classification is based on the premise that there is an interaction between worker characteristics (e.g., aptitudes, interests, motivation) and job characteristics (e.g., technical content, working conditions), making personnel classification a particular type of person-treatment interaction in which the treatment is occupation (Cronbach & Gleser, 1965; Cronbach & Snow, 1977; Ward, 1983). Both person-job and student-course matching processes attempt to capitalize on the interactions between individual characteristics and differential treatments. However, personnel classification researchers have focused heavily on optimizing the matching process to obtain gains in performance. In contrast, ATI researchers mainly have focused on trying to identify ATIs in different learning settings with a large number of measures of learner characteristics. (See Maldegen, Statman, Gribben, and Yadrick [1996] for a recent review of ATI research.)

In this report we propose that the personnel classification paradigm be used to study ATIs in learning settings. Our rationale was drawn from the observations described in the paragraph above that the classification proposition, which holds that worker and occupational characteristics interact, is equivalent to the ATI hypothesis. As we stated above, this hypothesis is that some contextual factors (e.g., method of instruction or difficulty of the material) differentially impact a student's learning-related characteristics to produce varying levels of success in different instructional settings. In other words, if every person were to perform equally well in every occupational or learning setting, then no person-treatment interactions would be present. If, however, some people tend to do better in some environments and worse in others, then some type of person-treatment interaction is responsible for this intra-individual variation in performance across settings.

Overview of the Personnel Classification Paradigm

The classification paradigm is a method for evaluating the benefits from optimally matching people to jobs. It produces a measure that compares optimally assigning people to one of several occupations with random assignment using no personnel or job information. Classification theory and methodology were developed through a continuously evolving process over a 50-year period beginning shortly after World War II. A number of researchers (Alley & Darby, 1995; Brogden, 1946, 1951, 1954, 1955, 1959, 1964; Horst, 1954, 1956; Hunter &

Schmidt, 1982; Johnson & Zeidner, 1991, 1994; Lord, 1952; Schoenfeldt, 1982; Thorndike, 1950) worked on different aspects of the problem, namely:

- the psychometric model of classification efficiency
- requirements of assessment instruments specifically designed for OPJM
- sampling and statistical considerations associated with measuring the benefits of OPJM
- development of assignment algorithms for fitting people to jobs
- research methods for measuring results

We provide an overview of personnel classification theory in the following section and a detailed description of the classification research paradigm in the Method Chapter.

The Benefits of Using the Personnel Classification Paradigm to Study ATIs

We believe that transporting the personnel classification paradigm to training will create important advances in ATI research for three reasons. First, the paradigm is based on a psychometric theorem developed by Hubert Brogden (1959) that delineates the mathematical basis for optimally matching people with treatments. Since Brogden's classification theorem is a general formula for characterizing any person-treatment interaction involving individual assessment measures and performance criteria for multiple treatments, we believe it will be as useful for measuring and interpreting ATI research findings as it is for person-job matching results.

Second, the classification paradigm is well-researched. It has been used to study empirical person-job matching questions since the 1960s (Zeidner & Johnson, 1994). More importantly, it provides a systematic approach for quantifying the practical effects of ATIs on student training performance. We have coined the term *mean predicted training performance* (*MPTP*) for the measure of person-treatment interaction in training settings. (The measure of benefit in the person-job matching context is referred to as *mean predicted* [job] *performance* [*MPP*]). MPTP is an estimate of the average training performance (across multiple course settings) produced by some method of placing students in training environments. Optimal assignment is the process of matching students to the settings that best match their aptitudes and learning strategies. The MPTP obtained from optimal matching should be compared to the MPTP obtained from other types of assignment processes (e.g., random or actual class assignments) to evaluate the potential practical improvements of optimal person-treatment matching compared to the other strategies.

Recent classification research has led to the development of a cross-validation procedure that has improved the accuracy of OPJM estimates and added a utility analysis capability that provides the opportunity to link performance benefits to dollar estimates of human resource costs (Nord & Schmitz, 1991). Both of these procedures can be transported to ATI research. We include cross-validation within the method we propose in this report. Further research will be needed to apply the OPJM utility analysis methods to training evaluation. The capability to

4

employ utility analysis to evaluate alternative technical course designs in terms of training dollars would be a major benefit to the Air Force.

Third, we believe that our adaptation of the personnel classification paradigm for ATI research will improve the detection of ATIs, if they are present. Moreover, we expect that the classification-ATI paradigm will provide a means for illuminating the causes of conflicting results that historically have been obtained with the traditional ATI research design. We modified the personnel classification method to produce a highly sensitive measure of ATIs using a twofold approach.

One, we designed an instrument we call the TCS, which measures specific learning context variables that we hypothesize will account for ATIs in alternative technical training settings. Two, we propose that multilevel regression (MLR) be employed to quantify and test the statistical significance of specific ATIs involving variables identified by the TCS. MLR requires the explicit formulation of interaction terms, and provides tests of their significance. Consequently, our method will identify which hypothesized ATIs are statistically significant and which are nonsignificant in predicting training performance. The TCS and MLR are described in detail in the Method Chapter of this document.

In conclusion, the direct parallel between the person-job interaction of classification and the aptitude-treatment interaction of training offers the opportunity to transport the classification paradigm, with modifications, to training evaluation research. Adapting a classification approach to the study of ATIs will move this area of research beyond the simple comparison of prediction functions across instructional methods. We believe that the classification-ATI paradigm can produce major advances in ATI research because it will improve the sensitivity with which ATIs are detected, if they are present, and shed new light on the exact nature of any ATIs detected.

A final advantage of the classification-ATI paradigm is that it will enable researchers to quantify the potential benefits of capitalizing on ATIs by simulating the optimal matching of students to training treatments. We anticipate that this quantification of the practical effects of ATIs will provide a basis for improving the effectiveness of training design. Snow and Lohman (1984) described the importance of ATI research to training evaluation as follows:

> Educational treatment comparisons, including program evaluations, must at least incorporate tests of plausible ATI hypotheses in order to interpret their intended main effect conclusions properly. Any treatment environment can serve some learners well and others poorly. Research on treatment design should thus always use what is known about individual differences to determine for whom any particular instructional method is appropriate and for whom it is not appropriate (pp. 358-359).

Personnel Classification Theory and Research

Personnel classification theory formally states the propositions underpinning OPJM and provides the backdrop for the methodology we propose in this report. The major premises are

that the nature of performance differs across occupations and that these differences interact with a worker's job-related characteristics to produce a range from low to high success in different occupations. Specifically, the theory holds that different occupations require different combinations of cognitive aptitudes, psychomotor abilities, personality characteristics, interests and other job-related variables (e.g., job knowledge). In turn, people vary in their patterns of these variables. Consequently, a person's success in a given occupation will depend upon the strength of the interaction (or match) of his or her profile on these variables with the occupational requirements for on-the-job performance (Statman, 1993).

As mentioned earlier, the significance of capitalizing on the interaction between individual aptitudes and interests and the differential performance requirements of occupations was recognized by Brogden (1946, 1951, 1954, 1955, 1959, 1964), Thorndike (1950), Horst (1954, 1956) and others (e.g., Lord, 1952) during and immediately after World War II. Brogden (1959) and Horst (1954) recognized that large organizations often face complex decisions in which personnel can be considered simultaneously for multiple treatments (e.g., career paths, jobs, training, and development opportunities). However, the person-job matching problem is usually simplified from a classification decision to a simple select/reject decision for a single treatment.

Brogden (1946, 1951, 1954, 1955, 1959) developed a mathematical model of differential classification between 1946 and 1959. This model, in greatly simplified form, became the basis for the Military Services' operational classification systems. However, little empirical research was conducted on Brogden's theorem after the 1960s. Researchers agree that this was due in large part to the complexity of the psychometric classification model and the person-job matching procedures that underlie classification decision-making processes (Hunter & Schmidt, 1982; Johnson & Zeidner, 1991; Zedeck & Cascio, 1984).

Recent advances in linear programming (LP) technology and in personal computer capacity led Johnson and Zeidner to revive the seminal work of Brogden (1959) and Horst (1954) in 1991. They proposed the first formally stated theory of classification efficiency called *differential assignment theory (DAT)*. In addition, they refined the research paradigm for studying classification efficiency through computer-based simulation of the person-job matching process, which had been developed in the 1960s.

We describe Brogden's classification theorem and DAT, and briefly review recent research in the next two sections of this chapter.

Brogden's Classification Model

Brogden (1959) proved algebraically that the gain in job performance from optimal matching of people to jobs compared to random assignment is a function of three variables: a) the predictive validity coefficients of the prediction equations for every job in the problem, b) a negative function of the intercorrelations of the equations, which is a measure of differential prediction efficiency, and c) the number of jobs (i.e., treatments) to which people are matched. His proof is based on several assumptions, including that the matching process is optimal (i.e.,

each person is assigned to the job for which he or she has the highest predicted performance score).

Brogden's (1959) measure of classification efficiency is the following:

$$MPP = R(1 - r)^{1/2}Z_m.$$

where:

MPP	=	the mean predicted performance standard score of a group of applicants optimally assigned to m jobs,
R	=	the average predictive validity of ordinary least squares (OLS) estimates for all jobs,
r	=	the average intercorrelation of the OLS estimates, and
Z_m	=	the mean criterion standard score of the group after assignment to m jobs with equal vacancies (called quotas).

This equation is fundamental to classification. It shows that classification efficiency is positively related to the predictive validity coefficients of the prediction equations for a set of jobs, and negatively related to the intercorrelations of the equations according to the function $(1 - r)^{1/2}$. This term, $(1 - r)^{1/2}$, is a measure of the effect of differential prediction across jobs on average job performance. Stated differently, it is a measure of the effect of person-treatment interactions on average performance across a range of occupations. Brogden's (1959) classification theorem is useful in constructing maximally efficient OPJM systems, because it instructs the researcher to maximize the predictive validities of the performance prediction equations, and to minimize their intercorrelations.

Although Brogden developed his classification theorem to estimate the benefits of OPJM systems, it applies to all person-treatment interaction situations in which one or more measures of individual characteristics are used to predict success in two or more treatments. Therefore, this theorem applies equally well to the study of ATIs in training. Further, the research paradigm that evolved from Brogden's theorem, which uses computer simulation to measure the benefits of OPJM, applies equally well to measuring the practical effects of capitalizing on ATIs to optimally match students to the best-fitting learning environment. As we discuss in the Method Chapter, we modified the classification paradigm for training settings to provide specific information on the nature and strength of hypothesized ATIs.

The most important term in Brogden's theorem is $(1 - r)^{1/2}$, the differential prediction function, because it measures the effect of person-treatment interactions on average performance when people are optimally matched to treatments. Understanding the differential prediction function allows the researcher to manipulate systematically the content of the assessment battery or the type of criterion variable in their investigation of ATIs.

The differential prediction term shows that (holding all else constant [e.g., the predictive validity of the equations and the matching process]) the strongest person-treatment effect is obtained when $r = 0.00$. In this case, the prediction equations are independent, meaning that a

completely different set of aptitudes, interests, etc., are required to perform successfully in each treatment. Conversely, there is no person-treatment interaction when the predictor composites completely overlap, producing $r = 1.00$. In this case, no benefit is achieved from OPJM, because a single set of measures predicts equally well for all treatments. This means that each individual performs equally well in all treatments.

Close examination of the differential prediction efficiency function highlights the interesting relationship between r and $(1 - r)^{1/2}$, and is useful in getting a rough estimate of the potential benefits from OPJM derived from different strengths of person-treatment interactions. As the average intercorrelation among the prediction equations increases in increments of .10 from $r = 0.00$ to $r = .99$, the loss in differential prediction efficiency occurs at a significantly slower rate than the pace at which the average intercorrelation increases.

Table 1 shows this effect. As stated above, when $r = 0.00$, the person-treatment interaction effect is at its strongest. When $r = .10$, the differential prediction effect is only reduced by 5%. When r increases to $r = .50$, the interaction effect is only reduced by 29% to .71. At the extreme point where the average intercorrelation of the prediction equations for a set of treatments is very high (e.g., $r = .99$) we still obtain a 10% person-treatment interaction effect.

Table 1. Comparison of r with $(1 - r)^{1/2}$

$r = 0.00$	$(1 - r)^{1/2} = 1.00$
$r = .10$	$(1 - r)^{1/2} = .95$
$r = .20$	$(1 - r)^{1/2} = .89$
$r = .30$	$(1 - r)^{1/2} = .84$
$r = .40$	$(1 - r)^{1/2} = .75$
$r = .50$	$(1 - r)^{1/2} = .71$
$r = .60$	$(1 - r)^{1/2} = .63$
$r = .70$	$(1 - r)^{1/2} = .56$
$r = .80$	$(1 - r)^{1/2} = .45$
$r = .90$	$(1 - r)^{1/2} = .32$
$r = .99$	$(1 - r)^{1/2} = .10$

Inspection of Table 1 demonstrates that an OPJM algorithm that assigns each person to the treatment for which he or she has the highest predicted performance score will capitalize on even small person-treatment interaction effects in the assignment process. Consequently, the OPJM process will result in a gain in average performance compared to random assignment even when only minor ATIs are present (e.g., when $r = .90$). Of course, the construction of any set of differential prediction equations must be based on large enough samples to insure that the differences in the predictor weights across equations are stable and valid.

The last term in Brogden's classification theorem, Z_m, is a measure of the effect of the number of treatments to which people are assigned. Z_m is an estimate of the mean actual performance of a group of applicants after assignment to m treatments (holding all else constant).

Brogden (1959) used an order statistic for Z_m to estimate the effect of the number of treatments without conducting a person-treatment matching simulation study. He showed that the gain from OPJM increases as the number of treatments increases. The effect of the number of treatments is independent of both the predictive validities and the intercorrelations of the differential prediction equations in Brogden's classification theorem.[1]

Brogden (1959) also showed that performance gains increase according to a decelerating function as the number of treatments (e.g., jobs or courses) is increased. However, valuable improvements in average performance can be obtained with only a few treatments depending upon the purpose of the person-treatment matching procedure and the strength of the ATIs. In fact, the decelerating function means that the largest percentage increases in performance are achieved with a small number of treatments.

The number of treatments is an important factor in designing an ATI study that employs the classification-ATI paradigm. We do not believe that it is necessary to have a large number of alternative settings for the classification-ATI paradigm to be useful in measuring ATIs in technical training and other learning settings. Although the aggregate benefit from assessing students for 10 or more learning environments would be greater than for 2, the decelerating function always reduces the marginal improvement in adding another treatment. It is up to the organization to evaluate whether having 2 or 3 alternative training settings (e.g., classroom, *computer-based training* [CBT], and distance learning) would be of practical value. This will depend upon a number of factors, the expense of recruiting personnel, the cost of training, the amount and cost of attrition or washback, and the consequences of poor training, to name a few.

The following is a brief overview of Johnson and Zeidner's (1991) classification theory, DAT, followed by a review of major recent research.

Differential Assignment Theory

Zeidner & Johnson (1994) and Johnson & Zeidner (1991) formulated a theory of classification efficiency called Differential Assignment Theory (DAT), which is largely based on Brogden's (1959) theorem for quantifying the benefits of OPJM, and on an index of differential prediction efficiency developed by Horst (1954). DAT describes the psychometric basis for using assessment batteries to optimally match people to jobs. We outline the basic tenets below because they may be useful in developing a theory of ATIs in learning. Further, Zeidner and Johnson's (1994) guidelines for creating OPJM procedures should be considered when designing ATI research and developing training applications that capitalize on ATIs operationally.

The basic propositions of DAT are that success in different occupations requires different sets of skills, abilities, interests, and other job-related variables (e.g., conscientiousness) and that

---

[1] This relationship does not hold in practice, although increasing the number of treatments has been found to have a relatively small effect on the other two variables (i.e., R and r) (Statman, 1993).

people vary in their profiles of these variables (Johnson & Zeidner, 1991; Zeidner & Johnson, 1994; Zeidner, Johnson, & Scholars, 1997). Thus, the theory holds that employing an OPJM strategy, which capitalizes on the stable variation in cognitive and non-cognitive predictors of performance, will improve average performance across all jobs, when compared to a simple selection strategy in which individuals are assessed for only a single occupational category.

Zeidner and Johnson (1994) developed a set of guidelines for designing OPJM procedures (Johnson & Zeidner, 1991). Three of the most important principles are the following:

(1) A classification battery must be multi-dimensional; i.e., it should measure a range of individual characteristics.

(2) Given adequate sample sizes, the highest level of classification efficiency will be obtained by computing OLS equations separately for each target job. This procedure maximizes classification efficiency because the OLS estimates have high (shrunken) predictive validity coefficients and low intercorrelations. Thus, Brogden's classification function ($R(1 - r)^{1/2}$) is maximized.

(3) Third, increasing the number of occupations (i.e., treatments) for which individuals are assessed will increase the benefits gained from OPJM at a decelerating rate, holding all else constant.

Summary of Recent Classification Research

The classification work of Johnson, Zeidner, and colleagues described below was directed toward validating Brogden's 1959 index of classification efficiency and identifying a set of principles to guide the development of OPJM batteries, treatment-specific prediction equations, and occupational groupings. The results support the validity of Brogden's classification measurement model, upon which our proposed classification-ATI paradigm is based. Further, most of the studies cited used a variant of the classification research design we propose in this report. The most important findings from these studies are the following:

(1) The relationships of R, r, and m to classification efficiency contained in Brogden's equation held up empirically (Johnson, Zeidner, & Leaman, 1992; Statman, 1993).

(2) Increasing the dimensionality of a mainly cognitive predictor battery (i.e., Armed Services Vocational Aptitude Battery [ASVAB]) by adding perceptual and psychomotor tests, a job-related personality measure, and an interest inventory produced a large increase in classification efficiency, although the improvement in predictive validity was modest (Statman, 1993).

(3) Multidimensional OLS prediction equations, which were computed for each job from a single battery, produced gains in average performance over both a general ability measure (weighted by predictive validity across jobs) and unit-weighted specific aptitude composites (Darby, Skinner & Alley, 1995; Johnson, Zeidner, & Leaman,

10

1992; Nord & Schmitz, 1991; Nord & White, 1988; Statman, 1993; Wetzel, 1990). As in (2) above, Statman (1993) obtained this finding despite that the average predictive validity of the OLS composites was not much greater than the validity coefficients of the other equations.

(4) Increasing the number of treatments to which assignments are made has a strong positive effect on classification efficiency that is independent of average predictive validity or differential prediction efficiency (Scholarios, Johnson, & Zeidner, 1994; Statman, 1993).

(5) The cross-validated estimates of average performance across treatments obtained in these studies, when compared to random assignment, showed gains ranging from about .10 to .50 standard deviation units.

Several other classification studies have been conducted using Air Force and Navy data. Alley and Teachout (1992) found that separate OLS equations of the 10 ASVAB tests predicting hands-on criterion measures resulted in an improvement in average performance over random assignment for eight Air Force jobs. Darby et al. (1995) obtained similar results with a criterion of final technical school grade in a larger study that included all Air force jobs. Siem and Alley (1997) found that an OPJM strategy, compared to random assignment, improved the predicted performance of Air Force pilots assigned to four different types of aircraft. Schmidt, Hunter, and Dunn (1987) conducted a study for the Navy in which they grouped ratings into three general job families. They found that a two-variable composite of general cognitive ability (g) and psychomotor ability produced greater classification efficiency than g alone.

Recently, Alley and Darby (1995) have used simulation techniques to expand Brogden's (1959) table of performance gains for alternative classification strategies from 10 to 500 jobs. In addition, they found and corrected a mistake in his theorem that improves the accuracy of the estimates. Alley, Darby, and Cheng (1996) expanded the Taylor-Russell tables to estimate the proportion of successful employees obtained through optimal selection *and classification* in the multiple job context as a function of base rate of success, selection ratio, predictive validity and number of jobs.

Sager, Peterson, Oppler, and Rosse (1997) compared indices of selection efficiency, classification efficiency, and differences in subgroup means for all possible combinations of ASVAB tests and the experimental predictors included in the Enhanced Computer Administrated Test (ECAT) battery (Wolfe, 1997). They found that no one battery of tests simultaneously optimized all indices. Consequently, they concluded that when determining the content of an assessment battery, researchers must consider the purpose (i.e., selection, OPJM, or to increase minority or gender representation in an organization or occupation) for which it will be used. Further, researchers must be prepared to make tradeoffs among the alternative types of outcomes they desire when designing the battery.

The potential practical utility of selection and classification strategies, measured in dollars instead of performance, has received modest attention. Nord and White (1988) and Nord

and Schmitz (1991) developed several approaches to classification utility analysis and found significant savings associated with increments in mean performance due to OPJM. Harris, McCloy, Dempsey, DiFazio, and Hogan (1993) developed a Cost-Performance Tradeoff Model (CPTM) based on an OPJM simulation that provided dollar estimates of utility. The CPTM approach employed a cost-effectiveness index of classification efficiency with a number of operational constraints built into the OPJM process. The objective of the model was to minimize recruiting, training, and compensation costs through an OPJM strategy that met minimum performance standards in all jobs. Harris et al. found that increasing the number of dimensions in a test battery minimized costs. Further, different combinations of tests affected the recruiting and training/compensation costs in different ways. Statman, Harris, McCloy, and Hogan (1994) compared the Harris et al. cost-effectiveness OPJM strategy to the Brogden-Johnson-Zeidner approach of maximizing average performance and obtained generally the same results using both models.

In summary, differential classification theory has a sound psychometric basis in Brogden's (1959) classification theorem and has received a good deal of research in recent years due to improvements in LP and personal computer technology. The refinements in the research method developed by Johnson and Zeidner (1991) have produced a strong body of results that support the existence of ATIs in the person-job matching domain. Further, the classification research paradigm effectively captures the practical effects of job-related ATIs in terms of both performance and personnel costs.

The final section of the Introduction presents an overview of the ATI literature and a review of training research. We emphasized the training literature in this project because we felt it was critically important to use current research findings to guide our development of the TCS.

Brief Overview of ATI Research

The research findings on ATIs are quite mixed. Numerous studies have found that aspects of the training environment interact with learner characteristics to influence training performance outcomes, e.g., the instructional method (Cronbach & Snow, 1977), teaching strategies (Snow & Lohman, 1984), and course content (Mumford, Weeks, Harding, & Fleishman, 1988). However, large numbers of studies have found no statistically significant ATI effects (Maldegen et al., 1996). These apparently conflicting results make interpretation of the ATI literature difficult, especially because most studies relied on small samples and investigated unique treatment variables. Further, Maldegen et al. (1996) found very little replication of research.

Although an extensive review of the ATI literature was beyond the scope of the current project, we noted in our limited review process that the research as a whole lacked carefully designed methods and controls (Maldegen et al., 1996). Some of the studies that reported little evidence for ATI effects involved ATI analyses that were not planned; consequently, the study designs did not include control conditions or variables consistent with sound research design (Goldstein, 1993).

12

Campbell (1988) observed that we have only "scratched the surface" of ATI research. He suggested that our understanding of both individual differences and the relevant features of the training environment should have more elaboration. In the domain of learner characteristics, he stated that we must clarify the independent effects of cognitive abilities and prior achievement or experience on training performance, and the interactions of these variables with training content. In the domain of the training environment, Campbell stated that complexity of instructional method (which interacts with general ability) is confounded with training content. In other words, highly complex and unstructured training programs tend to reflect highly difficult content; while structured, less complex courses contain less difficult material. The implication of this observation for the study of ATIs is that analysis of the training environment should include independent measurement of instructional method *and* course content. As we describe below, we developed the TCS as an instrument to measure each of these training variables separately..

We believe that better designed research is needed to identify the person and training variables with the strongest interaction effects on training success, and to improve methods of quantifying their impact. Although better understanding and measurement of ATIs will improve the effectiveness of all types of training, the greatest gains may be made in adaptive training systems.

Adaptive training systems consist of a number of different paradigms that embody different teaching strategies (e. g., exploration vs. coaching). The goal of adaptive training is to use student abilities and knowledge gained within lessons to diagnose student learning needs and develop individualized instructional strategies that help students learn (Sleeman & Brown, 1982). Improvements in training achievement and reductions in learning time have been reported when adaptive training systems were compared to conventional methods of instruction (e.g., classroom, self-study, on-the-job training) or control groups.

In her meta-evaluation of four intelligent tutoring systems, Shute (1991) identified several learner characteristics that were related to performance on computer-based tutors: acquisition and retention were related to LISP performance; scientific inquiry skills were related to performance in microeconomics delivered by an intelligent tutor; working memory, two problem-solving abilities, and learning style were related to performance on a PASCAL intelligent tutor.

These findings suggest that the interaction of learner characteristics with instructional method (and probably course content) partially determine training outcomes in an adaptive training environment. Therefore, evaluation of adaptive training systems must address ATIs in order to improve our understanding of training performance, and to determine the most efficient applications of adaptive training technology.

Further, study of the interaction between learner characteristics and intelligent tutors will contribute to the future development of adaptive training systems and other methods of instruction. Baker and O'Neil (1986) noted that understanding the relationships between abilities and instructional options is relevant for the analysis and implementation of alternative student

models and tutoring strategies. They said that the interaction of intelligent tutors and cognitive style (e.g., the need for structure, the need for reflection, and the attribution of success and failure) is also important for the design and evaluation of adaptive training systems.

In summary, the ATI literature contains many conflicting results, little replication, and some studies with poor research designs (Maldegen et al., 1996). Campbell (1988) and others (Snow & Lohman, 1984) have long called for improvements in ATI research as a strategy for improving training design and evaluation. Any improvements achieved could have wide-ranging effects across the spectrum of instructional methods, but especially in the design of adaptive tutors, because they capitalize on ATIs as a teaching strategy.

In the next section we present our review of the training literature. We focused on identifying situational variables that may interact with student learning characteristics. The purpose of our review was to guide the development of the TCS, which we designed to measure the major training variables we identified through the review process. The following discussion provides an indication of the strengths and weaknesses of the training literature and the background for our choices of variables to include in the TCS.

Review of the Training Literature

Purpose of the Review

We conducted a review of several bodies of literature, including those of technical training, human factors, industrial and organizational psychology, educational psychology, and instructional design, as the preliminary phase in designing the classification-ATI research method and developing the TCS. We considered this review to be essential because it provided us with research-based guidance for identifying the specific characteristics of technical training environments that may interact with learner characteristics. As we describe in the Method Chapter, our proposed approach involves using the TCS and MLR to identify and quantify ATIs related to specific training variables. We believe that this strategy of elucidating the interactions of learner characteristics with a number of training variables will help to reconcile the conflicting findings of previous ATI studies, most of which did not carefully control the training settings or include quantitative measures of ATIs.

The TCS is contained in the Appendix and described below in the Method. It was designed to measure the aspects of entry-level Air Force technical training courses that might interact with learner characteristics to produce intra-individual differences in training performance in alternative learning environments. In designing the TCS we made the assumption that Air Force researchers studying ATIs in the near future probably would have access to only the individual difference variables as measured by the ASVAB. This is because data on other types of variables (e.g., motivation, interests, learning styles, self-efficacy, work values) were not available during TCS development, and no large-scale data collections outside of the cognitive domain were planned at that time.

However, this situation has since changed. As this report was being finalized, the Air Force began collecting data on a non-cognitive predictor of attrition called the *Assessment of Individual Motivation* (*AIM*). The AIM is a self-report measure of psychological temperament and motivation developed by the Army Research Institute (Young & White, 1998). It was based on an earlier Army instrument called the *Assessment of Background and Life Experiences* (*ABLE*), but is believed to be an improvement because it uses a forced-choice format to control for socially desirable response distortion and susceptibility to coaching. The AIM contains six scales that measure dependability, work orientation, adjustment, physical condition, dominance, and agreeableness. Since the data had not been analyzed before this report went to press, we do not have results that would provide us with any indication of the AIM's usefulness for detecting ATIs in Air Force technical training. However, we suspect that several of the scales (especially the first three) might be good predictors of training motivation, and may interact with instructional setting.

While we conducted a broad review of the training literature, our emphasis was mainly on the aspects of training that we believed would interact with the cognitive aptitudes and job-related, technical interests measured by the ASVAB. Although we concentrated less on how training environments interact with other student characteristics (e.g., motivation and learning styles) not measured by the ASVAB, we would like to see future ATI research based on the classification-ATI paradigm include more than cognitive and military interest variables.

Our reasoning stems from the differential prediction efficiency term in Brogden's 1959 classification efficiency theorem. Recall that this term indicates that optimal person-treatment matching is strongly influenced by the amount of differentiation in a set of equations created to predict performance across alternative treatments (whether jobs or courses). The greater the dimensionality of the battery (i.e., the more different types of variables measured), the greater the opportunity for differential prediction efficiency across treatments.

Among possible candidate variables for future ATI research, we recommend self-efficacy, career identity, learning style, cognitive style, and the various measures of motivation included in the AIM, to name a few. If the AIM or other instruments were to be used in a classification-ATI study, then the TCS should be expanded to include training variables that might interact with those measures, e.g., lateness records and participation in study groups.

Cannon-Bowers, Tannenbaum, Salas, and Converse (1991) note that "reviews of the training literature over the past 20 years have painted an increasingly optimistic picture of the field" (p. 281). They quote John Campbell as stating more than 25 years ago that the training and development literature was "nonempirical, nontheoretical." While more recent reviews indicate that much work has been accomplished in integrating theory with training applications, Cannon-Bowers and colleagues note that there is still a gap between what training practitioners do and what training theory suggests. To fill the gap they propose a framework to link training-related theory and techniques. Their framework is based on three questions relevant to training research:

- What should be trained?
- How should training be designed?
- Is training effective, and if so, why?

They state:

> Overall, the framework suggests that research can be conducted in both training theory and training techniques, so that (1) theoretical findings can be translated into specific training techniques, and (2) the study of techniques can help to confirm/refine/expand related theory (Cannon-Bowers et al., 1991, p. 284).

The Cannon-Bowers et al. (1991) framework illustrated the importance of examining literature related to both training theory and practice. Still, we found the literature to be lacking. In general, we found that the training literature contained either narrowly focused studies which were designed to examine a single, specific training variable (e.g., Bacdayan, 1994), or broad-based approaches to training that attempted to organize research methods and results (e.g., Ryder & Redding, 1993). The Mumford, Weeks, Harding, and Fleishman (1988) study is an exception to this generalization. It was very comprehensive and provided a great deal of detailed information for the design of the TCS.

The following is a description of the training variables we identified as candidates for producing large interactions with student characteristics. The Mumford et al. (1988) study included a thorough examination of course content variables. Most of the variables identified in other studies could be categorized as aspects of method of instruction. However, a small number were difficult to categorize. Our discussion is organized around course content variables, variables related to method of instruction, and a miscellaneous category that includes variables related to course content and skill acquisition.

Course Content Variables

Mumford et al., (1988) conducted a comprehensive study of student and course variables related to technical training performance for the Air Force. They collected 6 measures of student characteristics, 16 measures of course content, and 7 measures of training performance. These variables cover a much greater range of the training environment than most studies. They are important descriptors of the Air Force training process (see Table 2). Most measures were readily available from programs of instruction and administrative records. They did not have access to student characteristics (e.g., learning style, preferred learning strategies, and interest) nor did they have measures of teaching style or motivational techniques.

Using measures of the student, course, and outcome variables from Air Force trainees in 39 entry-level technical training courses,[2] Mumford et al. (1988) were able to develop a

---

[2] The 39 training courses examined by Mumford et al. (1988) appear to be primarily lecture-based classroom instruction.

hypothetical model of the relationships among these variables. They found three primary course content factors: subject matter difficulty, occupational difficulty, and manpower requirements.[3]

Table 2. Student Characteristics, Course Content, and Training Performance Variables

Student Characteristics	Course Content	Training Performance
Aptitude Reading level Academic achievement motivation Educational level Educational preparation Age	Course length Diversity Practice Abstract knowledge requirements Reading difficulty Programmed attrition Student-faculty ratio Instructor experience Instructional quality Instructional aids Hands-on practice Feedback Yearly flow Manpower requirements Day length Occupational difficulty	Assessed quality of performance Special individualized assistance Academic counseling Nonacademic counseling Washback time Academic attrition Nonacademic attrition

The primary course content variables had a stronger impact on training performance than did other course content variables (e.g., course length, feedback, student-faculty ratio, hands-on practice). However, Mumford et al. suggest that the other course content variables may exert a greater effect on performance than they observed in their study when the other and primary course content variables are not consistent. For example, a course with difficult material is usually long or provides much feedback to students. When a difficult course is short, then length is expected to play a larger role in training outcome than when the course is long.

The authors (Mumford et al., 1988) concluded that the Air Force training process is complex and multivariate in nature and that "optimal prediction and sound understanding of training performance will be obtained only when both student characteristics and course content are considered" (p. 455). Their results indicated that training performance is a function of a large set of variables and no single variable will fully explain training outcomes. Their results also suggested that weak findings in previous research may reflect a limited focus on the setting

---

[3] Subject matter difficulty is measured by abstract knowledge requirements, programmed attrition, reading difficulty, and diversity. Occupational difficulty is measured directly by an occupational difficulty variable consisting of "aggregate evaluations of entry-level task-learning time weighted by the percentage of total time spent in task performance among individuals entering an occupational field" (Mumford et al., 1988, p. 447). Manpower requirements are measured by yearly flow and manpower requirements.

for learning (e.g., lecture course vs. CBT), rather than on variables that "condition the nature of the learning process" (e.g., subject matter difficulty).

Training Variables Related to Instructional Strategies

We define instructional strategy broadly as the manner in which material is presented and learned, and the medium of instruction used. The instructional strategy for a particular course consists of a large number of variables that characterize the learning situation, including teaching method; medium of instruction; role of the learner (i.e., active or passive); class size; type and amount of structure; amount and frequency of feedback to students; and control and flexibility of course content, sequence and pace. The distinction between teaching method and medium of instruction is often blurred, although a given instructional method may be used with a variety of media. For example, a human instructor or a computer may provide tutoring. Numerous instructional strategies are used in technical training and are referred to by their most salient characteristic—lecture, hands-on training, adaptive training, and distance learning (see Kearsley, 1977; Reynolds & Anderson, 1992; Thompson, Simonson, & Hargrave, 1992).

Our description of characteristics of instructional strategies relevant to developing the TCS is organized into studies that examine the effects on student performance of training methods and medium of instruction; class size; amount of course structure; feedback to students; and control of course content, sequence and pace.

Training methods and medium of instruction. Shute (1991) found that students trained with an intelligent tutoring system learned faster and performed at least as well as or better than students in traditional training programs (e.g., human tutoring, classroom training, on-the-job training). Kozlowski (1995) compared mastery training to performance goal training. In mastery training, the "emphasis is on acquiring essential knowledge and skills, instead of achieving success and errorless performance" (Kozlowski, 1995, p. 8). Performance goal training, on the other hand, is characterized by the reinforcement of correct, errorless performance that promotes "short-term and surface processing strategies, such as memorization and rehearsal" (Kozlowski, 1995, p. 8).

Controlling for ability and learning orientation preferences, mastery training led to faster learning of basic task knowledge than performance goal training. Also, mastery trainees showed improved development of meta-cognitive structure (i.e., comprehension of concepts, strategies linked to concepts, etc.); performance goal trainees showed little improvement. While performance goal trainees performed better than the mastery trainees during the training trials, they were not as successful as the mastery trainees were in adapting to the novel task.

Although comparisons of one training strategy against another provide important information about the effects of the training environment on learning, they do not provide a complete picture of the relationships between student, training, and outcome. Consider how interactions between student characteristics and the training environment may affect the results of comparison studies. McCombs and McDaniel (1981) and Savage, Williges, and Williges (1982) demonstrated the impact of individual differences on training performance. Students

adaptively assigned to instructional modules (i.e., assigned to modules based on prior knowledge and learning style to maximize match between student and instructional module) completed lessons an average of 6.9 percent faster and received lesson scores an average of 2.1 percent higher than students randomly assigned to modules (McCombs & McDaniel, 1981). Savage et al. (1982) used motor and information processing tests to match individual characteristics and training type. Using adaptive training with fixed difficulty, Savage et al. found that matched students completed training 47 percent faster than randomly assigned students and 53 percent faster than mismatched students.

Class size. Several researchers have studied the effects of class size on training performance for different types of learning. Smith, Neisworth, and Greer (1978) found that student participation is directly related to group size. Peterson and Janicki (1979) found that there is an interaction between class size and ability in retention of mathematics instruction at the elementary school level. High-ability elementary school children retained more mathematics instruction when taught in small groups, while their low-ability counterparts retained more when learning in a large-group setting (Peterson & Janicki, 1979).

Shute, Lajoie, & Gluck (in press) suggest that class size should differ as a function of the type of task being learned. For example, performance-based tasks, such as flying an airplane, require individualized practice on component skills. Knowledge-rich tasks, such as troubleshooting or diagnosis, "tend to require associative learning skills and elaborative processing, and are typically well-suited to small-group instruction" (Shute et al., in press, p. 36).

Kramer and Korn (1996) suggest groups of four to nine students for class discussion. Shute et al. (in press) state that the optimal size of groups for collaborative and cooperative small group learning environments is two–three individuals.

Amount of course structure. A learning environment high in structure tends to be teacher-centered, uses pre-organized material, and includes very specific instructions and expectations (e.g., math classes) (Hunt, 1979). The need for structure is considered to be a learning style. Not only do students vary in their need for structure, but different subjects or disciplines tend to vary in their amount of structure. For example, mathematics tends to be more structured than the social sciences. There is a tendency for students with structured learning styles to perform better in engineering and math (structured subjects) and for students with less need for structure to perform better in social sciences (less structured subjects). However, the types of tests used with these subjects may confound this finding. Math tests tend to favor structure while social science tests tend to favor less structure (Hunt, 1979).

Snow and Lohman (1984) concluded that there is evidence of a significant interaction between general academic ability and the degree of structure in a learning environment. "[M]easures of intelligence ... correlate more highly with learning when instruction is incomplete, complex, and relatively unstructured, and less highly as instruction is more complete, carefully structured, and controlled by teachers" (Snow & Lohman, 1984, p. 118).

There is also evidence that there are interactions between structure and preference for type of structure and between structure and student anxiety. Students in a college-level psychology course who reported a high preference for structure but were placed in a class low in structure scored lower than students who were placed in classes matching their preference for structure or those with a low preference for structure who were placed in high structure classes (Shaw & Bunt, 1979). De Leeuw (1983) found that more global teaching methods, characterized by less structure and larger steps, were beneficial for less anxious students, while analytic methods, including more structure and smaller instructional steps, were beneficial for more anxious students. Similarly, there was a significant interaction of software self efficacy and type of instruction with managers and administrators learning to use computer software with either video-modeling training or a one-on-one interactive tutorial on diskette. All trainees performed similarly in the video-modeling condition, but the low computer efficacy group scored significantly lower than the others in the tutorial condition (Gist, Schwoerer, & Rosen, 1989).

Leeds On-Line Advisor (LOLA) is an example of a computer-based educational advisory system that provides advice to students who are learning on their own (Arshad & Kelleher, 1990). LOLA is designed according to the notion that students who have been in teacher-centered learning environments may have some adjustment problems in higher-level education where there is less support and more choices. It advises students what to study (content, curriculum), how to study (methods, strategies), and when to study (schedule). Essentially, LOLA provides structure for the student who is learning on his or her own. LOLA incorporates five different methods—exposition, consolidation, remediation, test-diagnosis, and introduction. LOLA provides structure by suggesting one of the five methods based on the student's previous responses (Arshad & Kelleher, 1990).

Feedback. Feedback is one of three fundamental factors that Taylor (1987) describes for selecting effective courseware. It may be informative or motivational. If a course provides feedback to the students, the feedback should be appropriate. It should be in line with the objectives of the course and the objectives and needs of the students taking the course.

Knowledge-of-results feedback provides both motivation and guidance that enhance performance (Mark & Greer, 1995; Salmoni, Schmidt, & Walter, 1984). Trainees prefer immediate feedback (Reid & Parsons, 1996). However, while immediate feedback aids initial task performance, slight delays in feedback (i.e., 10-30 seconds) or other disruptions in initial learning may actually benefit transfer of training (Schroth, 1995). Also, feedback that is too frequent can interfere with the learning process and degrade performance (Salmoni et al., 1984).

Brophy (1986) presents the idea of feedback intensity in the classroom: specific, immediate feedback at each stage of teacher-student interaction. For example, the instructor first presents information to the class; students then receive feedback as they discuss, answer and ask questions; the teacher then assigns practice exercises; and, finally, students receive additional feedback as the teacher monitors their individual work.

While a complete review of the relationship of goal setting and feedback to training is not within the scope of this review, we briefly mention how feedback and goal setting work together

in the context of training design. Feedback on the extent of goal achievement is necessary, but not sufficient, for goal setting to have an effect. Hence, the pairing of feedback with specific and challenging, but attainable, goals is an important component of good training design (Goldstein, 1993).

Student control of course content, sequence and pace. The amount of control that a student has over the content, sequence, and pace of instructional material can vary from course to course. Taylor (1987) includes learner control as an important factor in the evaluation of courseware. Content control includes selection of the curriculum, objectives, and lessons. Control of learning strategy includes selection of the number of examples, practice exercises, and level of elaboration (Taylor, 1987).

Thompson et al. (1992) stated that there may be optimal levels of learner control that should not be exceeded. They cite two studies that support this view. First, Tennyson (as cited in Thompson et al., 1992) demonstrated that adaptive programs are superior to programs that give the learner total control. Second, Allred & Lotactis (as cited in Thompson et al., 1992) found that although learner control may facilitate intrinsic motivation, learning outcomes may suffer.

Kearsley and Hillelsohn (1982) report that high achievers or extremely goal-oriented students complete self-paced training programs faster than traditional training programs with their lock-step sequence and pace. Additionally, they report that distributed practice leads to better retention than massed practice, particularly for lower aptitude trainees.

Other Training Variables Related to Course Content and Skill Acquisition

This section includes studies that were difficult to categorize, but which addressed a number of variables related to course content and their potential for interacting with student characteristics to produce different learning outcomes either at different points in a course or in different training settings.

Relationship of course content and training techniques to cognitive demands and skill acquisition. Schneider (1985) defines high-performance skills as those where the training requires trainees to expend considerable time and effort to acquire the skill, a substantial number of motivated individuals will fail the training, and there are substantial qualitative differences in performance between novices and experts. In high-performance skills, performance changes qualitatively over time, therefore training techniques compatible with initial skill acquisition may not be effective during later stages of skill learning. Schneider's work with high-performance skill training prompts the question: Which training techniques are best at different stages of skill acquisition?

According to Anderson (1985), there is a three-phase sequence in skill acquisition: acquisition of declarative knowledge, knowledge compilation, and acquisition of procedural knowledge. In phase one, general intelligence is required. In phase two, perceptual speed is tapped. And in phase three, psychomotor abilities are needed. In the initial stage of skill

acquisition, learning the steps to perform difficult, novel, or complex tasks places high demands on cognitive resources. That is, the individual's cognitive workload is high and he or she cannot process additional information or do additional tasks. Therefore, skill acquisition is a sequential, and not a simultaneous, process. Ackerman, Sternberg and Glaser's (1989) three stages of practice—cognitive, associative, and autonomous—mirror Anderson's phases of skill acquisition. They note that learning or training a novel task requires basic content knowledge and cognitive ability. After some practice, applying the content knowledge requires perceptual speed. Finally, after sufficient practice, psychomotor abilities are needed for expert performance.

Similarly, Kraiger, Ford, and Salas (1993) identified three general categories of cognitive measures used in training evaluation—verbal knowledge, knowledge organization, and cognitive strategies—which are sequential in the sense of skill training and acquisition. Verbal knowledge is taught and learned first, and is needed to move into the knowledge organization stage. The basic subject material must be learned and organized before cognitive strategies are applicable.

In summary, the work of Schneider (1985), Anderson (1985), and others on skill acquisition and training stimulates questions about the relationship of training techniques to learning and the types of techniques which maximize learning at different stages of skill acquisition.

Ryder and Redding (1993) created an Integrated Task Analysis Model (ITAM) as a framework for integrating cognitive and behavioral task analysis methods in the design and development of training using alternative approaches like instructional systems design (ISD). The ITAM skill taxonomy considers a large number of variables, for example, "demands on working memory, knowledge requirements (long-term memory), internal code (verbal or spatial), stimulus complexity and predictability, and overall mental workload" (p. 84). The memory requirements for different types of training are important considerations in the ITAM. Memorization ability can be an important prerequisite for training.

A completely different aspect of training concerns tests, which are not usually thought of as part of the course content. However, tests have been shown to influence teacher and student performance (Frederiksen, 1984). Frederiksen suggests that different types of test items require different cognitive processes. Therefore, the type of test used in a class may influence teaching and learning strategies beyond merely teaching to and studying for a test. For example, a course that includes tests which ask the students to apply a principle will generally include both learning the principles, and teaching and practice of the application of principles. Further, test questions that prompt students to apply a principle generally require more thorough cognitive processing than items that require them to recall a principle.

Goals and learning styles. Different learning (and training) strategies may be optimal for different training goals (Donchin, 1989) and different course content (Sein & Bostrom, 1989). Abstract learners performed significantly better than concrete learners on transfer tasks while learning an electronic mail system (Sein & Bostrom, 1989). According to Kanfer and Ackerman (1989), ability plays a role in how learning/teaching strategies are used by trainees. High ability students were more able to disregard the use of non-optimal learning/teaching strategies than

22

were low-ability students. In addition, task complexity interacts with the relationship between training goals and performance. For example, goal setting affects performance on simple tasks more than on complex tasks (Kanfer & Ackerman, 1989).

Instructional method and learning styles. Gregorc (1979) suggests that different types of instruction should be used for students with different learning styles. Student learning styles are defined as:

> characteristic cognitive, affective, and physiological behaviors that serve as relatively stable indicators of how learners perceive, interact with, and respond to the learning environment.... Styles are hypothetical constructs.... They are persistent qualities in the behavior of individual learners regardless of the teaching methods or content experienced (Keefe, 1979, p. 4).

According to Gregorc (1979), there are four learning patterns: concrete sequential, concrete random, abstract sequential, and abstract random. Training characterized by direct, hands-on experience, with step-by-step directions and clearly ordered presentations of material, is best suited for concrete sequential learners. Trial and error instruction and independent or small group training is characteristic of the concrete random style. Training emphasizing written, verbal, and symbolic tasks, and presentations with substance is most effective for abstract sequential learners. Holistic, unstructured multisensory training is most successful with abstract random learners.

Table 3 presents the type of course materials and teaching strategies suggested by Gregorc (1979) for each of his four learning styles. Students tend to prefer training that reflects their favored learning method (i.e., lecture, demonstration, discussion, film, print, etc.) (Dixon, 1982). However, research on whether matching training techniques to learner preferences increases the amount learned has led to equivocal results. As previously mentioned, Allred & Lotactis (as cited in Thompson et al., 1992) found that although giving the learner control may increase intrinsic motivation, learning outcomes may suffer.

Table 3. Types of Instruction by Learning Types

Learning Types	Types of Instruction
Concrete sequential	workbooks, manuals, demonstration, programmed instruction, hands-on, field trips
Abstract random	movies, group discussion, short lectures with question and answer and discussion, television
Abstract sequential	extensive reading assignments, substantive lectures, audio tapes, analytical "think-sessions"
Concrete random	games and simulations, independent study projects, problem-solving activities, optional reading assignments

Motivational strategies. Smith-Jentsch, Jentsch, Payne, and Salas (1996) suggest that pretraining experiences can influence posttraining performance by increasing students' motivation to learn. They found a positive relationship between pretraining motivation to learn and gains due to training.

In summary, several sets of situational training variables were identified as being important contributors to success in technical training and other learning contexts. However, the question of the existence and importance of ATIs is still in doubt. We believe that this is at least in part due to the following: the large number of variables that have been studied; the lack of replication of methods and studies; and limitations in the designs of many studies (e.g., small sample sizes). In the Method Chapter, which follows, we present a proposed strategy for improved precision in measuring ATIs and assessing multiple ATIs in a single study.

Chapter 2: Method

Adaptation of the Personnel Classification Paradigm for Studying Aptitude-Treatment Interactions (ATIs)

Overview of the Method

We present in this Chapter an approach to the study of ATIs that employs a person-treatment matching research paradigm taken from personnel classification, and that quantifies ATIs in a new way. The method will produce a measure of ATIs that accounts for their practical effects on learning achievement, and, with further development, can be linked to training budgets.

A key component of our method is the Training Characteristics Survey (TCS). It is a structured questionnaire that asks training subject matter experts (SME) to quantify the aspects of courses that we hypothesize will interact with learner characteristics to produce intra-individual variation in training achievement in different course settings (e.g., classroom/lecture, distance learning, computer-based training [CBT], adaptive tutors). The TCS is presented in the Appendix and described in detail below.

The TCS data are entered into a multilevel regression (MLR) procedure that uses them to construct course-specific prediction equations. We considered MLR a useful technique for studying ATIs because it allows a researcher to compute a separate ATI term for each predictor-training -variable combination in a study. MLR also enables the researcher to identify the statistical significance and strength of interactions involving specific training variables. In contrast, the traditional ATI research method only permits the identification of global ATIs.

The most important difference between the classification-ATI and traditional research paradigms is that the former uses optimal person-treatment matching software to assign students to courses. In contrast, students are randomly assigned to treatments in traditional ATI research. The matching software allows the researcher to simulate the benefits that could be obtained in real settings if ATIs were used to match each student to the most effective learning setting for him or her. If strong ATIs were identified by the MLR procedure, then a large gain in average performance across settings would be obtained from optimal matching in comparison to random assignment. If weak or no ATIs were found, then optimal and random assignment would produce equivalent levels of average performance.

The description of the classification-ATI method is divided into the following sections:

- development of the TCS
- estimation of prediction equations: MLR analysis
- selection of courses for a classification-ATI study
- selection of criterion variables
- selection of predictors
- simulation of the student-course matching process

25

Development of the TCS

The purpose of the TCS is to obtain quantitative ratings of training characteristics for entry-level technical training courses in the Air Force. It will allow a researcher to describe training in terms of those aspects that differentiate one course or course setting from another ,just as personal characteristics can be described with measures of individual differences. When used in conjunction with individual differences variables (e.g., the tests of the Armed Services Vocational Aptitude Battery [ASVAB]), the TCS will provide the data to identify specific learner-training-variable interactions in Air Force technical training. However, the TCS can be adapted easily for other types of training settings and evaluation research, because it was designed to measure major situational variables.

We envisioned that training SMEs, who could include Air Force instructional system designers, course managers, and instructors, would complete the TCS. Ideally, each course selected for study would have approximately 10 independent SME ratings. If 10 SMEs are not available, then the instrument, with modifications, could be administered to students. Students would bring a different perspective to the training variable ratings, but they would not have the instructional knowledge of the SMEs. Thus, we would expect student ratings to provide less and somewhat different information than the SME ratings. If the TCS were administered to students, we would recommend having approximately 20 to 30 student respondents per course. Separate and combined analyses of SME and student responses would be necessary.

We conducted several internal reviews of the TCS with training research and development experts to refine the instrument. We also conducted an external review with Air Force research psychologists. The instructions and a number of items were clarified as a result of the reviews. Before administering the survey, we recommend that a pilot test be conducted with a sample of potential respondents.

Training variables included in the TCS. The TCS contains five sections:

- Background Information
- Occupational Area
- Method of Instruction
- Course Difficulty
- Course Content

Multiple items were included in all sections except Occupational Area, which asked for the Air Force Specialty Codes (AFSCs) associated with the course being surveyed. We varied the types of items and response formats, and designed items to tap important aspects of the general areas with which they were associated. Some items (e.g., reading grade level) probably could be obtained more efficiently from training materials (e.g., the program of instruction) instead of from SMEs. If this is the case, we recommend that the researcher obtain all information possible

from existing Air Force materials and data bases. This would result in reduction of the size of the TCS and a concomitant reduction in survey time.

As noted in the review of training literature above, we used the information we gleaned from it as the main guide for our TCS development process. However, we focused on training characteristics we believed would be well matched to individual characteristics measured by the ASVAB. For example, we considered mechanical ability and electronics knowledge, but not motivation or impulsivity, which are not measured by the ASVAB, and for which the Air Force currently does not have available instruments or data. We took this approach because the general view among Air Force researchers at the time we were developing the TCS was that ASVAB tests would be the only individual difference variables available for large samples of recruits in the near future.

However, his situation changed unexpectedly late in the project, when a large data collection was begun on work motivation variables captured by the Assessment of Individual Motivation (AIM). Refer to the section above entitled *Review of the Training Literature* for a description of the AIM, and to the section below entitled *Selection of Predictors* for mention of a cognitive information processing battery, the Advanced Personnel Testing (APT) battery, which also may be appropriate to include in a future classification-ATI study.

In general, our item development process revolved around the major sections of the survey, course content and difficulty, and method of instruction. The work of Mumford et al. (1988) provided the basis for the course content section because they identified and carefully analyzed 16 variables. The remaining training studies provided a range of variables that we used for the method of instruction and course difficulty sections. The training characteristics included pace of the class (see Kearsley & Hillelsohn, 1982), sequence of the instruction (see Taylor, 1987), flexibility to change the pace or sequence (see Allred & Lotactis, as cited in Thompson et al., 1992), instructional methods (see Dixon, 1982; Kearsley, 1977; Thompson et al., 1992), and level of abstraction of course concepts (see Gregorc, 1979). Several variables such as pace and structure, seemed to belong in two categories, so we placed items in both sections when appropriate.

Additionally, research on individual differences was reviewed and considered from a training perspective to fill some of the gaps we found in the training literature. Several cognitive abilities defined by Fleishman and Reilly (1992), including written comprehension, mathematical reasoning, inductive reasoning, and perceptual speed, served as stimuli for developing corresponding training variables for the course content section. We also used the learning styles literature, which included variables such as need for structure (see De Leeuw, 1983; Hunt, 1979; Snow & Lohman, 1984), to suggest several items.

Analysis of the TCS. We recommend that the TCS items be subjected to principal components analysis with varimax rotation to identify the underlying dimensions of variability in

training environments.[4] After varimax rotation to simple structure, we suggest that the first several factors, which account for the greatest proportion of variance and make conceptual sense, be selected. The training factors taken from the TCS data would serve as course-specific variables and would be entered into the MLR procedure described below to produce a set of differential course prediction equations that reflect ATIs, if they are present.

Based on previous findings with job analysis data and MLR in personnel classification research (Harris, McCloy, Dempsey, Roth, Sackett, Hedges, Smith, & Hogan ,1991; Harris et al., 1993), we would expect to find that three–five factors will describe the training environment adequately. The knowledge we gleaned from the training literature leads us to anticipate that the factors would reflect aspects of the method of instruction, course content, and job (see, for example, McCombs & McDaniel, 1981; Mumford et al., 1988; Snow & Lohman, 1984). Specifically, two of the factors probably would be measures of course cognitive demands and prior technical knowledge or experience needed (see Anderson, 1985; Kanfer & Ackerman, 1989; Mumford et al., 1988).

Estimation of Prediction Equations: MLR Analysis

An example. Suppose that some new selection measures have been developed for predicting performance and it is of interest to investigate their predictive validity for several jobs. In this example, we have a criterion (e.g., a score from a hands-on test of job performance) P_{ij} for person i in job j. We assume that P_{ij} depends on an individual's aptitude test score (call it A_{ij}; this could be a set of test scores) and some other set of other individual characteristics such as education and time in service (call this O_{ij}). We further assume that the effects of these independent variables could differ across jobs and that the jobs are a random sample of the total set of jobs. Thus, the model is:

$$P_{ij} = \alpha_j + \beta_j A_{ij} + \gamma_j O_{ij} + e_{ij} \qquad (1)$$

where α_j is a job-specific intercept, β_j and γ_j are job-specific slopes, and e_{ij} is an error term. This model says that α_j, β_j, and γ_j can, *in principle*, vary across jobs. *Multilevel regression* allows one to quantify the variation in these parameters and to determine if the variation is statistically significant. The variation is addressed by assuming that the parameters themselves have a stochastic structure. Namely:

$$\alpha_j = \alpha + a_j , \quad where \quad a_j \sim N (0, \sigma^2_a) , \qquad (2)$$

---

[4] Note that Mumford, Weeks, Harding, and Fleishman (1987) reported range restriction in the reading difficulty of technical training manuals. There is likely to be restriction of range on the reading grade level item in the TCS. Other items may show restriction of range as well. Range restriction is inherent in the Air Force training system due to selection on AFQT scores. Since the TCS ratings would be used to provide measures of training characteristics in a sample of Air Force courses, range restriction will not be an issue. However, restriction in range in the predictor and training criterion variables should be statistically corrected for the calculations of the correlation coefficients of the prediction equations for each course in a study.

$$\beta_j = \beta + b_j \ , \quad \textit{where} \quad b_j \sim N\left(0, \sigma^2_b\right) \ , \tag{3}$$

$$\gamma_j = \gamma + c_j \ , \quad \textit{where} \quad c_j \sim N\left(0, \sigma^2_c\right) \ . \tag{4}$$

Equation 2 says that the intercept for job j (α_j) has two components: α, the mean of all the α_js (note the lack of the j subscript), and a_j, a component that can be viewed as the amount by which job j's intercept differs from the average job's intercept (i.e., differs from α). Note that the model assumes the distributions of a_j, b_j, and c_j to be normal; their joint distribution is assumed to be multivariate normal. Although a_j, b_j, and c_j are completely determined for any specific job, the multilevel model conceives of these components as random, because the sample of jobs is assumed to be chosen at random. If the jobs are picked at random, these components are likewise random. Thus, coefficients modeled to vary across groups (here, jobs) may be labeled "random effects" (indeed, multilevel models are sometimes called random effects models), whereas coefficients modeled to remain constant across groups may be labeled "fixed effects." The variance components represent the variance of the random effects across jobs. For example, σ^2_a is the variance across jobs of the a_j's, and therefore of the α_j's, because α is the same for all jobs.

Why MLR? A multilevel regression model was suggested for the current project because the data are multilevel, or "nested." Specifically, individuals are nested within training courses (i.e., each individual takes one training course rather than all training courses). Individuals represent the first level (level one) and training courses the second level (level two). Returning to the example above, we need simply substitute "training course" for "job" such that P_{ij} is the performance of individual i in training course j. Equation 1 is a first-level equation: it models those observations nested within a higher level (i.e., individuals nested within training course). Specifically, the level-one equation models individual performance in a training course as a function of individual characteristics. Equations 2–4 are second-level equations: they model the variation in the first-level parameters.

Ordinary least squares (OLS) regression models are inappropriate for multilevel data. To see why this is so, consider a simpler version of Equation 1 in which only the intercept (α) is allowed to vary across training courses. That is, we wish to estimate α_j. The model is:

$$P_{ij} = \alpha_j + \beta A_{ij} + \gamma O_{ij} + e_{ij} \ , \tag{5}$$

and α_j is modeled by Equation 2. Substituting Equation 2 into Equation 5 results in a residual term of:

$$a_j + e_{ij} \ , \tag{6}$$

implying that the residuals from two individuals in the same training course are correlated (i.e., individuals within a training course share the same error component, a_j). The same situation obtains for the other parameters, as well. Therefore, applying the ordinary regression model to

these data would result in biased standard errors for the regression parameters (generally, biased downwards, increasing the chance of a Type I error).

Rather than treating the variation in the job-specific parameters as error, we usually try to model this variation as a function of other variables. Hence, Equations 2–4 (the second-level equations) are typically presented in the following form:

$$\alpha_j = \alpha + \pi_\alpha M_j + \eta_{\alpha j} , \tag{7}$$

$$\beta_j = \beta + \pi_\beta M_j + \eta_{\beta j} , \tag{8}$$

$$\gamma_j = \gamma + \pi_\gamma M_j + \eta_{\gamma j} , \tag{9}$$

where α, β, and γ are the mean values of the parameters across all courses (note the lack of the j subscript). The π's are vectors of coefficients constrained to be the same across courses (i.e., they are "fixed" coefficients); M_j is one or more variables that describe characteristics of the training course (e.g., method of instruction, content), and the η's are random variation.[5] (To generalize the model to the universe of training courses, the course-level coefficients—the πs— cannot be course-specific.)[6]

This structure for the model parameters assumes that some of their variation is due to characteristics of the training courses. The M_j variables represent characteristics of courses believed to influence an individual's performance in that course. Note that the training factors derived from the TCS will be used to provide the M_j variable scores. The inclusion of such course characteristic information allows one to generalize from a small sample of training courses to the population of military training courses. The amount of variance in the parameters that is unaccounted for can be reduced, when some portion of the parameter variation is due to course characteristics and the proper course characteristic variables (M_j s) are included in the multilevel model. This will increase the accuracy of prediction or, equivalently, decrease the standard error of estimate.

The M_j variables reduce the uncertainty in the course-specific parameters by absorbing some of the variation across courses that would be part of the random effect if the M_j variables were not in the model. For example, for the course-specific intercept α_j, the term $\pi_\alpha M_j$ models part of the variation in intercept parameters across courses that otherwise would be part of the random effect a_j. Including the second-level variables should reduce the uncertainty in the estimation of the α_j's. This same logic holds for all other model parameters.

---

[5] In this multilevel parameter specification, the course-level variables (i.e., the M_j variables) do not need to be the same for all parameters. In addition, the random error terms may covary.
[6] Those more familiar with analysis of variance will recognize this as a mixed model—one having both random and fixed effects.

The multilevel model may be approximated by a fixed-effects (i.e., conventional OLS) regression model. Substituting Equations 7–9 into 1 gives the following:

$$P_{ij} = (\alpha + \pi_\alpha M_j + \eta_{\alpha j}) + (\beta + \pi_\beta M_j + \eta_{\beta j})A_{ij} + (\gamma + \pi_\gamma M_j + \eta_{\gamma j})O_{ij} + e_{ij} \ , \qquad (10)$$

Multiplying through and collecting terms yields:

$$P_{ij} = \alpha + \beta A_{ij} + \gamma O_{ij} + (\pi_\alpha M_j + \pi_\beta M_j A_{ij} + \pi_\gamma M_j O_{ij}) + Z \ , \qquad (11)$$

where:

$$Z = \eta_{\alpha j} + \eta_{\beta j}A_{ij} + \eta_{\gamma j}O_{ij} + e_{ij} \ . \qquad (12)$$

Thus, a model containing course characteristics obtained from the TCS, and interactions between course characteristics and individual difference variables, may be used to estimate the structural parameters (regression coefficients) in the multilevel analysis. The standard errors of the parameter estimates for this model will be biased, however, due to the failure of the fixed effects regression to adequately model the correlations among errors in the multilevel error structure. The standard errors will typically be smaller than they should be, thereby increasing the probability of a Type I error.

Deriving course-specific equations. One of the principal advantages of the multilevel regression approach is that it allows performance predictions for courses having no criterion data. Using ordinary regression, performance scores can be estimated for individuals without criterion data by weighting their predictor information by the appropriate regression coefficients. However, performance data are needed in ordinary regression for some individuals in that course before the course-specific equation may be estimated. By including course characteristics in our multilevel model, course-specific parameters can be derived for any course having course characteristic data without performance data. These parameters are functions of the course characteristic variables.

For example, let us assume that the mean effect of A across courses is $\beta = .074$ and that we have four course characteristic variables (mean=0, sd=1.0). Also assume that the respective weights for these course characteristic variables (i.e., the π_β coefficients) are $-.030, .001, -.020,$ and $-.036$. Substituting these values into Equations 7 through 9 allows the estimation of course-specific parameters. Equations 7 through 9 also demonstrate that these estimated course-specific parameters are deviations from the mean parameter estimate—the degree of deviation being a function of the course's factor scores. If we assume that the scores on the four course characteristics for a given training course are $-0.68, -2.41, 2.33,$ and 0.18, then substituting these M_j values and the multilevel parameter estimates just given into (8) yields the A parameter (β_j) for predicting performance of individuals in this training course:

$$\beta_j = \beta + \pi_\beta M_j + \eta_{\beta j}$$

31

$$= .07 + [(-.030)(-.68) + (.001)(-2.41) + (-.020)(2.33) + (-.036)(-.32)]$$

$$= .07 + (-.03)$$

$$= .04.$$

This procedure thus affords course-specific parameters for courses without criterion data (see McCloy, 1994, for a description of generating job-specific performance equations for jobs that have no criterion data). Note that the value for β and the four π_β values remain constant in the β_j equations for all training courses; the equations differ only in the M_j values.

The model also may be amended to include additional or different individual and course characteristics. All that is required is to reestimate the multilevel regression equation with the new variables in the model so that new parameter values may be obtained. The procedure just described still applies.

Selection of Courses for a Classification-ATI Study

Selecting the treatment sample is an important part of the classification-ATI research method. In traditional classification research, the sample is comprised of jobs or job families. In the training context, the treatment sample will be comprised of courses. The purpose of this section is to present the major issues that should be considered when designing a sampling plan for a classification-ATI study. We discuss both general topics and those that are specific to the Air Force's technical training system. First, we define the scope of a course in the context of Air Force technical training. Second, we discuss a major obstacle we encountered in selecting a sample of courses for a potential study. Third, we suggest a set of Air Force Specialties (AFSs) from which courses could be selected if a study were conducted in the near term, and the general issues we considered in making our suggestions.

Definition of course. We defined course in this research method to include all instructional units in the Air Force's training pipeline. The training pipeline includes all fundamental and specialized units of instruction after basic training up through completion of 3-level course. We focused on 3-level courses only, which provide fundamental skills training to qualify recruits in a particular career field. We limited our proposed sample to this level of training because it provides ample numbers of students and is delivered in a fairly standardized manner across instructors. More importantly, selection of only 3-level courses limits our student sample to enlisted, entry-level personnel and levels the playing field in terms of what they already know going into training.

Further, limiting our sample to entry-level recruits and courses provides a rationale for the student-course matching simulation, which provides the estimate of the practical effects of ATIs on training performance. We describe the matching procedure below in the section entitled Description of the Student-Course Assignment Simulation. It would not make sense to

match Air Force enlisted personnel of different ranks to courses at various levels without considering experience, which is not a variable in our model.

Practical considerations in selecting a sample of Air Force courses for a classification-ATI study. As the first step in designing the treatment sampling plan, we conducted an informal survey of the Air Force 3-level technical training system. This included talking to Air Force training researchers and training managers at the technical schools about the types of courses available across the major occupational areas, student flow rates and other details about specific technical courses. Additionally, we reviewed the course catalogs within each technical area and discussed with training managers new courses and changes in existing courses.

Before presenting our suggestions for sampling Air Force courses, we describe a major constraint we encountered in designing the sampling procedure: very little variation in instructional methods. We found that most Air Force courses are taught in the classroom, with some having CBT or interactive videodisk (IVD) modules. In many cases the CBT or IVD modules are supplementary, rather than integral, parts of the course. A large number of courses include simulation modules. Distance learning is becoming increasingly prevalent in Air Force training. However, courses were just going on-line during this project, so no distance learning data were available. Finally, we found no operational courses based on adaptive tutors.

When we first proposed this project, our goal was to focus on method of instruction as our treatment variable. We expected to be able to compare methods of instruction within course content area (e.g., electronics courses presented in classroom, CBT and distance learning settings). However, we could not find any existing Air Force technical courses simultaneously presented by different methods of instruction. We did identify two or three courses that were changed from substantially classroom to mainly CBT, and one that was in the process of being reversed from CBT back to the classroom. But they were not adequate to fit our design for a variety of reasons (e.g., differences in the sampling timeframe for the two instructional methods).

Further, we could not sample instructional method across occupation. Although we found a large number of courses with CBT, IVD or simulation modules, we did not find a sufficient number that were completely, or even mainly, presented in any of these media. Consequently, we had to drop our notion of focusing on method of instruction as the main training characteristic and modify our sampling plan.

Our first idea was to sample *modules* within a given course that differed in medium of instruction. However, we rejected this approach because it does not fit the student-course matching procedure that forms the foundation of the classification-ATI paradigm. The matching procedure is based on the assumption that the treatments are equivalent in nature. Since modules are presented sequentially within a course (with many modules dependent on material learned in earlier modules) and all modules must be taken for course completion, the optimal matching of students to one of several modules did not make sense.

We finally settled on a compromise sampling design that meets all the assumptions and requirements of the classification-ATI paradigm and will produce a meaningful estimate of the effects of ATIs on mean predicted training performance (MPTP). The approach we recommend is to sample AFSs (each with an associated course) across the four main Air Force occupational areas: mechanical (M), administrative (A), general (G), and electronics (E). Further, we suggest that the researcher choose AFSs with courses that vary on as many of the training characteristics in the TCS as possible. We believe that by obtaining a good deal of variability in training environments, a researcher using this sampling approach would be able to identify a few strong training factors outside of occupational specialty.

We realize that student-course matching across occupational area is not practical, or even desirable, within the Air Force training environment, and we do not mean to suggest it as a change in policy. We suggest it only as a sampling procedure that solves the applied research problems of obtaining enough variation in technical training variables, and a large enough sample of courses, to provide an adequate test of the classification-ATI paradigm in the Air Force.

Although the compromise sampling procedure is not optimal for policy makers, and not one we would recommend if enough courses with different instructional methods were available, it will produce a good test of the classification-ATI paradigm, and one that is easy to communicate to a variety of audiences. Ideally, the Air Force will develop some courses with alternative methods of instruction (e.g., adaptive tutors and distance learning) in the near future so that a more realistic course sampling plan can be devised to test the classification-ATI paradigm.

In summary, we want to stress that our proposal of an ATI study that assesses student performance in alternative occupational areas was due solely to the absence of a variety of instructional methods in the Air Force technical training system. We attempt to moderate the influence of occupational specialty in the analysis by suggesting that the researcher choose courses that also vary on a large number of other training-related variables. Again, the design would produce a good initial test of the usefulness of the classification-ATI research paradigm for investigating ATIs in technical training environments.

Selection of the course sample. Tables 4 through 7 present the AFSs that we propose be included in a classification-ATI study by mechanical, administrative, general, and electronics (MAGE) occupational category. Our criteria for selecting AFSs (with their associated technical courses) were that they varied in terms of the major training variables measured by the TCS, namely, course content, level of difficulty, occupational area, and wherever possible, method of instruction. Examination of the TCS in the Appendix shows that we attempted to capture variation in instructional method by considering other variables in addition to media. For example, we asked about student-teacher ratio, number of tests and quizzes, and pace of the course.

Table 4. Selected Mechanical Air Force Specialties

AFSC	Title	Notes *
2A3X3	Tactical Aircraft Maintenance	"shredded" AFS, possible base course
2A5X0	Strategic Aircraft Maintenance	"shredded" AFS, possible base course
2A5X1	Airlift Aircraft Maintenance	"shredded" AFS, possible base course
2A6X1	Aerospace Propulsion	"shredded" AFS, possible base course
2A6x5	Aircraft Pneudraulic Systems	"shredded" AFS, possible base course
2A4XX	Weapon Control Systems	"shredded" AFS, possible base course
2W1XX	Aircraft Armament Systems	"shredded" AFS, possible base course
2A6X4	Aircraft Fuel Systems	
2A7X1	Aircraft Metals Technology	
2A7X3	Aircraft Structural Maintenance	
2A7X2	Non-destructive Inspection	
2M0X2	Missile Maintenance	drawdown impacted
2M0X3	Missile Facilities	drawdown impacted
2T3XX	Vehicle Maintenance	"mechanic", "shredded" AFS, possible base course
2E3X1	Structural Specialist	

* "Shredded" AFS refers to the differentiation of AFSCs into specialties that reflect particular aircraft. Possible base course indicates that all AFSCs with the same first three digits are likely to share a single set of preliminary courses.

Table 5. Selected Administrative Air Force Specialties

AFSC	Title	Notes
3A0X1	Information Management	large AFS
6F0X1	Financial Management	large AFS
3S0X1	Personnel	large AFS
3S0X2	Personnel Systems	
2T0X1	Traffic Management	
2S0X1	Inventory Management	
6C0X1	Contracting	
2R1X1	Maintenance Scheduling	
1C0X1	Airfield Management	
1C0X2	Operations Resource Management	
2S0X3	Materiel Storage and Distribution	
2T2X1	Air Transportation	

Table 6. Selected General Air Force Specialties

AFSC	Title	Notes *
3P0XX	Security & Law Enforcement	large general area, 2 AFSs, should have common basic course
1A2XX	Loadmaster	large aircrew areas, high math ability
1A0XX	In-Flight Refueling	aircrew, requires hand-eye coordination
1N0X1	Intelligence Ops	requires high general ability
1N4X1	Signals Intelligence	requires high general ability
1N0X2	Target Intel	requires high general ability
1N3XX	Cryptolinguist	"shredded" AFS, possible base course
1W0XX	Weather	high math ability
1C1XX	Air Traffic Control	high electric ability
5J0X1	Paralegal	
1T0XX	Survival Training	requires both content knowledge and teaching ability
4N1X1	Surgical Service	"shredded" AFS, possible base course
4N0X1	Medical Service	"shredded" AFS, possible base course
4T0X1	Medical Laboratory	"shredded" AFS, possible base course
4R0X1	Radiology	
4PX01	Pharmacy	
4Y0X1	Dental Assistant	

* "Shredded" AFS refers to the differentiation of AFSCs into specialties that reflect particular aircraft. Possible base course indicates that all AFSCs with the same first three digits are likely to share a single set of preliminary courses

Table 7. Selected Electronics Air Force Specialties

AFSC	Title	Notes *
2A0XX	Avionics Test Station and Component	"shredded" AFS, possible base course
2A3XX	Avionics System	"shredded" AFS, possible base course
2E0X1	Air Traffic Control Radar	"shredded" AFS, possible base course
2E0X2	Aircraft Control and Warning Radar	"shredded" AFS, possible base course
2E1X1	Wideband Communications Equipment	"shredded" AFS, possible base course
2E8X1	Instrumentation and Telemetry Systems	"shredded" AFS, possible base course
2E6X1	Systems Installation/Maintenance	"shredded" AFS, possible base course
2E1X2	Meteorological and Navigation Systems	
2E0X1	Electrical Systems	basic electrician skills
1N5XX	Electronic Intelligence	high general , high electronics abilities
1A4XX	Airborne Warning Command and Control System Operator	aircrew position--high general ability
2M0X1	Missile Systems Maintenance	impacted by drawdown

* "Shredded" AFS refers to the differentiation of AFSCs into specialties that reflect particular aircraft. Possible base course indicates that all AFSCs with the same first three digits are likely to share a single set of preliminary courses.

Given that a variety of occupational types is represented in our sample, we would expect to find some variation in training methods across courses because the methods will be at least partially adapted to course content. For example, courses in the administrative career field might rely heavily on drill and practice, while courses in the electronics and mechanical career fields might rely heavily on hands-on performance tasks. By selecting equal numbers of courses within different occupations, we attempted to tap whatever variation there is in method of instruction in Air Force 3-level technical training.

Another consideration in selecting AFSs and the courses associated with them is sample size. In a classification-ATI study, sample size refers to both the number of students within a course or treatment and the number of treatments. Concerning the number of students who have attended and completed a course (i.e., student flow) for which predictor and criteria data are available, it is always advantageous to obtain large sample sizes. However, small within course samples are not an insurmountable problem with the proposed classification-ATI design.

The MLR procedure we described above in the section entitled *Estimation of Prediction Equations: MLR Analysis* was developed specifically for educational research. It allows the use of courses with small samples because the individual difference parameters shown in Equation 1 are estimated from the total sample. In other words, the samples within courses are pooled for estimation of the predictor weights. This permits the inclusion of small samples without creating the deleterious effects of sampling error on the standard error of the predictor weights.

Regarding the number of courses, the classification-ATI research paradigm can be applied to a large number of courses, or to as few as two or three. However, including many courses can enhance the potential for obtaining person-treatment interactions, when the courses vary substantially in the training characteristics under investigation. In other words, when training settings are very different, having a large number of courses increases the chance that a student will perform differently in at least two settings.

A large number of treatments is needed for the MLR procedure to obtain precise measurement of course characteristics when computing course-specific prediction equations. This is because course characteristics are sampled in the same manner as individual difference variables, and the same rule of thumb about the ratio of number of variables to sample size applies. In other words, MLR requires about 8 to 10 courses per training variable for accurate measurement. Since we would expect to find three–five relevant training factors in the Air Force, the sample should have at least 20–40 courses. Although we recommend adhering to this rule of thumb, Harris et al. (1993) obtained stable estimates of person-treatment interactions in an OPJM study that had a sample of only 10 treatments with four treatment variables. They did not report any explanation for this finding, but it suggests that it may be worthwhile to try MLR with as few as 10 courses.

When fewer than 10 courses are available, say two, the classification-ATI paradigm can be used with traditional multiple regression, instead of with MLR. The downside is that traditional regression will not provide the detailed information on specific learner-training

interactions that MLR does, because it cannot employ the TCS data in forming the prediction equations.

In summary, Tables 4 through 7 contain a total of 56 AFSs, each associated with a separate course as defined above. The AFSs were selected to provide variation in occupational area, course content and difficulty, and method of instruction. In addition, these AFSs have high student flow rates, which would produce large within-course samples. If other AFSs with small course samples would add substantial differentiation, we suggest they be considered, since MLR compensates for small samples. Finally, if the Air Force expands the instructional media it employs in the near future to include adaptive tutors and distance learning, then courses presented in these formats also should be given strong consideration in designing a classification-ATI sampling plan.

Criterion Variables

When the classification research paradigm is used in employment testing, the criterion variable typically is a measure of performance on the job. This is the traditional criterion in personnel research because improving productivity is the major reason for instituting employment testing procedures. Further, job performance is considered to be a good indicator of global organizational effectiveness that can be tied to dollar estimates of a test's utility. Other criteria (e.g., attrition) have received less attention. Harris et al. (1993) incorporated both attrition and job performance into their model of classification.

We suggest that the criterion variable in a classification-ATI study be numerical final course grade. Other possible criteria could be training time, washback rate, and number of extracurricular tutoring sessions. We believe that a measure of training achievement is superior to the other criteria because it is a global measure of learning success that represents performance in the entire course. Additionally, because it is comprehensive, final course grade probably is less biased by variables outside the control of the student than are training time, remedial tutoring sessions, and washback rate. Training performance measures have been used in some OPJM research as a surrogate for job performance when that criterion was not available (Alley & Teachout, 1992; Darby et al., 1995; Johnson, Zeidner, & Leaman, 1992). These studies showed positive results with OPJM strategies compared to random assignment, thus providing a precedent for use in a classification-ATI study.

In creating the criterion variable for a classification-ATI study, we suggest that only those units that assign grades be included in the analysis; all units that assign "pass/fail" scores should be excluded because they do not provide enough information about performance to be useful for identifying statistically significant ATIs.

Selection of Predictors

The major objectives in constructing a differential prediction battery are to maximize the potential for differential prediction across courses (reflected in the term $[1 - r]^{1/2}$ in Brogden's 1959 classification theorem) and the average validity of the prediction equations (i.e., R).

Johnson and Zeidner (1991) specified that the objectives are accomplished by selecting a multidimensional set of individual difference measures with a view toward covering as much of the criterion domain as possible.

A measure of general cognitive ability (*g*) is the best single predictor of both job and training performance (Hunter, 1986; Hunter & Hunter, 1984; Ree & Earles, 1991). However, the addition of other measures, (e.g., psychomotor ability, job-related personality variables, and interests) has improved both differential prediction efficiency across treatments and predictive validity with both criteria in person-job matching studies (Hunter & Schmidt, 1982; Schmidt, Hunter, & Dunn, 1987; Statman, 1993; Statman et al., 1994; Wise, McHenry, & Campbell, 1990).

Traditionally ATI research is designed to investigate a single predictor across diverse training environments. Often it is a measure of *g,* but Maldegen et al. (1996) found a large number (44) of other predictors (e.g., working memory, motor skills, anxiety, conformity, impulsivity, and self-efficacy) in ATI research, and little replication of studies. The lack of consistency in the selection of predictors (and training settings—another finding by Maldegen et al. [1996]) may be partially responsible for the confusion of results in ATI research.

The classification-ATI paradigm we designed may provide a strategy for addressing this limitation, because the MLR procedure allows us to examine the statistical significance and strength of multiple predictor-training-variable interaction terms simultaneously. By studying more than one learner characteristic in a single ATI study, we may gain insight into the reasons for the variation in ATI results obtained in separate studies of these predictors.

Although the results of the ATI and training literatures are far from unequivocal about the presence of ATIs, our review and that of Maldegen et al. (1996) indicated that general cognitive ability, cognitive and learning styles (especially verbal learning ability), prior knowledge of the course material, psychomotor skills, visual-spatial ability, and working memory would make good candidates for inclusion in a battery designed to detect training ATIs. Since the focus of our proposed classification-ATI research is job-related technical training, measures of vocational interest and job-related personality characteristics (e.g., those measured by AIM) might also interact with training variables.

We recommend use of a highly diversified battery of cognitive and non-cognitive predictors with the classification-ATI method. However, the Air Force only had data available for the ASVAB, which is fundamentally a cognitive test, across a broad range of 3-level courses during this project. Consequently, we propose that initial classification-ATI research be conducted with the ASVAB.

The Air Force's APT battery may be considered in the future because predictor and criterion data were collected recently in 18 AFSs. Since the APT is an information processing battery, which includes measures of working memory and processing speed for verbal, quantitative and spatial abilities (Kyllonen, 1994), it may provide additional sources of variance in training performance that cannot be obtained from the ASVAB. Another possibility for

inclusion in future classification-ATI research is the AIM, which was described above in the *Review of the Training Literature.*

In brief, the ASVAB comprises eight power and two speeded tests. Factor analytic studies consistently indicate that most of the variance in the 10-test space is accounted for by four factors: verbal ability, speeded performance, quantitative ability and technical knowledge (which includes mechanical, electronics and auto shop information) (Welsh, Kucinkas, & Curran, 1990). As mentioned in the description of the TCS development process, we designed that survey to tap elements of the training environment that are congruent with the ASVAB to maximize the potential for finding ATIs. (William Alley, Ph.D., of the Air Force made this valuable suggestion at the start of the project.) If other measures of learning characteristics are incorporated into Air Force research, then the TCS should be expanded to include training characteristics related to those variables. For example, if the AIM were to be used, then the TCS should be modified to include additional training characteristics that researchers hypothesize would tap motivation, dependability and work ethic (e.g., absences and attendance at extra-curricular activities).
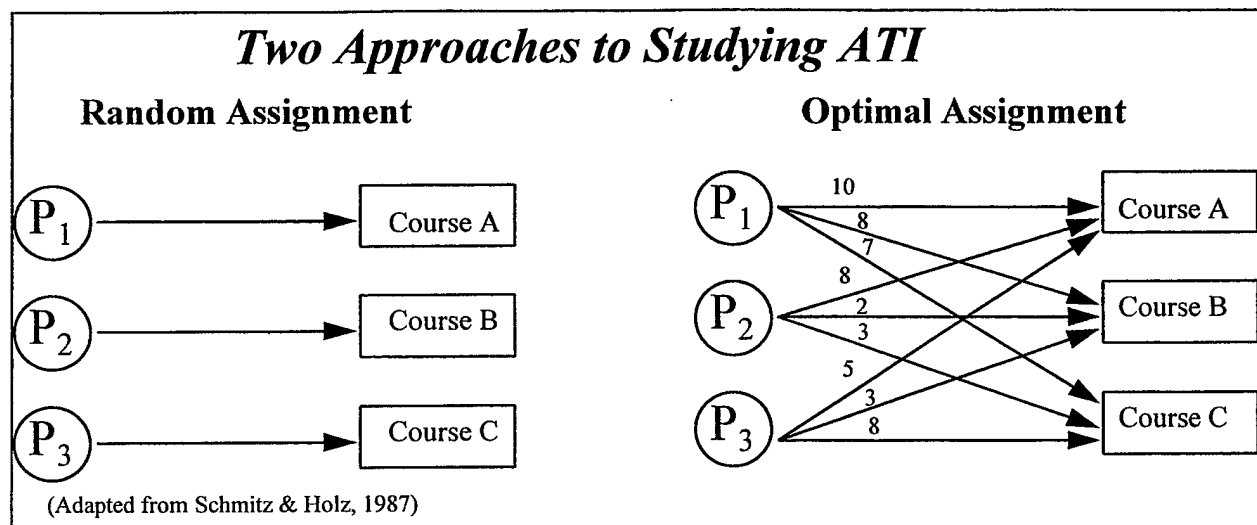
Simulation of the Student-Course Matching Process

Simulation of a student-course matching process is the core of the classification-ATI research paradigm. We divide our description of the process into five sections:

- description of the student-course assignment simulation
- measurement of student-course matching simulation results
- specification of the experimental conditions
- the classification cross-validation procedure
- use of synthetic samples for cross-validation

Description of the student-course assignment simulation. Figure 1 presents a schematic diagram that compares the traditional ATI research design to the classification-ATI method. In the traditional ATI study depicted on the left of Figure 1, students are randomly assigned to courses (treatments). Pretest and posttest (i.e., criterion) measures are obtained for each student. A separate regression equation is computed for each course by regressing the criterion (e.g., training achievement) on the pretest measure. Significant differences in the slopes of the regression lines across courses indicate the presence of an ATI.

The traditional ATI method has two significant limitations that are addressed by the classification paradigm we propose. First, it does not produce a quantitative measure of the effect of an ATI on training performance. Second, it presents a global indication of differences in training environments, but does not provide a means for identifying the exact nature of the training characteristics that may be producing intra-individual differences in learning across settings

Figure 1. Two Approaches to Studying ATIs

The right side of Figure 1 depicts the classification-ATI methodology, which employs a very different approach for detecting ATIs. This method uses an optimal student-course matching process to assign individuals to treatments. The objective of the assignment procedure is to place each student in the course in which he or she is expected to perform best.

A student's predicted performance in each course is estimated by a course-specific prediction equation, which is a weighted composite of predictor information and predictor-by-training-variable interaction terms (see *Estimation of Prediction Equations: MLR Analysis* for how to compute the test weights and ATI terms).[7] Each student receives a separate predicted performance score for each course. If the TCS and MLR procedure successfully detect ATIs, then each student will have a different score for each course.

The differences in a student's scores across courses will be a direct function of the ATIs. This is because the MLR procedure computes one set of test weights for all courses, with only the interaction terms varying according to the variation in training characteristics across courses. (Remember that the MLR procedure provides a statistical test of the significance of the interaction terms, which is one indication of the presence of ATIs.)

---

[7]Note that we standardize the criterion scores within course to control for differences in the difficulty level of the performance measures. We also use standardized test weights (removing the regression constant from the prediction equations) to control for the effect of different within-course mean criterion scores on assignment. Variation in within-course mean scores would indicate that courses differ in difficulty level. The TCS contains items on course difficulty. Therefore, if any significant differences in difficulty among courses do exist, their effects will be seen in the interactions of the course difficulty factor with the predictors.

In the example on the right side of Figure 1, all three students have a different score for each treatment setting. For instance, Person one has scores of 10 in course A, 8 in course B, and 7 in course C. The variation in the scores of the three students indicates the presence of ATIs.

Linear programming (LP) software is used to conduct person-job matching simulations in employment testing. We suggest that the same type of software be used to conduct the student-course matching process. Depending upon the purpose of the ATI study, the LP can be designed to control or account for organizational constraints (e.g., differences in course sizes). If the purpose is to conduct an experimental study comparing different methods of instruction (e.g., classroom, CBT, distance learning), then organizational variables should not be included in the design of the LP. In this case, the simulation simply should assign each person to the treatment for which he or she has the best score. The result will be optimal assignment and optimal average performance in all courses.

However, if the purpose is to evaluate ATI effects under fairly realistic conditions, then the LP should reflect practical organizational constraints. Important variables to consider might be course size and seasonal variation in student sub-populations (e.g., graduating seniors vs. recruits who enter the Air Force during the school year). The constraints and procedures for making trade-offs between achieving optimal performance and meeting other organizational goals are programmed directly into the software, which is a mathematical model designed to simulate the organization's policy. When organizational constraints on optimal assignment are included in the matching LP, average performance after assignment is reduced. This is because the LP will make tradeoffs between producing the highest average performance and accommodating factors like course size or seasonal variation in size of the Air Force applicant pool.

Measurement of student-course matching simulation results. Figure 2 presents an overview of the variables, procedures and sequence of operations that make up the proposed classification-ATI paradigm. The process of preparing the data requires selecting the course-specific criterion variable and the predictors of learner characteristics. When MLR is employed, the training characteristic variables in the TCS must be logically matched to learner characteristics and the hypothesized relationships stated a priori. Finally, a representative sample of courses, which is hypothesized to vary along the dimensions under investigation, must be selected.

As mentioned above, two or more course settings can be studied with the classification-ATI design. However, if MLR is employed, then a large number of courses is needed to provide an adequate number of observations for the training characteristic variables. In MLR the two levels of variables for which samples must be obtained are individual difference characteristics and treatment characteristics. If only a small number of courses are available or desirable for study, then traditional regression analysis can be used instead of MLR in our proposed design. However, the TCS cannot be used with traditional regression and the researcher will not be able to obtain information about the specific training characteristics involved in ATIs.
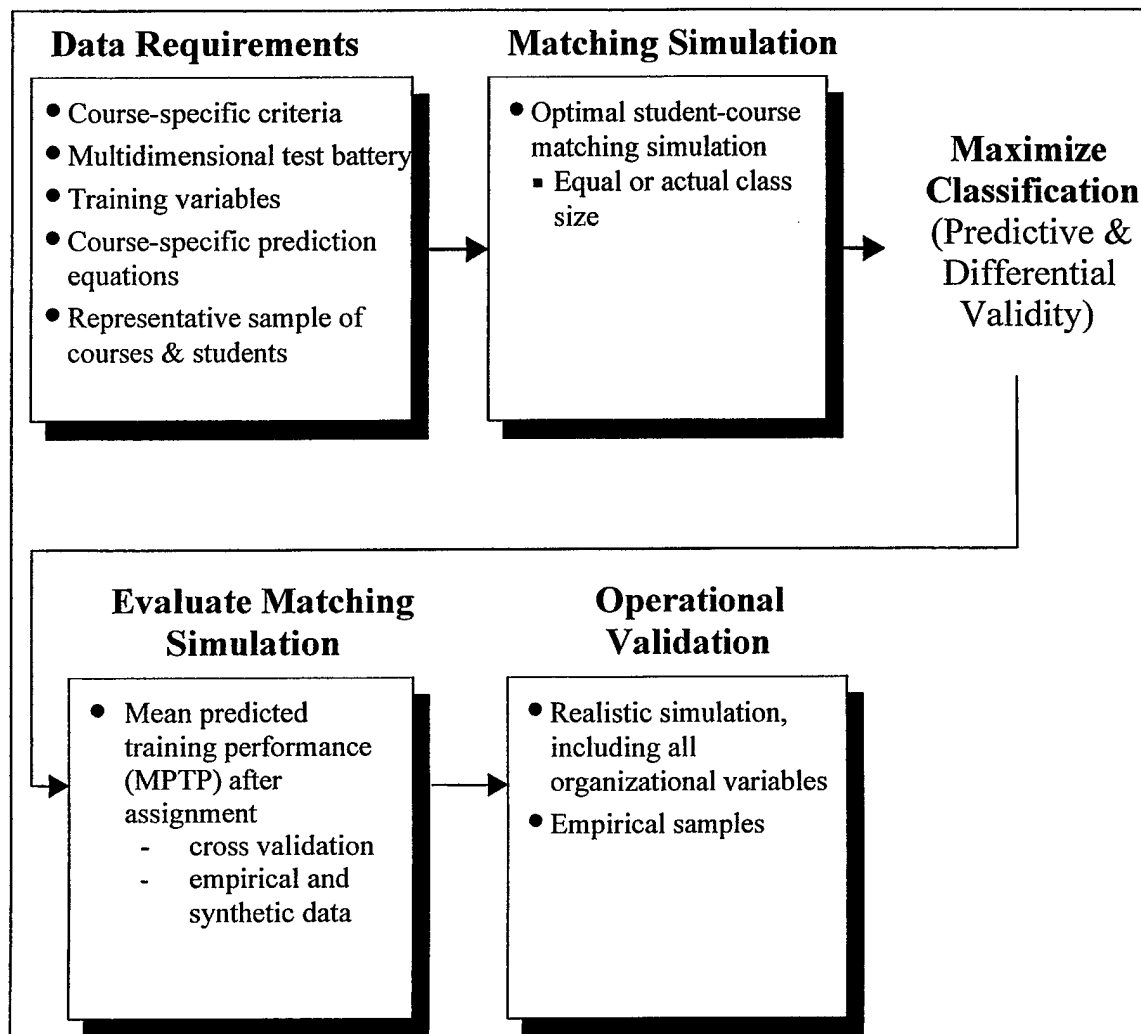
Figure 2. Overview of the Classification-ATI Research Paradigm

As we will discuss under the classification cross-validation procedure, the total sample of students in all courses is randomly segmented into subsamples that are used to construct the course-specific prediction equations, provide the pool of students for optimal person-treatment matching, and evaluate the ATI effects after the assignment simulation.

An example of a student-course matching simulation that uses the TCS contained in the Appendix, MLR, and the 3-level courses from the AFSs listed in Tables 4–7 follows. Select a representative sample of students from each course for a given time period. Use two-thirds of the sample to compute the predictor weights for the first level prediction equation. Administer the TCS to a sample of 5 to 10 training SMEs in each course (e.g., training developers and managers, and instructors). Compute a principal components analysis and a varimax rotation to simple

structure of the TCS items. Obtain mean principal component scores for each course on factors with eigenvalues ≥ 1.00. These principal components will be the training characteristic variables. Compute the interaction terms between the training and learner variables using the MLR procedure described in *Estimation of Prediction Equations: MLR Analysis*. This will result in a set of course-specific regression equations that reflect both learning characteristics and statistically significant ATIs.

Once the course equations are obtained, compute a separate course score for all members of the one-third hold-out sample. Then run this sample through the LP matching software. The outcome will be the assignment of each student to the course for which he or she had the best score. Since our proposed design samples across occupational areas and training characteristics, we suggest incorporating variation in course size as a constraint into the LP.

The measure of the effect of any ATIs identified in the MLR procedure is MPTP. As stated in the Introduction, this is a measure of the average performance of all students in all courses after assignment. As in the personnel classification paradigm, the dependent variable should be a standardized score that is obtained by standardizing the criterion variables within each course. (See Footnote 7 for a more detailed discussion of this issue.) We suggest using a mean of 0.00 and an SD of 1.0 for ease in interpreting the results.

If no significant ATIs are present, then each student would have about the same score in each course, and all students would be *randomly assigned* to courses. This would produce an MPTP standard score of 0.00, the mean of all the standardized criterion scores for the assignment pool. Thus, any MPTP significantly greater than 0.00 would indicate the presence of an ATI.

The level of MPTP obtained is a measure of the practical effects of ATIs on training performance. As mentioned in the Introduction, assignment simulation results from personnel testing have been linked to human resource budgets using a variety of approaches (Harris, McCloy, DiFazio & Hogan, 1993; Nord and Schmitz, 1991; Nord and White, 1988; Schmidt, Hunter, & Dunn, 1987). Similar approaches could be used to estimate the budgetary savings achieved by optimally assigning Air Force recruits to different training settings.

As a supplement to MPTP scores, the interaction terms in the course-specific MLR equations identify the specific training factors and predictor variables that produce interactions. Further, the terms indicate whether the interactions are statistically significant and quantify the strength of those interactions. Thus, the adaptation of the personnel classification paradigm to ATI research produces quite a bit more information than is provided by the traditional ATI research design.

Specification of the experimental conditions. We think it is valuable to compare the MPTP produced by different sets of predictors (and the accompanying predictor-training-variable interaction terms), and suggest comparing batteries made up of the following combinations of ASVAB factors:

- verbal composite alone
- verbal and quantitative composites (i.e., AFQT)
- verbal, quantitative, and technical composites
- verbal, quantitative, technical and speed composites

These four batteries should be compared to two baseline conditions: actual and random assignment. This comparative analysis will provide information about the relative differences in the practical benefits of different combinations of ATIs for training performance. If the results are positive, they could be used to develop technical training courses (including lecture, CBT, distance learning, and adaptive tutors) that capitalize on the specific learner-training-variable interactions identified by the classification-ATI research paradigm. If data bases of new predictor batteries that appear to be relevant to ATI research become available to the Air Force, then we would suggest creating a set of conditions that make comparisons among complete batteries (e.g., ASVAB vs. AIM vs. APT).

The classification cross-validation procedure. Johnson and Zeidner (1991) strongly recommend using a classification cross-validation procedure to control for overfitting the prediction equations, which causes inflation of the predicted performance measure (i.e., MPTP). Since the classification research method (more specifically, the assignment simulation) uses prediction equations differently from traditional regression analysis procedures (like those used in typical ATI and test validation research), three independent samples from the same population are needed. (If MLR is employed then only two samples are needed, but they are not used in the same way as in traditional cross-validation research [see below]).

The first sample is used to form the treatment-specific prediction equations for the assignment simulation. The second sample (which does not need to have scores on performance measures) is the student-course matching pool that is run through the person-treatment matching simulation. The second sample should be fairly large and divided into 20 or 30 batches. This strategy provides a distribution of MPTP scores. The scores can be entered into an analysis of variance procedure that compares the various conditions under investigation.

The third sample should be the same size as the first. It is used to compute an independent set of test weights for the treatment-specific prediction equations. These prediction equations are used to reestimate MPTP after the assignment is conducted. Reestimation of MPTP is an additional control for overfitting of the original set of prediction equations. When several different batteries are compared, a single set of prediction equations that includes all of the tests in the study should be used so that MPTP scores are equivalent across conditions.

We suggest using MLR in the proposed research design, because it circumvents the weakness of small within-treatment samples. Thus, it alleviates the need for the third sample. In traditional testing research MLR employs the full sample of test data to compute predictor weights. In the classification-ATI procedure, MLR can be used with two-thirds of the sample to compute the weights for both the assignment equations and for computation of MPTP after assignment, based on all predictors in the study. The hold-out sample of one-third of the observations will be used to provide subjects for the student-course matching pool.

Use of synthetic samples for cross-validation. Because classification cross-validation procedures need large sample sizes, Johnson, Zeidner and others (e.g., Johnson & Zeidner, 1991; Nord and Schmitz, 1991; Statman, 1993; Statman et al., 1994) have employed a Monte Carlo technique to produce additional samples of synthetic data. Their general approach is to map the variance-covariance structure of the population of interest onto a random normal distribution. This procedure is used extensively by statisticians for many different types of simulations. However, in the classification context in which we are attempting to simulate operational organizational conditions, it tends to produce inflated results. This is because the actual military applicant and recruit populations vary from a strictly normal distribution and because it is impossible to synthesize all of the random characteristics of real data. The Johnson-Zeidner classification design requires one empirical sample, which is used to compute the differential equations for assignment, and two synthetic samples, one for the assignment pool and one to evaluate MPTP after assignment.

However, we suggest a different approach. Balancing our concerns about the limitations of synthetic data with those of overfitting prediction equations due to small samples, we recommend using MLR to eliminate or reduce the need for synthetic samples. If the overall sample is large, two-thirds of the subjects can be used to compute the prediction equations for assignment. The one-third hold out sample then will be used as the matching pool. If the overall sample is small, then the full data base can be used to create the training-specific prediction equations and to compute MPTP. Only one synthetic sample will be needed—for the person-treatment matching pool.

Conclusion

We have described a classification-ATI research method that is designed to improve the detection and measurement of ATIs, and to provide an estimate of their practical effects on training performance. With further development, this method can be extended to include estimates of the savings in training dollars due to optimal matching of students to training settings (e.g., classroom lectures, CBT, distance learning, and adaptive tutors).

The classification-ATI method is composed of four major procedures:

- selection of the set of learner variables hypothesized to interact with training settings
- measurement of specific training variables with the TCS developed in this project
- computation of course-specific prediction equations that quantify and statistically test ATIs using MLR analysis
- simulation of a student-course matching process that capitalizes on ATIs, if they are present

We believe that the classification-ATI method developed in this project will improve ATI research by providing a means of simultaneously analyzing multiple ATIs in a single setting. This should shed some light on the conflicting findings in the traditional ATI literature. Further, the improved identification and measurement of the practical effects of ATIs will be useful in both training design and evaluation research. Finally, we mentioned above that the classification-ATI paradigm can be expanded to include cost-benefit analysis of the savings captured by optimal student-course matching (or of the gains due to higher technical performance) through use of ATIs in training development.

References

Ackerman, P.L., Sternberg, R.L., & Glaser, R. (Eds.). (1989). *Learning and individual differences.* New York, NY: W.H. Freeman and Company.

Alley, W. E., & Darby, M. M. (1995). Estimating the benefits of personnel selection and classification: An extension of the Brogden table. *Educational and Psychological Measurement, 55*(6), 938-958.

Alley, W. E., Darby, M. M., & Cheng, C. (1996, November). The practical benefits of personnel testing: An extension of the Taylor-Russell tables to multiple job categories. Paper presented at the International Military Testing Association, San Antonio, TX.

Alley, W. E., & Teachout, M. S. (1992, August). *Differential assignment potential in the ASVAB.* Paper presented at 100th Annual Convention of the American Psychological Association, Washington, DC.

Anderson, J. R. (1985). *Cognitive psychology and its implications* (2nd ed.). New York: W.H. Freeman and Company.

Arshad, F. N., & Kelleher, G. (1990). Learning with a computer-based advisor. *Current Psychology Research and Reviews, 9*(2), 131-140.

Bacdayan, A. W. (1994). Time-denominated achievement cost curves, learning differences and individualized instruction. *Economics of Education Review, 13*(1), 43-53.

Baker, E. L., & O'Neil, H. F., Jr. (1986). *Evaluation approaches to intelligent computer assisted instruction.* Paper presented at the Workshop on Potential Applications of Intelligent Computer Assisted Instruction in Military Training, Los Angeles, CA.

Brogden, H. (1946). An approach to the problem of differential prediction. *Psychometrika, 11,* 139-154.

Brogden, H. (1951). Increased efficiency of selection resulting from replacement of a single predictor with several differential predictors. *Educational and Psychological Measurement, 11,* 173-196.

Brogden, H. (1954). A simple proof of a personnel classification theorem. *Psychometrika, 19*(3), 205-208.

Brogden, H. (1955). Least squares estimates and optimal classification. *Psychometrika, 20*(3), 249-252.

Brogden, H. (1959). Efficiency of classification as a function of number of jobs, percent rejected, and the validity and intercorrelation of job performance estimates. *Educational and Psychological Measurement, 19*(2), 181-190.

Brogden, H. (1964). Simplified regression patterns for classification. *Psychometrika, 29,* 393-396.

Brophy, J. (1986). Teacher influences on student achievement. *American Psychologist, 41*(10), 1069-1077.

Campbell, J. P. (1988). Training design for performance improvement. In J. P. Campbell, R. J. Campbell, & Associates (Eds.), *Productivity in organizations* (pp. 177-215). San Francisco: Jossey-Bass.

Cannon-Bowers, J. A., Tannenbaum, S. I., Salas, E., & Converse, S. A. (1991). Toward an integration of training theory and technique. *Human Factors, 33*(3), 281-292.

Cronbach, L. J., & Gleser, G. C. (1965). *Psychological tests and personnel decisions.* Urbana: University of Illinois.

Cronbach, L. J., & Snow, R. E. (1977*). Aptitudes and instructional methods: A handbook for research on interactions.* New York: Irvington.

Darby, M. M., Skinner, J., & Alley, W. E. (1995, November). *Classification potential in current Air Force test composites.* Paper presented at the International Military Testing Association, Toronto.

de Leeuw, L. (1983). Teaching problem solving: An ATI study of the effects of teaching algorithmic and heuristic solution methods. *Instructional Science, 12,* 1-48.

Dixon, N. (1982). Incorporating learning style into training design. *Training and Development Journal, 36*(7), 62-64.

Donchin, E. (1989). The Learning Strategies project: Introductory remarks. *Acta Psychologica, 71*(1-3), 1-15.

Fleishman, E. A., & Reilly, M. E. (1992). *Handbook of human abilities: Definitions, measurements, and job task requirements.* Bethesda, MD: Management Research Institute.

Frederiksen, N. (1984). The real test bias: Influences of testing on teaching and learning. *American Psychologist, 39*(3), 193-202.

Gist, M. E., Schwoerer, C., & Rosen, B. (1989). Effects of alternative training methods of self-efficacy and performance in computer software training. *Journal of Applied Psychology, 74*(6), 884-891.

Goldstein, I. L. (1993). *Training in organizations: Needs assessment, development, and evaluation* (3rd ed.). Pacific Grove, CA: Brooks/Cole Publishing Company.

Gregorc, A. F. (1979). Learning/teaching styles: Their nature and effects. In National Association of Secondary School Principals (Eds.), *Student learning styles: Diagnosing and prescribing programs* (pp. 19-26). Reston, VA: Author.

Harris, D. A., McCloy, R. A., Dempsey, J. R., DiFazio, A., & Hogan, P. F. (1993, November). *Personnel enlistment testing, job performance and cost: A cost-effectiveness analysis* (HumRRO FR-PDR-93-35). Alexandria, VA: Human Resources Research Organization.

Harris, D. A., McCloy, R. A., Dempsey, J. R., Roth, C., Sackett, P. R., Hedges, L. V., Smith, D. A., & Hogan, P. F. (1991). *Determining the relationship between recruit characteristics and job performance: A methodology and a model* (HumRRO FR-PRD-90-17). Alexandria, VA: Human Resources Research Organization.

Horst, P. (1954). A technique for the development of a differential prediction battery. *Psychological Monographs: General and Applied, 68*, (5, Whole No. 8).

Horst, P. (1956). Multiple classification by the method of least squares. *Journal of Clinical Psychology, 12*, 3-16.

Hunter, J. E. (1986). Cognitive ability, cognitive aptitudes, job knowledge, and job performance. *Journal of Vocational Behavior, 29*, 340-362.

Hunter, J. E., & Hunter, R. F. (1984). Validity and utility of alternative predictors of job performance. *Psychological Bulletin, 96*, 72-98.

Hunter, J. E., & Schmidt, F. L. (1982). Fitting people to jobs: The impact of personnel selection on national productivity. In E. A. Fleishman and M. D. Dunnette (Eds.), *Human performance and productivity, Vol. 1: Human capability assessment* (pp. 233-284). Hillsdale, NJ: Erlbaum.

Hunt, D. E. (1979). Learning style and student needs: An introduction to conceptual level. In National Association of Secondary School Principals (Eds.), *Student learning styles: Diagnosing and prescribing programs* (pp. 27-38). Reston, VA: Author.

Johnson, C. D., & Zeidner, J. (1991*). The economic benefits of predicting job performance: Vol. II. Classification efficiency.* New York: Praeger.

Johnson, C. D., Zeidner, J., & Leaman, J. A. (1992*). Improving classification efficiency by restructuring Army job families* (Technical Report 947). Alexandria, VA: Army Family Research Institute.

Kanfer, R., & Ackerman, P. L. (1989). Motivation and cognitive abilities: An integrative/ aptitude-treatment interaction approach to skill acquisition. *Journal of Applied Psychology, 74*(4), 657-690.

Kearsley, G. P. (1977). *Instructional design in CAI.* Edmonton, CA: Division of Educational Research Services, Alberta University. (ERIC Document Reproduction Service No. ED 152 293)

Kearsley, G. P., & Hillelsohn, M. J. (1982). Human factors considerations for computer-based training. *Journal of Computer-Based Instruction, 8*(4), 74-84.

Keefe, J. W. (1979). Learning style: An overview. In National Association of Secondary School Principals (Eds.), *Student learning styles: Diagnosing and prescribing programs* (pp. 1-17). Reston, VA: Author.

Kozlowski, S. W. J. (1995). Enhancing the training and development of adaptive expertise. *Psychological Science Agenda*, (Sept-Oct), 7-9.

Kraiger, K., Ford, J. K., & Salas, E. (1993). Application of cognitive, skill-based, and affective theories of learning outcomes to new methods of training evaluation. *Journal of Applied Psychology, 78*(2), 311-328.

Kramer, T. J., & Korn, J. H. (1996). Class discussions: Promoting participation and preventing problems. *APS Observer, 9*(5), 24-25.

Kyllonen, P. C. (1994). Cognitive abilities testing: An agenda for the 1990s. In M. G. Rumsey, C. B. Walter, & J. H. Harris (Eds.), *Personnel selection and classification* (pp. 103-125). Hillsdale, NJ: Erlbaum.

Lord, F. M. (1952). Notes on a problem of multiple classification. *Psychometrika, 17*(3), 297-304.

Maldegen, R., Statman, M. A., Gribben, M., & Yadrick, R. (1996). Analysis of aptitude-treatment interactions from a personnel classification perspective. Paper presented at the International Military Testing Association, San Antonio, TX.

Mark, M. A., & Greer, J. E. (1995). The VCR tutor: Effective instruction for device operation. *Journal of the Learning Sciences, 4*(2), 209-246.

McCloy, R. A. (1994) Predicting job performance scores for jobs without performance data. In B. F. Green & A. S. Mavor (Eds.), *Modeling cost and performance for military enlistment.* Washington, DC: National Academy Press.

McCombs, B. L., & McDaniel, M. A. (1981). On the design of adaptive treatments for individualized instructional systems. *Educational Psychologist, 16*(1), 11-22.

Mumford, M. D., Weeks, J. L., Harding, F. D., & Fleishman, E. A. (1988). Relations between student characteristics, course content, and training outcomes: An integrative modeling effort. *Journal of Applied Psychology, 73,* 443-456.

Mumford, M. D., Weeks, J. L., Harding, F. D., & Fleishman, E. A. (1987). Measuring occupational difficulty: A construct validation against training criteria. *Journal of Applied Psychology, 72*(4), 578-587.

Nord, R., & Schmitz, E. (1991). Estimating performance and utility effects of alternative selection and classification policies. In J. Zeidner and C. D. Johnson, *The economic benefits of predicting job performance. Vol III: The gains of alternative policies.* New York: Praeger.

Nord, R., & White, L. (1988). The measurement and application of performance utility: Some key issues. In B. G. Green, Jr., H. Wing, & A. K. Wigdor (Eds.), *Linking military enlistment standards to job performance: Report of a workshop* (pp. 215-243). Washington, DC: National Academy Press.

Peterson, P. L., & Janicki, T. C. (1979). Individual characteristics and children's learning in large-group and small-group approaches. *Journal of Educational Psychology, 71*(5), 677-687.

Ree, M. J., & Earles, J. A. (1991). Predicting training success: Not much more than g. *Personnel Psychology, 44,* 271-278.

Reid, D. H., & Parsons, M. B. (1996). A comparison of staff acceptability of immediate versus delayed verbal feedback in staff training. *Journal of Organizational Behavior Management, 16*(2), 35-47.

Reynolds, A., & Anderson, R. H. (1992). *Selecting and developing media for instruction* (III ed.). New York: Van Nostrand Reinhold.

Ryder, J. M., & Redding, R. E. (1993). Integrating cognitive task analysis into instructional systems development. *Educational Technology Research and Development, 41*(2), 75-96.

Sager, C., Peterson, N. G., Oppler, S. H., & Rosse, R. L. (1997). An examination of five indexes of test battery performance: Analysis of the ECAT battery. *Military Psychology, 9*(1), 97-120.

Salmoni, A. W., Schmidt, R. A., & Walter, C. B. (1984). Knowledge of results and motor learning: A review and critical reappraisal. *Psychological Bulletin, 95*(3), 355-386.

Savage, R. E., Williges, B. H., & Williges, R. C. (1982). Empirical prediction models in training group assignment. *Human Factors, 24*, 417-426.

Schmidt, F. L., Hunter, J. E., Dunn, W. L. (1987). *Potential utility increases from adding new tests to the Armed Services Vocational Aptitude Battery (ASVAB).* Contract No. Delivery Order 0053). San Diego, CA: U.S. Navy Personnel Research and Development Center.

Schneider, W. (1985). Training high-performance skills: Fallacies and guidelines. *Human Factors, 27*(3), 285-300.

Schoenfeldt, L. F. (1982). Intra-individual variation and human performance. In M. D. Dunnette & E. A. Fleishman (Eds.), *Human performance and productivity: Vol I: Human capability assessment.* (pp. 107-134) Hillsdale, NJ: Erlbaum.

Scholarios, D. M., Johnson, C. D., & Zeidner, J. (1994). Selecting predictors for maximizing the classification efficiency of a battery. *Journal of Applied Psychology, 79*(3), 412-424.

Schroth, M. L. (1995). Variable delay of feedback procedures and subsequent concept formation transfer. *Journal of General Psychology, 122*(4), 393-399.

Sein, M. K., & Bostrom, R. P. (1989). Individual differences and conceptual models in training novice users. *Human-Computer Interaction, 4*, 197-229.

Shaw, R., & Bunt, M. M. (1979). Behavior of college students when matched or mismatched for course structure. *Journal of Educational Research, 73*(1), 41-45.

Shute, V. J. (1991). *Meta-evaluation of four intelligent tutoring systems: Promises and products.* (Final Technical Paper AL-TP-1991-0040). Brooks AFB, TX: Human Resources Directorate, Armstrong Laboratory.

Shute, V. J., Lajoie, S. P., & Gluck, K. A. (in press). Individualized and group approaches to training. In S. Tobias & D. Fletcher (Eds.), *Handbook on training.*

Siem, F. M., & Alley, W. E. (1997). Optimal personnel assignment: An application to Air Force pilots. *Military Psychology, 7*(4), 253-263.

Sleeman, D., & Brown, J. S. (1982). *Intelligent tutoring systems.* London: Academic Press.

Smith-Jentsch, K. A., Jentsch, F. G., Payne, S. C., & Salas, E. (1996). Can pretraining experiences explain individual differences in learning? *Journal of Applied Psychology, 81*(1), 110-116.

Smith, R. M., Neisworth, J. T., & Greer, J. G. (1978). *Evaluating educational environments.* Columbus, OH: Charles E. Merrill.

Snow, R. E., & Lohman, D. F. (1984). Toward a theory of cognitive aptitude for learning from instruction. *Journal of Educational Psychology, 76*(3), 347-376.

Statman, M. A. (1992). Developing optimal predictor equations for differential job assignment and vocational counseling. In J. Zeidner (Chair), *Differential assignment: The classification-efficient person-job matching technology.* Symposium presented at the 100th Annual Convention of the American Psychological Association, Washington, DC.

Statman, M. A. (1993). Improving the effectiveness of employment testing through classification: Alternative methods of developing test composites for optimal job assignment and vocational counseling. *Dissertation Abstracts International, 54*(08), 4430B. (University Microfilms No. AAG94-03360).

Statman, M. A., Harris, D. A., McCloy, R. A., & Hogan, P. F. (1994, August). *Two approaches to classification: Maximizing performance vs. minimizing costs.* Paper presented at the 102nd Annual Convention of the American Psychological Association, Los Angeles, CA.

Taylor, R. (1987). Selecting effective courseware: Three fundamental instructional factors. *Contemporary Educational Psychology, 12,* 231-243.

Thompson, A. D., Simonson, M. R., & Hargrave, C. P. (1992). *Educational technology: A review of the research.* Washington, DC: Association for Educational Communications and Technology.

Thorndike, R. L. (1950). The problem of classification of personnel. *Psychometrika, 15,* 215-235.

Ward, J. H., Jr. (1983). Strategies for capitalizing on individual differences in military personnel systems. In R. C. Sorenson (Ed.), *Human individual differences in military systems.* (NPRDC SR 83-30). San Diego, CA: U.S. Navy Personnel Research and Development Center.

Welsh, J. R., Kucinkas, S. K., & Curran, L. T. (1990). *Armed Services Vocational Aptitude Battery (ASVAB): Integrative review of validity studies* (AFHRL-TR-90-22). Brooks AFB, TX: Air Force Human Resources Laboratory.

Whetzel, D. L. (1990). *Multidimensional screening: Comparison of a single-stage personnel selection/classification process with alternative strategies.* Unpublished doctoral dissertation, Washington, DC: George Washington University.

Wise, L. L., McHenry, J., & Campbell, J. P. (1990). Identifying optimal predictor composites and testing for generalizability across jobs and performance factors. *Personnel Psychology, 43,* 355-366.

Wolfe, J. H. (Ed.). (1997). Enhanced Computer-Administered Test (ECAT) Battery [Special Issue]. *Military Psychology, 9* (1).

Young, M. C., & White, L. A. (June, 1998). *Development of a measure to predict soldier attrition and motivation.* Paper presented at the 21st Army Science Conference, Norfolk, VA.

Zedeck, S., & Cascio, W. F. (1984). Psychological issues in personnel decisions. *Annual Review of Psychology, 35,* 461-518.

Zeidner, J., & Johnson, C. D. (1994). Is personnel classification a concept whose time has passed? In M. G. Rumsey, C. B. Walker, & J. H. Harris (Eds.), *Personnel selection and classification* (pp. 103-125). Hillsdale, NJ: Erlbaum.

Zeidner, J., Johnson, C. D., & Scholarios, D. (1997). Evaluating military selection and classification systems in the multiple job context. *Military Psychology, 9*(2), 169-186.

APPENDIX

Training Characteristics Survey

Training Characteristics Survey

January 1997

This survey has been developed under contract (F41624-95-C-5027) with the Air Force Armstrong Laboratory by the Human Resources Research Organization. The survey is being used to collect information about Air Force technical training. We are distributing it to course managers, instructors, curriculum chiefs, and training developers. This information is needed for research on the assignment of recruits to entry-level technical training courses.

APPROVED FOR PUBLIC RELEASE; DISTRIBUTION UNLIMITED

What's in This Survey?

The Training Characteristics Survey has five parts.

Part 1 requests brief information about you -- this information will only be used to group responses. Parts 2 through 5 ask for information about a particular training course.

Part 2 asks you to identify the Air Force Specialties associated with the training course.

Part 3 asks you to describe the methods of instruction used in the training course.

Part 4 asks questions about the difficulty of the training course.

Part 5 asks you to describe the content of the training course, specifically what kinds of activities must students do and what skills and abilities are needed.

General Instructions:

The purpose of this survey is to collect descriptive information about a sample of Air Force training courses. Specifically, we are interested in the characteristics of the training environment that differentiate courses from each other.

Some of the survey questions ask for subjective responses. We want your best estimates based on your experience in military training. There are no right or wrong answers. We are interested in your perceptions of the characteristics of the technical training environment.

Throughout this inventory we are concerned only with the course identified as:

[COURSE NUMBER AND TITLE]

-- .

*Do **not** consider other courses when answering.*

What You Should Do With The Completed Inventory:

After you finish the survey, place it in the pre-addressed envelope provided and put it in the mail to your base enlisted specialty training monitor. If you misplace the envelope, please return the survey to the following address:

[INSERT BASE ENLISTED SPECIALTY TRAINING MONITOR ADDRESS HERE]

Please return your survey within ten (10) days from the date you receive it.

Survey Monitor:

If you have any questions or comments about this survey, please call survey monitor name and phone number. Thank you very much for your participation.

Part 1: Background Information

1. What best describes your position?
   (*Mark one*)

   _____ course manager
   _____ instructor or trainer
   _____ curriculum chief
   _____ training developer
   _____ other (describe)
   _____

2. How many years of experience do you
   have in training development, research,
   or instruction?

   _____ years   _____ months

Part 2: Occupational Area

In this part of the survey, you will find questions about the occupational area associated with the training course. Only consider the course named above when answering questions.

3. Mark the Air Force Specialty Code(s)
   for which this course provides training:

   _____ xxxxx
   _____ xxxxx
   _____ xxxxx
   _____ xxxxx
   _____ others (list)
   _____

Part 3: Method of Instruction

In this part of the survey, you will find questions about the methods of instruction, media, and materials used in the course. Only consider the course named above when answering questions.

4. What percentage of course time is
   devoted to the media used in this
   course?
   (*Percentages should sum to 100*)

   Example:
   85%   face-to-face instruction
   15%   computer-based instruction (CBI)

   100%   TOTAL

   _____ face-to-face instruction
   _____ computer-based instruction (CBI)
   _____ interactive videodisc (IVD)
   _____ simulator
   _____ distance learning technology
   _____ other (describe) _____

   _____ TOTAL

5. What percentage of course time is
    devoted to the methods of instruction
    used in this course?
    (*Percentages should sum to 100*)

    Example:
    _70%_ lecture
    _0%_ discussion
    _30%_ instructional game

    _100%_ TOTAL

    _____ lecture
    _____ discussion
    _____ demonstration
    _____ hands-on performance
    _____ simulation
    _____ tutorial
    _____ drill and practice
    _____ instructional game
    _____ modeling
    _____ problem solving
    _____ other (describe)
    _____

    _____ TOTAL

6. How many hours of instruction are
    included in this course?

    _____ hours

7. How many blocks of instruction are
    in this course?

    _____ blocks

8. What is the student/teacher ratio (i.e.,
    average student flow per instructor
    for classroom course)?

    _____ student/teacher ratio

9. How many quizzes, tests, hands-on
    performance exercises, and other graded
    activities are included in this course?

    _____ number of tests, quizzes, etc.

10. How much verbal or written feedback,
    apart from tests and graded activities,
    do students typically receive during
    the course?
    (*Mark only one*)

    ___ 1-No feedback (until end of course)
    ___ 2-Very little feedback
    ___ 3-Some feedback
    ___ 4-A lot of feedback
    ___ 5-Very extensive feedback

11. Describe the learning environment.
    Students work mostly:
    (*Mark only one*)

    _____ individually
    _____ in small groups (2 to 3)
    _____ in moderate groups (4 to 9)
    _____ in large groups (10 or more)
    _____ in some combination of the above
          (describe) _____

12. Who usually controls the *pace* of the
    instruction (i.e., how quickly is material
    presented/learned)?
    (*Mark only one*)

    _____ instructor

    _____ students

13. Who usually controls the *sequence* of
    instruction (i.e., the order of lessons

    _____ instructor

4

or units)?
(*Mark only one*)

_____ students

14. How much flexibility or variability is permitted in the *pace* of the instruction? (*Mark only one*)

_____ 1-No variability
_____ 2-Slight variability
_____ 3-Moderate variability
_____ 4-High variability
_____ 5-Very high variability

15. How much flexibility or variability is permitted in the *sequence* of the instruction? (*Mark only one*)

_____ 1-No variability
_____ 2-Slight variability
_____ 3-Moderate variability
_____ 4-High variability
_____ 5-Very high variability

16. How structured is this course? (Structure is a function of the level of control assigned to the instructor [i.e., person or computer] as opposed to the student.) (*Mark only one*)

_____ 1-Completely structured
_____ 2-Somewhat structured, somewhat unstructured
_____ 3-Completely unstructured

Part 4: Course Difficulty

In this part of the inventory, you will find questions related to the difficulty of the course. Course difficulty is a subjective concept. Please give your best estimates based on your experience with military technical training. There are no right or wrong answers. Only consider the course named above when answering questions.

17. What is the average reading grade level of the course materials (e.g., lectures, books, study guides, workbooks, handouts, self-study materials, computerized text)?

_____ Reading grade level

18. What percentage of students require special individualized assistance from the instructor(s)? (*Give your best estimate*)

_____ percent

19. What percentage of students repeat
one or more blocks of this course after
failing quizzes or tests or due to poor
academic performance?
(Give your best estimate)   _____ percent

20. What percentage of students fail this
course based on academic
performance?
(Give your best estimate)   _____ percent

21. How much does this course emphasize
learning abstract concepts and
principles?
(Mark only one)

_____ 1-No emphasis
_____ 2-Slight emphasis
_____ 3-Moderate emphasis
_____ 4-High emphasis
_____ 5-Very high emphasis

22. How quickly is the instruction paced
(for example, in a very highly fast-paced
course, students learn a very large number
of facts, concepts, or procedures in a
very short amount of time)?
(Mark only one)

_____ 1-Not fast-paced
_____ 2-Slightly fast-paced
_____ 3-Moderately fast-paced
_____ 4-Highly fast-paced
_____ 5-Very highly fast-paced

23. How difficult or challenging is this
course? (Difficulty is a function of
the amount, complexity, or novelty
of information, and the pace of
instruction.)
(Mark only one)

_____ 1-Extremely easy
_____ 2-Somewhat easy
_____ 3-Neither easy nor difficult
_____ 4-Somewhat difficult
_____ 5-Extremely difficult

24. If you rated this course as somewhat or extremely difficult in question 23, please describe what
makes this course difficult.

_____

_____

_____

Part 5: Course Content

In this part of the inventory, you will find a list of characteristics that may describe activities required of the students (e.g., discussion, hands-on practice) or abilities and skills needed to learn the course material (e.g., speaking ability, problem solving). We would like you to tell us how important each characteristic is to this training course. Only consider the course named above when answering questions.

Use the following scale to describe the importance of each item:

NA = **Not applicable** (item is not related to the training course)
1 = **Not important** (item is associated with the course, but is not important)
2 = **Somewhat important**
3 = **Important** (item is an important characteristic/requirement of the course)
4 = **Very important**
5 = **Extremely important** (item is a critical characteristic of the course)

Circle only one response for each student activity or skill/ability.

Student Activities

	Not Applicable	Not Important	Somewhat Important	Important	Very Important	Extremely Important
25. Discussion between students and instructor	NA	1	2	3	4	5
26. Discussion among students	NA	1	2	3	4	5
27. Learning concepts and principles	NA	1	2	3	4	5
28. Learning facts	NA	1	2	3	4	5
29. Learning step-by-step procedures	NA	1	2	3	4	5
30. Hands-on performance	NA	1	2	3	4	5
31. Drill and practice	NA	1	2	3	4	5
32. Self study (out of class activities, not assigned reading)	NA	1	2	3	4	5
33. Outside reading assignments	NA	1	2	3	4	5

Skills and Abilities

	Not Applicable	Not Important	Somewhat Important	Important	Very Important	Extremely Important
34. Speaking	NA	1	2	3	4	5
35. Listening	NA	1	2	3	4	5
36. Writing	NA	1	2	3	4	5
37. Reading	NA	1	2	3	4	5
38. Mathematical ability	NA	1	2	3	4	5

7

	Not Applicable	Not Important	Somewhat Important	Important	Very Important	Extremely Important
39. Creativity or originality	NA	1	2	3	4	5
40. Spatial abilities	NA	1	2	3	4	5
41. Problem solving	NA	1	2	3	4	5
42. Troubleshooting	NA	1	2	3	4	5
43. Memorization of words, numbers, procedures	NA	1	2	3	4	5
44. Quickness/speed of performance	NA	1	2	3	4	5
45. Accuracy or precision of performance	NA	1	2	3	4	5
46. Knowledge of mechanical concepts	NA	1	2	3	4	5
47. Mechanical ability	NA	1	2	3	4	5
48. Electronics knowledge	NA	1	2	3	4	5
49. Knowledge of cars (parts and how they work)	NA	1	2	3	4	5
50. Knowledge of shop equipment and procedures	NA	1	2	3	4	5
51. Hand-eye coordination	NA	1	2	3	4	5
52. Interpersonal interaction	NA	1	2	3	4	5

Thank you for completing this survey.