THE UTILITY OF THE INTERACTION OF APTITUDE WITH INSTRUCTIONAL SUPPORT IN TRAINING

By

Ellen Marie Whitener

A DISSERTATION

Submitted to
Michigan State University
in partial fulfillment of the requirements
for the degree of

DOCTOR OF PHILOSOPHY

Department of Management

1988

ABSTRACT

THE UTILITY OF THE INTERACTION OF APTITUDE WITH INSTRUCTIONAL SUPPORT IN TRAINING

by

Ellen Marie Whitener

Cronbach and Gleser's suggestion that performance could be increased if organizations adapted their training treatments to match individuals' knowledge, skill or ability levels rests on the hypothesis that aptitudes interact with treatments to influence performance. This assumption is tested in two ways. First, a meta-analysis of education studies yielded results that were consistent with this hypothesis for the interaction of ability and instructional support treatments. Second, a study assessed the effect on learning for 118 subjects of the interaction of cognitive ability and instructional support treatments in training for tax preparation. This interaction was statistically significant (standardized regression coefficient = .19, p=.043). In addition, the nature of the interaction was ordinal such that while all individuals performed better in the high rather than low instructional support treatment, those with high ability enhanced their performance even more than those with low ability. Future researchers need to investigate the influence of different types of instructional support and abilities, study the effect of interactions on job performance and estimate the utility of developing programs that adapt to individuals' aptitudes.

DEDICATION

To Smokey and Spunkey

and

in Memory of

Taddybogle, Twinklebell, Grape, Powwacket,

Shadow, Moccasins, Rumble Bum Ditty and Nubbin

ACKNOWLEDGEMENTS

Many thanks go to a number of individuals for their input and support on this project. The most thanks go to my committee. John Hollenbeck took a risk on me four years ago, when we were both rookies at Michigan State. I have benefitted tremendously from working with him as he has taught me about research, mentoring, professionalism, and keeping a sense of humor. Dan Ilgen supported me, by giving me the opportunity to spend essentially all of my time finishing my dissertation and by providing all the funds for this project. Georgia Chao helped me by making sure I "sweated the details"--from typos to integrity of treatments. Many thanks.

I am indebted to Curt Hildorf of H & R Block for providing the instructional materials for the training treatments and to Valetta Ball for teaching the high instructional support training class.

I also am grateful to Anne O'Leary, Parshotam Dass and Jon and Barb Werner for helping with data collection and to the TAs of Management 310 and 302 for encouraging their students to participate in the research project by providing extra credit points.

And finally, words cannot express the deep appreciation to my family and friends for their encouragement, support and belief in my eventual success. You are truly special, and I am truly blessed.

TABLE OF CONTENTS

LIST OF TABLES..... ix

LIST OF FIGURES..... xi

CHAPTER 1: INTRODUCTION

Problem Overview..... 1

CHAPTER 2: A META-ANALYTIC REVIEW OF THE EFFECT OF THE INTERACTION BETWEEN ABILITY AND INSTRUCTIONAL DESIGN ON EDUCATIONAL ACHIEVEMENT

Introduction..... 10
Foundations of the ATI Paradigm..... 13
Criticisms of the ATI Paradigm..... 17
Empirical Evidence of ATI..... 19
Conceptual and Methodological Problems with Research on ATI..... 20
Method..... 23
Meta-Analysis..... 23
Meta-Analysis in general..... 23
Meta-Analysis with standardized regression weights..... 25
Data Collection..... 29
Results..... 33
Statistical Artifacts..... 33
Potential Moderators..... 43
Discussion..... 47

CHAPTER 3: LITERATURE REVIEW OF ADAPTATION WITH TRAINING

Overview..... 55
Fixed-treatment Selection in Training..... 56
Selection Based on Samples of Training Tasks..... 61
Trainability..... 61
Early Training Success/Time in Training..... 63
Aptitude-treatment Interaction in Training..... 64
Interactions between Ability and Training..... 65
Prior achievement..... 65
General ability..... 69
Interactions between Other Aptitudes and Training..... 71
Summary..... 75

TABLE OF CONTENTS (CONT'D)

CHAPTER 4: METHOD

Overview 83
Subjects and Tasks 84
Subjects 84
Task 85
Variables 90
Tests 90
Verbal comprehension and numerical ability ... 90
General intelligence 92
Aptitudes 94
Performance 95
Reaction to Training 96
Instruction time and time spent studying 96
Procedure 99
Analysis 100

CHAPTER 5: RESULTS

Description of Sample 108
Hypotheses 113
Moderated Regression 113
Hypothesis 1 118
Hypotheses 1a and 1b 118
Hypothesis 2 119
Hypothesis 3 121
Hypothesis 4 122
Summary 124

CHAPTER 6: DISCUSSION

Summary and Implications of Results 126
Hypothetical Utilities 131
Application of Results 133
Comparison of Meta-analytic Results and Empirical Results 135
Effect Sizes 135
Nature of Instructional Support 135
General and Specific Ability 137
Summary 139
A Taxonomy of Instructional Support 141
Alternative Explanations of Results 145
Power 145
Selection 146
Restriction of Range 146
Manipulation Failure 148
Generalizability of Results 149
Future Research 151
Conclusion 153

TABLE OF CONTENTS (CONT'D)

APPENDICES

Appendix A - Decision-Theoretic Foundations of Adaptive Treatment
Decision theory and personnel decisions.....154
Types of personnel decisions.....154
Characteristics of decision processes.....155
Evaluation of personnel decisions.....160
Evaluation of predictive utility of test score...161
Adapting Training Treatments in Selection Decisions.....164
Fixed- and adaptive treatment in selection decisions.....164
Fixed-treatment selection.....166
Adaptive-treatment selection.....169

Appendix B - Formulae and Procedures used in Meta-analysis.....177

Appendix C - Measures.....178

LIST OF REFERENCES 209

LIST OF TABLES

Table 1: Results of Meta-analysis of Standardized Regression Coefficients..... 34

Table 2: Results for Incremental Variance associated with ATI..... 36

Table 3: Data and Summary for Error of Measurement. 39

Table 4: Results of Moderator Analyses..... 44

Table 5: Empirical Studies of Aptitudes Predicting Training Success..... 57

Table 6: Descriptive Statistics for Teacher Behaviors Inventory..... 88

Table 7: Results of Principle Components Factor Analysis of Aptitude Tests..... 94

Table 8: Basic Demographics of Evaluation of Instructional Materials..... 97

Table 9: Intercorrelations for Scales and Variables.....101

Table 10: Comparison of Subjects for Differential Attrition.....108

Table 11: Descriptive Statistics for Whole Sample.....110

Table 12: Descriptive Statistics by Treatment.....112

Table 13: Moderated Regression on Learning for General Ability Factor Controlling for Hours.....114

Table 14: Moderated Regression on Learning for General Ability Factor Not Controlling for Hours.....114

Table 15: Moderated Regression on Learning for Numerical Ability Factor Controlling for Hours.....115

Table 16: Moderated Regression on Learning for Numerical Ability Factor Not Controlling for Hours.....115

LIST OF TABLES (CONT'D)

Table 17: Moderated Regression on Learning for Verbal Ability Factor Controlling for Hours.....116

Table 18: Moderated Regression on Learning for Verbal Ability Factor Not Controlling for Hours.....116

Table 19: Full Regression Equations for Learning with Ability Factors.....117

Table 20: Coefficients of Determination for Regression Equations.....119

Table 21: Standardized Regression Coefficients for Interaction Terms.....121

Table 22: Full Regression Equations for Learning with Ability Tests.....123

Table 23: Hypothetical Differences in Utility.....132

LIST OF FIGURES

Figure 1: Graph of a Disordinal Interaction..... 11

Figure 2: Graph of an Ordinal Interaction..... 11

Figure 3: Ordinal Interaction (James, 1962)..... 70

Figure 4a: Expected Disordinal Interaction for Remedial and Compensatory Approaches..... 79

Figure 4b: Expected Ordinal Interaction for Preferential Approach..... 79

Figure 5: Graph of Ordinal Interaction of General Ability Factor and Treatment on Learning.....120

Figure 6. Hypothesized Interactions of Ability with Different Dimensions of Instructional Support.....143

Figure 7: Personnel Decision Problems.....156

Figure 8: Schematic View of Decision Process for Personnel Decisions.....158

Figure 9: Payoff functions for Different Treatments.....163

Figure 10: Selection Decision Process.....165

Figure 11: Adaptive Treatment Selection Decision Problem.....170

Figure 12a: Expected Payoff as a Function of Aptitude and Treatment.....172

Figure 12b: Expected Payoff and Treatment for Aptitude Level S_1.....172

Figure 12c: Expected Payoff and Treatment for Aptitude Level S_2.....175

Figure 12d: Aptitude-Treatment Interaction for aptitude S.....175

CHAPTER 1: INTRODUCTION

Problem Overview

Demographic and employment security changes in today's labor force have the potential to influence industry's selection and training methods in the long run. Roger D. Semerad, U.S. Assistant Secretary of Labor for Employment and Training, described the major demographic changes that are expected to influence the labor market in the next 15 years (Semerad, 1987). An imbalance in the demand and supply for labor is predicted. By the year 2000 demand for white collar, professional and technical workers will have increased while demand for skilled and unskilled production workers will have decreased. The labor supply, however, will be seriously inadequate to meet this type of demand, primarily because of an overall decline in the birth rate. Since fewer individuals are being born, there will be fewer individuals available for the work force.

Growth in the size of the labor force has already slowed and will continue to decrease. The work force will be older and perhaps more reliable and less adaptable to the rapid changes in the economy. In addition, because of the relatively higher birth rate in disadvantaged households, it is projected that more workers may be functionally illiterate, at the same time that businesses are eliminating positions that these workers can hold. While the

standard for workplace literacy is currently a 9th grade education; the standard for 2000 is expected to be two years of college. The Department of Labor is projecting, therefore, that there will be a shortage of professional/technical workers and a surplus of unskilled/low skilled workers.

Thus, businesses with a demand for highly skilled workers may face a shrinking applicant pool. These businesses may be forced into hiring large numbers of workers with a variety of skills and aptitudes who will need extensive training or retraining provided by the organization. The type of training individuals need in order to perform their jobs successfully will depend on the knowledge, skills and abilities or, more generally, on the aptitudes these individuals possess.

Not only will organizations face large applicant pools that need specialized or differential treatment, they may also face large pools of unskilled/low skilled workers already employed with their companies who need retraining. This pool of employees will be generated because unions have been negotiating job or employment security clauses into their contracts, protecting workers' jobs after plant closings and preventing layoffs (Schlesinger, 1987; "UAW's Bieber," 1987). These workers may also possess inadequate skills or preparation for the new jobs in which they will be placed. Organizations will need to treat (i.e., train or reassign) these workers in different ways depending on their aptitudes.

Organizations need to plan to deal with the effects of these changes. Semerad points out that "the nation has almost no choice

but to work with the people who will be available, some of whom are now ill-prepared for the workforce of 2000. None of the transitions will take place by chance. They will require the most sophisticated strategic planning and resolute implementation to educate, train, retrain, and more important, to instill hope and the necessary work ethic" (1987, p. 40).

These demographic and employment changes suggest that organizations will need to provide different treatments for different workers. Indeed, it makes common sense to treat workers differently in industrial training when they possess a variety of aptitudes, just as people with different aptitudes and skills are treated differently in sports training or education.

Despite the sensibleness of this suggestion, it has rarely been put into practice (Cronbach and Gleser, 1965). In the past, organizations have developed selection and training programs assuming what Cronbach and Gleser refer to as fixed-treatment conditions in which "...the treatment is specified a priori, without regard to information from the particular persons tested" (1965, p. 29). Individuals are selected on the basis of some information about their aptitudes and then experience the same treatment (i.e., training program or job assignment) without regard for their different knowledge, skill or ability levels.

Such practices ignore several critical issues which suggest that a continued reliance on fixed treatment is inappropriate. First, the goal of maximizing the benefit to individuals which occurs when assignment to treatments is made on the basis of

individuals' aptitudes is not incompatible with industry's goal of institutional benefit. Generally, all other things being equal, when individuals' job performance increases, the organization benefits; indeed management always seems to be searching for techniques such as goal setting, merit pay, or piece rate pay to increase workers' performance levels. Adapting treatments to workers' aptitudes, perhaps through different job assignments, modified job characteristics or different training techniques, provides another viable way to increase individual performance levels and, ultimately, to benefit the organization.

Second, the practices based on fixed treatment assumptions fail to take into account the fact that there is a great deal of variance in performance among those who are identified as successful employees. As Cronbach and Gleser (1965) point out, the focus on satisfactoriness of performance in evaluating the utility of selection (e.g., H. C. Taylor and Russell, 1939) has failed to recognize that there may be wide differences in quality of service among those individuals who are counted as successful (as hits). If the variance of performance among selected individuals is great, then the benefit to the organization or utility may be further enhanced by reducing the variance and raising the performance of employees through adapting treatments to the aptitudes of individuals.

Third, many organizations already have to hire a large proportion of their applicants. As the model proposed by H. C. Taylor and Russell (1939) indicates, a high selection ratio (the

proportion of employees hired from the applicant pool) reduces the success ratio (that proportion of selected individuals who subsequently perform satisfactorily) for any test with even some predictive value. These organizations with high selection ratios may be able to increase their success ratios by testing their applicants and using the test results to adapt their treatments to fit applicants' aptitudes. In addition, organizations facing high selection ratios may be incurring high selection costs with little performance benefit. The testing cost/performance benefit ratio could be reduced if organizations could lower their selection ratios to the optimum (.50) suggested by Brogden (1949). One way organizations could lower their selection ratios is by increasing recruiting efforts to generate more applicants. They could then be more selective in their hiring, reducing the selection ratio. However, because of labor market conditions or economic conditions, many organizations cannot affect significant change on their selection ratios or subsequent mean performance levels. When this is the case, organizations may choose to place new employees in different treatments designed to fit their special skills and needs, and ultimately, to provide greater benefits for future performance on the job.

The suggestion that individuals should be assigned to treatments on the basis of their aptitudes which are defined broadly as "any characteristic of a person that forecasts his [or her] probability of success under a given treatment" (Cronbach & Snow, 1977, p. 6) is what Cronbach and Gleser defined adaptive-treatment.

"In adaptive treatment,...the choice of... [employee]... and choice of treatment are both made in the light of test data....Speaking generally, the aim of adaptive treatment is to choose the treatment for an accepted person in order to fit his aptitudes as well as possible..." (p. 29). Cronbach and Gleser argued that such adaptation may have greater value than has previously been assumed. They indicated that benefit to the organization will be increased through enhancing individuals' job performance by offering different treatments which are varied through matching the aptitudes of tested and selected individuals. [Details of Cronbach and Gleser's (1965) theories and arguments are given in Appendix A.]

However, if adaptive treatment is going to benefit the organization, there must be an interaction between individuals' aptitudes and the treatment that they experience. For example, high ability individuals may be able to perform a job successfully through on-the-job training; whereas, individuals with relatively low ability may not perform the job successfully until they have also had some classroom training.

Such an aptitude by treatment interaction (ATI) has been studied and applied extensively in education (Cronbach & Snow, 1977) because education's primary frame of reference is to maximize the benefit for the individual. ATIs have been studied with the goal of finding instructional treatments that will bring the greatest performance (i.e., learning) for each individual. In industry, however, ATIs have not been studied to the same extent because industry's frame of reference is essentially reversed. Where

education has focused on the benefit or performance gain for the individual, industry has focused on the benefit to the institution with relatively less concern for the individual's development. Industrial organizations tried to get the most out of people given a treatment designed to maximize the performance of the average person. If non-average individuals did not perform sufficiently well given that treatment, then they were terminated and the treatment was tried on new individuals.

However, as discussed above, individual benefits certainly are not incompatible with institutional benefits. Individuals and institutions both benefit under adaptive treatment if two conditions hold. First, individuals' aptitudes and different treatments must interact such that one treatment enhances performance for low aptitude individuals and another treatment enhances performance for high aptitude individuals. Individuals can reap the benefits associated with being able to perform their jobs more successfully (e.g., increased compensation, promotion and other rewards). The institution benefits if the mean performance of their workers increases. Second, the value of the performance increase to the organization must exceed the costs of providing different treatments. The utility of organizational programs is reflected in the balance of benefits to costs. Increased performance must lead to greater productivity or decreased costs that outweigh the costs of getting that increased performance.

In summary, fixed treatment has been the standard paradigm in determining the benefit or utility of personnel programs, and

particularly of selection. However, for the reasons noted above, fixed treatment assumptions may be inadequate to deal with the changing conditions. Industry needs to increase consideration and investigation of the utility of adapting treatments to individuals' aptitudes. Where the focus on institutional benefit has been the excuse for failing to consider individual benefit in the past, it has now become the driving force to seek how to increase individual benefit or performance and institutional benefit.

This research project investigates several aspects of adapting a training program to individuals' aptitudes. The investigation focuses on the two issues described above--ascertaining whether there is an ATI and estimating the balance of costs to benefits.

The utility of adaptive treatment within the industrial setting is developed and investigated in this research project in the following manner. Because most of the empirical work on adaptive treatments or ATI has been conducted in educational settings, the educational literature is reviewed for evidence of the nature and strength of ATIs (Chapter 2). Even advocates of ATIs such as Cronbach and Snow (1977) have admitted that the evidence is weak. This lack of support for ATIs, may be due to inadequate tests of the model. However, meta-analysis, a quantitative approach to summarizing empirical results and correcting for statistical artifacts that weaken results, can be combined with a critical literature review to evaluate the strength and nature of ATIs in the educational context. In Chapter 3, the sparse literature that

investigates adaptive treatment or ATIs in training settings is reviewed. Several reviewers of the training literature have called for more research on ATIs or adaptive treatment (e.g., Baldwin & Ford, 1988; Goldstein, 1980; Wexley, 1984); however, few studies have directly tested their effect. Rather, a number of studies have indirectly tested aspects of ATI by assessing the use of aptitudes as predictors of training success. Finally, a few studies have tested aptitude-treatment interactions in military settings. These studies, as well as findings from education, provide insight into how training might be adapted in industry. In Chapter 4, the procedures followed in the current project are described. The results of the current investigation are given in Chapter 5 and are discussed in Chapter 6.

The utility of adaptive treatment in industry rests on whether aptitudes and treatments interact in determining performance. This research project attempts to assess that utility in one situation by determining the strength of an ATI and by estimating the benefits and costs of different treatment strategies. The contribution of this line of research may be to enhance the productivity of workers, to increase consideration of the individual employee and to evaluate the linkage between the selection and training personnel functions.

CHAPTER 2:

A META-ANALYTIC REVIEW OF THE EFFECT OF THE INTERACTION BETWEEN ABILITY AND INSTRUCTIONAL DESIGN ON EDUCATIONAL ACHIEVEMENT

Educators have identified meeting individual learners' needs as a major goal and have focused on individualized instruction as a primary means to accomplish that goal (Carrier & McNergney, 1979). Individualized instruction relies on the theory that ATIs exist, that is, that learning is not simply a function of learners' aptitudes or of educational treatments, but of both interacting to influence learning. Aptitudes and treatments interact in two ways which may influence how instruction is adapted to individual learners' needs.

The first type of ATI, a disordinal interaction, as shown in Figure 1, indicates that one treatment yields high achievement for individuals at one end of the aptitude continuum whereas a different treatment yields high achievement for individuals at the other end. Achievement would be maximized, then, if the individuals at each end of the aptitude continuum receive different instructional treatments. The second type of ATI, an "ordinal" interaction, shown in Figure 2, indicates that one treatment is better than another treatment regardless of aptitude, but the magnitude of the difference varies at different aptitude levels. If the treatment that consistently yields higher achievement is more expensive than the other treatment, then the more expensive treatment would be used

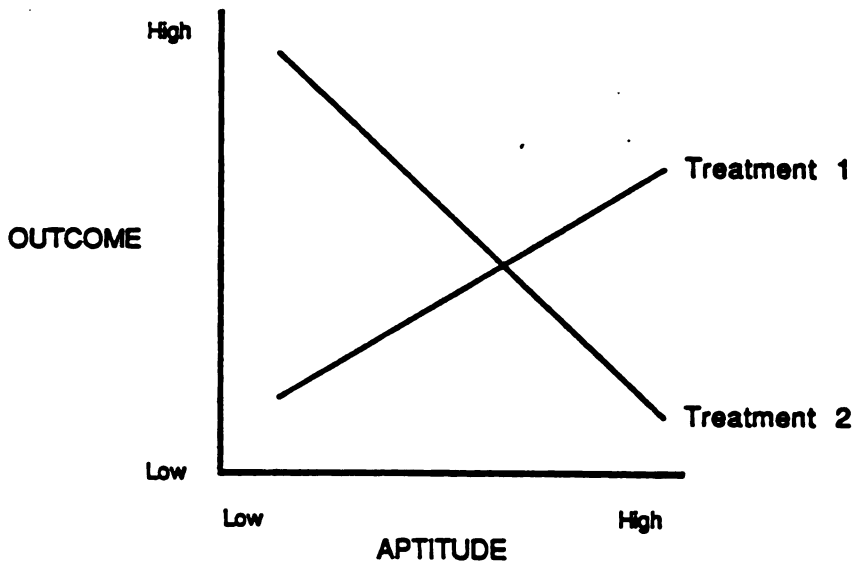


Figure 1. Graph of a Disordinal Interaction.

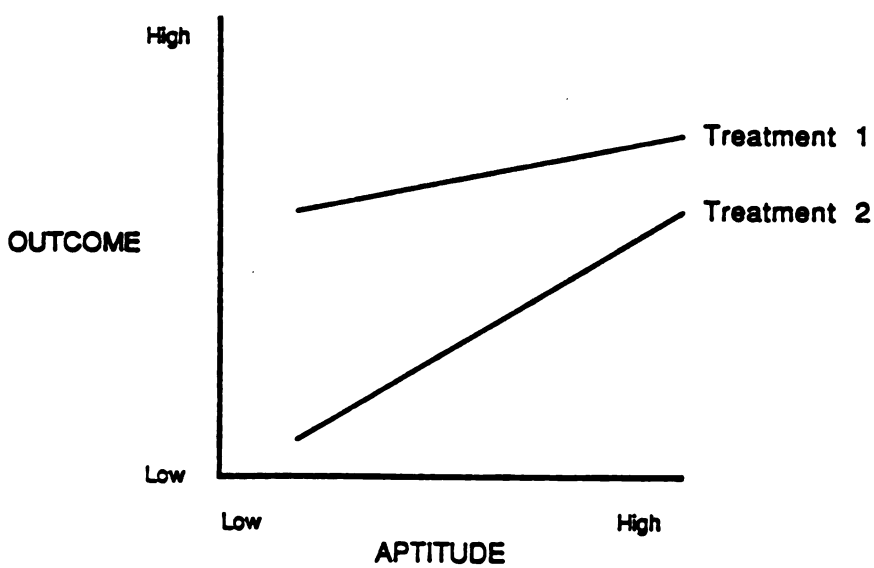


Figure 2. Graph of an Ordinal Interaction.

only for those individuals who benefit from it the most. Those individuals who achieve at approximately the same level regardless of treatment would experience the less expensive instructional treatment.

Although there is agreement that learners' needs should be met, there has been an active debate over the existence and practicality of ATIs. Researchers who argue against ATIs focus on either one of two issues. First, some argue that the empirical evidence is inconsistent because ATIs probably do not exist (Bracht, 1970). Second, others concede that ATIs exist, but argue that designing different treatments for different individuals is impractical or inappropriate (Gehlbach, 1979). On the other hand, advocates of ATIs discount the relatively unsupportive empirical research for not being sophisticated or theoretically sound (e.g., Berliner & Cahen, 1973; Cronbach & Snow, 1977; Snow & Lohman, 1984). These arguments are explained in more detail below.

Underlying these arguments are the critical questions regarding the existence and strength of ATI effects. Individual studies testing the existence of ATIs have contained methodological and conceptual difficulties that inhibit interpretation of the studies individually and of reviews of collections of the studies (Cronbach & Snow, 1977). Kulik (1981) suggested that some of these difficulties may be overcome through the use of meta-analytic techniques of reviewing related groups of ATI studies. For example, meta-analysis procedures devised by Hunter & Schmidt (Hunter, Schmidt & Jackson, 1982) can statistically correct for some

methodological problems such as insufficient sample sizes and error of measurement. In addition, focusing a meta-analysis on a specific subset of ATI studies can control for some of the conceptual difficulties. Meta-analysis, as a method of quantitatively summarizing a body of literature, can therefore be used to estimate the strength of an ATI effect.

The purpose of this review is to use meta-analytic procedures to test for the presence and magnitude of an ATI effect, in which the aptitude construct is ability, the treatments are a form of instructional method, and the outcome construct is achievement. In addition, to place this hypothesis in context, this review will include some of the theoretical and empirical arguments, as well as methodological difficulties related to the ATI paradigm.

Foundations of the ATI Paradigm

Three perspectives, common sense, science, and research methodology, have provided the foundation on which the aptitude-treatment interaction paradigm has been built. First, common sense has identified the basic elements and relationships of ATI which are rooted in the truism that different people respond differently in different situations (Solomon, 1980). As Ragsdale (1980) noted: "it is certainly intuitively appealing that there should be 'different strokes for different folks' in instruction as well as in golf and tennis" (p. 44). Thus, ATI constitutes a specific application of this truism in which abilities or aptitudes differentiate among people, instructional methods characterize the different situations, and measures of learning or achievement

constitute the responses. Intuition or common sense suggests that people of different abilities will achieve or learn different amounts under different instructional methods.

This common sense notion has motivated a number of researchers to be more scientific and theory-based in their attempts to build and refine the ATI paradigm. Some ATI theories have been built on general person-environment interaction theories that have been refined in other psychological disciplines. For example, in personality theory, Lewin (1935) proposed that the interaction between the person and the environment influenced behavior through subjective evaluations or perceptions of the environment. Similarly, in ATI research, Snow (1978) proposed that "individual differences in performance on ability tests and learning tasks are manifestations of cognitive processes common to each" (p. 227). Snow suggests that individuals will achieve more when the learning method matches or capitalizes on the cognitive processes underlying abilities and will achieve less when the learning method requires use of cognitive processes that are weak or absent.

Second, ATI research has also been linked with scientific theories or models of instruction. For example, Salomon (1972) identified a major function of ATI research as helping "...to develop better explanatory principles concerning the nature of instruction" (p. 328). He also contributed to this end by proposing that there are three complementary models of how achievement is enhanced by designing instruction for the individual. These models suggest three different, though not necessarily competing,

explanations for why different instructional treatments might yield different performance levels for different people. The first model, the remedial approach, involves treatments that compensate for individuals' learning deficiencies, particularly insufficient information. This model is based on the assumption that individuals will not be able to progress in their learning until they have mastered some important background information. Individuals in remedial programs receive extra instruction in the background information that other individuals already know. The second model, the compensatory approach, involves treatments that are designed to provide instructional support for learners who cannot provide it for themselves. For example, highly structured learning programs provide organization and support for poor learners; whereas, good learners can provide this organization and support for themselves and can even be inhibited in their learning by such programs. The last model, the preferential approach, is similar to Snow's (1978) theory in that it suggests that treatments can be designed to capitalize on what the student already knows or is capable of doing. Treatments that match the learner's preferred learning style or information processing strategy will lead to better performance.

Third, testing such theories as Salomon's or Snow's requires a focus on research methodology assessing interactions of persons and environments. Traditionally social science research has concentrated on the main effects of person or environment (Cronbach, 1957; Solomon, 1980); however, relatively recent changes in emphasis in research design and methodology have focused more directly on

interactions. Cronbach (1957) is generally credited with being the first to challenge researchers to change their research strategies to test the joint effects of persons and environments (Hunt, 1975). In addition he suggested the aptitude-treatment interaction paradigm as an appropriate replacement strategy. He proposed ATI as a means to use the best of the two dominant research perspectives, experimentalism and correlationism, which utilized competing methods to attain the same goal of increasing average performance. As Cronbach pointed out:

> The program of applied experimental psychology is to modify treatments so as to obtain the highest average performance when all persons are treated alike--a search, that is, for the "one best way." The program of applied correlational psychology is to raise average performance by treating persons differently--different job assignments, different therapies, different disciplinary methods. The correlationist is utterly antagonistic to a doctrine of "the one best way..." (p. 678).

He based his recommendation of ATI on evolutionary theory noting that "if for each environment there is a best organism, for every organism there is a best environment" (p. 679). Using experimental and correlational methods together and considering treatments and persons simultaneously, researchers could then identify the best "environment" for each "organism." Cronbach did not detail a theoretical position as much as he proposed a research strategy. The result of this is that investigators have responded to his challenge by studying a number of ATIs in diverse areas, such as social behavior and drugs and therapy as well as education (Cronbach, 1975).

Criticisms of the ATI paradigm

Some researchers have generally accepted the idea that ATIs exist, but question whether the ATI paradigm can be applied to a theory of instruction. They base their challenges on two arguments. The first argument is that ATIs have limited practical scope because they may affect only a small number of individuals for only a short period of time. Millman (1974) argued that teaching should not be adapted to student characteristics because such adaptation only benefits a few individuals. He pointed out that in a disordinal interaction, ATIs only apply to a few cases at the extremes of the normal distribution of aptitudes. However, the vast majority of cases are in the middle where there are small differences between the outcomes of the different treatments. It could be impractical to adapt treatments to students since so few might benefit. In addition, since aptitudes change over the course of a treatment, assignment based on an ATI may only be valid for a limited period of time (Burns, 1980). In this case, aptitudes would have to be re-measured periodically and individuals would have to be reassigned to a different treatment if their aptitudes had changed. This requires a great deal of organization and precision that may be difficult or impractical to attain. Both of these issues, the small number of beneficiaries and the dynamic nature of aptitudes, imply that the cost of adapting instructional treatments to individuals may outweigh the potential benefits.

The second argument against applying ATIs to instruction focuses on undesirable side effects of adapting treatments to meet

individual needs. One undesirable side effect would be that students could become overly dependent and expect environments or situations to change in order to meet their individual needs. Merrill (1975) proposed that the responsibility of adaptation should be on the student, not on the environment or institution. He felt students should select an educational environment and learn to adapt to the situation since most environments outside of school will not adapt to meet the individuals' needs. A second undesirable side effect might be that students become categorized and labelled. Hunt (1975) pointed out that some educators fear that if treatments are differentiated according to aptitudes that this could lead to labels which acquire a value judgment. For example, persons with X level of Aptitude A who encounter Treatment T1 are considered better than those with Y level who get T2. These undesirable side effects of an ATI-based program may be serious enough to discount other, more positive outcomes.

These arguments suggest the need for two lines of research. First, research studies need to ascertain whether ATIs exist and to estimate the strength of the interaction effect. Second, as Cronbach and Snow (1977) also noted, the outcomes of programs based on ATIs need to be evaluated for their practical value, contribution to instruction and/or costs and benefits. Since Cronbach's (1957) paper a large number of studies have focused on the first goal of determining the existence of an interaction effect, but virtually none have pursued the second.

Empirical evidence of ATI

Several reviewers have collected and summarized the research evidence that has accumulated on ATI. Most of them have concluded that the research has tended to yield conflicting or nonsupportive results. One of the first to summarize the ATI literature was Bracht (1970) who systematically analyzed 90 ATI studies and found only 5 provided evidence of a disordinal interaction. The other 85 indicated either no interaction or ordinal interactions (which he considered to have no practical value). Based on this evidence he concluded: "Although there is an increasing interest in the topic of ATI among educational psychologists, very little empirical evidence has been provided to support the concept" (p. 627).

Other reviewers have come to somewhat similar conclusions. Berliner and Cahen (1973) summarized ATI (what they referred to as Trait-Treatment Interactions or TTI) research related to instruction. Although they found somewhat greater evidence of the TTI concept, they also pointed out that support was minimal. One reason they found more support is that they took a less conservative stance than Bracht as to what constituted evidence of an interaction by considering ordinal interactions along with disordinal interactions as evidence of TTI. They summarized their review:

> In general, significant interactions are not a rare occurrence, and interactions [both ordinal and disordinal] have important implications for the design of instructional treatments. Lest an overly optimistic view of the present status of TTI research be conveyed, we hasten to point out the many cases where hypotheses about interactions were not

> confirmed and where findings of interaction were contrary to the hypotheses that guided the study. In addition, most studies of interaction have not been replicated; when replicated, interactions have not been confirmed (pp. 84-85).

Other reviewers have found a number of studies that indicated significant interactions but have concluded that none of these interactions are consistent or strong enough to be used to direct instructional design. Cronbach and Snow (1977) reviewed a large variety and number of ATI studies and concluded that "no Aptitude X Treatment interactions are so well confirmed that they can be used directly as guides to instruction" (p. 492). In a more recent summary, Pintrich, Cross, Kozma & McKeachie (1986) stated that ATI research findings cannot be used with any confidence to construct general principles of instructional design.

At best, then, the research evidence for ATIs is conflicting and inconsistent. A number of explanations for this inconsistency exist. The two most probable, methodological problems in the design of ATI studies and sampling error are discussed below.

Conceptual and Methodological Problems with Research on ATIs

Many researchers are optimistic that ATIs exist and have a contribution to make to education (e.g., Berliner & Cahen, 1973; Cronbach & Snow, 1977). The major reason for their optimism is that they feel that methodological problems provide alternative explanations for the inconsistent results. For example, a typical ATI study might use a small sample, manipulate two treatments, subgroup on an aptitude by a median split, and disregard the impact of error of measurement and other procedural issues. As Cronbach

and Snow concluded after their reanalysis and summary, "many reports (of both positive and negative results) must be discounted because of poor procedures" (1977, p.6).

Cronbach and Snow identified a number of such problems and detailed what procedures are appropriate to test for ATIs and what information should be included in reports of those studies to help in future reviews of ATI findings. For example, they advocated the use of multiple regression to avoid subgrouping or blocking on the aptitude variable, often as a median split, in the use of analysis of variance procedures. They noted that subgrouping often leads to nonsignificant results when the full range would yield a significant ATI. In addition, multiple regression can facilitate interpretation of the interaction. For example, if equations are run for two treatments separately, regressing the dependent variable on the independent aptitude variable, then the output will consist of two regression equations, $y_{T1} = a_{T1} + b_{T1}x$ for Treatment 1 and $y_{T2} = a_{T2} + b_{T2}x$ for Treatment 2 where a is the y intercept and b is the regression weight on the aptitude variable of the slope of the equation. The difference between the regression weights ($b_{T1} - b_{T2}$) gives the difference in the slopes and thus an indication of the interaction effect. If moderated regression is used in which one multiple regression equation of the form, $y = a + b_1X + b_2T + b_3XT$, is computed and the treatments are entered as dummy variables, then the ATI effect is revealed in the regression weight on the aptitude-treatment interaction term (b_3). This regression weight would equal the difference between the regression weights on the aptitude for

the two treatments noted above ($b_{T1} - b_{T2} = b_3$). Cronbach and Snow noted that these regression weights should be unstandardized for comparisons across samples using the exact same set of variables. They argued, "standard deviations of predictors and criteria fluctuate from one sample to another, and consequently the standard score of +1... corresponds to a different 'raw'... score in each sample. Unstandardized weights have to be applied, in order to make the same prediction for persons in the two samples who have identical test performances" (p. 40).

Cronbach and Snow also called for consideration and correction for unreliability in the measurements of aptitudes and outcomes since unreliability in the aptitude measure can bias the nature of the interaction and unreliability in the outcome can reduce the probability of finding significant results. In addition, they pointed out that sample sizes for many studies were woefully inadequate to detect significant differences for interaction effects. They recommended that generally a sufficient sample size would have to include at least 100 subjects per treatment. Finally, one of their chief concerns had to do with the appropriate use of statistical techniques. They lamented "selective reporting of only those effects that appear to be statistically significant and suppression of studies with no significant result, testing of multitudinous hypotheses in some studies, [and] calculating in terms of the number of individuals when the actual sampling or treatment unit was the class." (p. 58).

For these and other reasons, single studies have not provided definitive evidence of the existence of ATIs. This necessitates consideration of the results of sets of studies. Kulik suggested the use of meta-analysis to accomplish this task: "I believe that in the long run we will learn a great deal about ATIs from meta-analysis. That is, to find out which treatments are most effective for which people, we should not only look within studies..., we should also look across collections of studies....If we use this strategy, I think that we will overcome the shortcomings of traditional ATI research. First, we'll get consistent results; second we'll get stimulating results; and third, we'll get results that we can use." (Kulik, 1981, p. 2).

Therefore, although empirical research and prior reviews of the ATI literature have proved inconclusive, recent developments in quantitative review methods or meta-analysis make it possible to reexamine the literature. This review uses meta-analytic techniques to look across a collection of studies that look at the effect of the interaction of ability and supportive instructional practices on achievement. The major intent is to test whether this particular ATI exists and to estimate the strength of the ATI effect by looking at the difference between the standardized regression slopes.

Method

Meta-Analysis

Meta-Analysis in general. Meta-analysis has been used in a number of areas in education (Johnson, Maruyama, Johnson, Nelson & Skon, 1981; Kulik, Kulik & Cohen, 1979; Wise & Okey, 1983). Several

meta-analytic techniques have been developed and are appropriate for different purposes (Rosenthal, 1978). The meta-analytic procedures described by Hunter, Schmidt & Jackson (1982) were chosen for this review because they provide methods to aggregate effect sizes common to ATI studies and to correct for statistical artifacts in order to estimate true relationships.

The details of this meta-analysis technique have been described in Hunter et al (1982), therefore, only the general procedures are given here. After collection of studies and coding of data, a sample-size weighted average effect size and confidence interval are calculated. The residual variance or variance not explained by sampling error is derived by subtracting the error variance from the variance observed across the accumulated effect sizes. If the residual variance equals zero, then sampling error explained all of the variation in the effect sizes. Hunter et al (1982) also provided corrections for statistical artifacts other than sampling error. If sufficient information has been provided in the collection of studies, the residual variance can also be reduced by the variation in the studies due to error of measurement in the dependent and independent variables and due to restriction in range. Other statistical artifacts which can significantly affect results cannot be corrected; these include computational and typographical errors, measurement contamination or deficiency; and differences in factor structures among different measures of the same constructs. If variation still remains after correction for statistical artifacts, then it is appropriate to search for moderators, either

study characteristics or theoretical variables, of the relationship under study. Potential moderator variables are investigated by calculating and comparing effect sizes for subgroups of these variables. If these comparisons indicate differences among the subgroups for a variable, then that variable should be identified as a moderator. The search for moderators theoretically continues until the residual variance is fully explained.

Meta-Analysis with standardized regression weights. Most meta-analyses have accumulated main effects using the techniques described above. It is, thus, a different use of meta-analysis to accumulate interaction effects. As discussed above, the interaction is carried in the regression weight for the interaction term in a multiple regression equation in which both aptitude and treatment main effects and their interaction are entered. It would be appropriate to accumulate this unstandardized regression coefficient on the interaction if the aptitudes and criteria were the exact same variables across all studies. However, these variables are somewhat different from study to study and their values differ depending on the unit of measurement. Thus, to compare regression weights across studies, it is necessary to standardize the weights using the aptitude and outcome standard deviations. Besides facilitating comparisons, standardized regression weights also have an advantage over unstandardized regression weights in that they provide an estimate of the strength of the interaction effect on achievement.

As discussed above, the standardized regression weight is found in either of two ways--as the regression weight on the interaction

term in a moderated regression equation or as the difference between the regression weights on the aptitude terms in the two separate treatment equations. In the first, the standardized regression coefficient for the interaction term is provided in the studies. In these studies outcome has been regressed on aptitude, treatment and the cross product or interaction of aptitude and treatment. In the second, the standardized regression coefficient for the interaction is calculated by subtracting the standardized regression coefficient for aptitude in one treatment from the standardized regression coefficient for aptitude in the other treatment. In these studies, separate regression equations have been generated for each treatment in which outcome has been regressed on aptitude for each treatment separately.

Unfortunately, few studies actually report the standardized regression weights for the interaction. However, if the studies provide the unstandardized weight on aptitude and the standard deviations for outcome and aptitude for each treatment, then the standardized regression coefficients can be calculated. For each treatment, the unstandardized regression coefficient for aptitude is multiplied by the ratio of the standard deviations for aptitude and criterion to calculate the standardized regression coefficient for aptitude. Then the difference between standardized regression coefficients for the two treatments is found to yield the standardized regression coefficient for the interaction. (The formulas and procedures used to calculate the statistics are provided in detail in Appendix B).

Meta-analysis can be used with the standardized regression weights to estimate the weighted average regression weight, observed variance in the regression weights, error variance and estimated true variance in the regression weights. These estimates can be used in the manner suggested by Hunter et al, (1982) to statistically correct for sampling error. However, it is not possible to make simple corrections for error of measurement. As Cohen & Cohen (1983) noted, "that measurement error may decrease or increase or even change the sign of a partial relationship holds for all our means of expressing partial relationships" (p. 407), including unstandardized regression weights, standardized regression weights, semi-partial correlations, and partial correlations.

The estimated standardized regression weight for an ATI is affected by unreliability in the measures of ability, treatment, the interaction of ability and treatment, and achievement. Cohen & Cohen pointed out that unreliability in the dependent variable acts only to attenuate regression weights. Corrections for unreliability in the dependent variable, therefore, inevitably result in an increase in the absolute value of the regression weight because these corrections do not change the numerator and only reduce the denominator of the formula for the "true" or corrected regression weight. The same is true of the independent variable that is being correlated with the dependent variable, from which other variables are partialled. Therefore, the net effect of unreliability in achievement outcome and in the interaction term is to increase the regression weight. This is not the case for corrections of

measurement in the partialled variables, aptitudes and treatments. Unreliability in the partialled independent variables affects the numerator and the denominator and can reduce or increase the absolute value of the regression weight. Because of the complexity of the relationships among the variables it is close to impossible to estimate what error of measurement in the independent variables will do to the standardized regression coefficient. Cohen & Cohen suggested that "the best way to deal with this problem is to finesse it entirely by using IVs [independent variables] whose reliability is, if not perfect, at least high" (1983, p. 411).

The meta-analyst cannot "finesse" the problem of error of measurement by selecting variables with high reliabilities since the reliabilities of those variables must be taken as found in the original studies. These reliabilities can be collected, however, even though a quantitative correction cannot be calculated. Rather, the reliabilities can be evaluated subjectively, using Cohen and Cohen's guidelines, as evidence of the impact of error of measurement on the variance among effect sizes.

The use of meta-analysis in this case is limited to statistical corrections for sampling error because the effects of error of measurement and range restriction can not be estimated. This limitation is not particularly unusual since other meta-analysts (e.g., Fisher & Gitelson, 1983; Schmitt, Gooding, Noe & Kirsch, 1984) have also been restricted to sampling error corrections only.

Data Collection

In meta-analysis data collection refers to the collection of relevant research studies and the coding of effects and study characteristics. For the present review, the search was restricted to published studies. Many reviewers also collect unpublished studies to ensure that they have obtained a representative sample of the studies that have actually been conducted. In many instances, a meta-analysis which relies solely on published studies can provide biased results because published studies are generally considered to be designed better and are more likely to have significant results than unpublished studies. In this instance, such "publication bias" is unlikely to be a problem since ATI studies are often published with nonsignificant interactions because the significant main effects are noteworthy. In addition, it is possible to assess the influence of the exclusion of unpublished studies by calculating Rosenthal's (1979) fail-safe number which estimates how many studies of null results must reside in researchers' file drawers to reduce the meta-analytic finding to nonsignificance.

The research studies were collected using several sources. First, the literature search was initiated by collecting studies which had been published between 1978, the year after the publication of Cronbach and Snow's handbook on ATI, and 1986, inclusively, and were identified in the Social Science Citation Index as having Cronbach and Snow in its reference list. Second, studies were identified by searching the Education Index and Current Index to Journals in Education under several headings, including

Aptitude Tests, Aptitudes and Aptitude-Treatment Interaction. Third, the reference sections of all the collected empirical studies and of related theoretical articles were examined for omitted studies.

After the studies were collected, they were evaluated more closely for inclusion in the meta-analysis. To be included, they had to meet several criteria. The study had to examine the effect of the interaction of ability and instructional technique on achievement. The measurement of ability was restricted to demonstration of ability through achievement tests or generally accepted ability measures. A number of studies which assessed cognitive style and personality traits were subsequently eliminated.

The study also had to have at least two definitive treatments (as suggested by Cronbach and Snow) in order to identify any differential effects of the treatments on the achievement of low versus high ability subjects. This eliminated studies in which the first group received instruction and the second, control group only completed measures, not encountering the instructional material at all. This also eliminated studies in which only one ability group experienced the treatment. In addition, the dependent variable had to be some assessment of achievement; therefore, studies whose dependent variables were attitudes or reactions were eliminated. Studies were also omitted if they conducted research on a special sample, such as learning disabled students, in order to enhance generalizability. Lastly, studies were included only if they partitioned variance for the aptitude, the treatment and the

interaction. The regression weights for the interaction should have the same variances removed (due to the main effects for aptitude and treatment) across the studies in order to accumulate consistent and similar statistics. This requirement eliminated studies which used procedures which removed covariates or independent variables in addition to the aptitude and treatment main effects.

In total, the review yielded 56 studies that tested some kind of ability by treatment interaction. However, many of these did not meet the criteria stated above, the most frequent reason for exclusion from the meta-analysis being that the treatments were not a manipulation of instructional support. Eliminating these studies yielded 30 that assessed the relevant ATIs. Twelve of these studies with 14 independent samples provided the statistical information necessary for conducting the meta-analysis using standardized regression weights. Information accumulated on these samples included the ATI statistics given in the study, sample size, sample description, reliabilities of measures of abilities and achievement, and various study characteristics (suggested by Cronbach and Snow, 1977) that might moderate the result. From this information the standardized regression weight was identified or calculated and its standard error (where possible) was identified. In order to maintain independence, only one regression weight per sample was used (Bangert-Drowns, 1986). In almost all cases with multiple independent and/or dependent variables, there was only one statistic that met the criteria stated above. In the few instances where

there were several (usually because of multiple ability measures) the measure that reflected general ability was used.

Because approximately a third of the studies did not contain sufficient information to estimate the standardized regression coefficient, a supplementary meta-analysis using 13 of the 30 studies was conducted assessing the incremental change in explained variance associated with the interaction term. This incremental change in explained variance is represented by the semi-partial correlation which is the correlation between the outcome and the interaction from which treatment and aptitude effects have been removed. This meta-analysis, therefore, is also only a slight variation from customary use as it accumulates semi-partial correlations instead of correlations.

The benefit of this analysis is that it does not restrict consideration just to those studies which analyzed data using multiple regression. Studies which analyzed data using analysis of variance in a comparable fashion could also be included. As Cohen (1968) pointed out, R-squared associated with multiple regression and eta-squared associated with ANOVA are identical when the same variables are being considered, since both are interpreted as the proportion of variance in the dependent variable accounted for by the independent variable under consideration. This analysis yielded 13 studies: 7 studies provided the incremental change in the squared multiple correlation associated with the interaction and 6 studies using ANOVA procedures provided eta-squared associated with the interaction.

Results

As shown in Table 1, for the total set of 14 standardized regression coefficients and a total sample size of 2092, the weighted average regression coefficient was .192. This was significantly different from zero as the confidence interval ranged from .044 to .341 and did not include zero. However, considerable observed variance was found across the set of studies (.018), of which approximately 33% (.006) was due to sampling error. The remaining variance could be due to either (a) statistical artifacts such as unreliability or (b) moderator variables. Due to limitations in reporting or meta-analysis methodology, it is impossible to ascertain with any certainty which of these two factors account for the remaining variance. However, each of these possibilities is explored below.

As shown in Table 2, the sample size weighted incremental change in explained variance was .03 for 13 studies with a total sample size of 1466. A confidence interval cannot be constructed because it is not possible to get an estimate of the error variance across studies. This estimate cannot be obtained because, as noted by Keppel (1982), the incremental change in explained variance is not a test statistic.

Statistical Artifacts

As discussed above, Hunter et al (1982) suggested that variation among effect sizes could be due to statistical artifacts other than sampling error: unreliability in the independent variable, unreliability in the criterion, restriction in range,

Table 1. Results of Meta-Analysis of Standardized Regression Coefficients

Study	Sample Size	High Support Beta	Low Support Beta	|Beta|	St.error
Adams & McLeod (1979)	97	.31	.06	.25	.08
Elawar & Corno (1985)	336	.31	.30	.005	
Greene (1980)	165	.46	.52	.06	.05
McLeod & Adams (1980)	47	.60	.53	.07	.12
McLeod & Briggs (1980)	66	.50	.06	.44	.11
Pascarella (1978)	248	.22	.43	.21	.06
	110	.22	.55	.32	.09
Peterson et al (1980)	145	.23	.60	.37	
	146	.32	.49	.17	
Ross & Rakow (1981)	63	.68	.41	.26	.07
Stinard & Dolphin (1981)	200	.08	.45	.37	

Table 1. (cont'd)

Study	Sample Size	High Support Beta	Low Support Beta	|Beta|	St.error
Threadgill (1979)	126	.16	.28	.12	
Winn (1980)	122	.05	.01	.04	.09
Winn (1981)	221	.09	.35	.25	
Total	2092				
Average	149			.19[1]	.08

[1] Average standardized regression coefficient weighted by sample size. Differences are due to rounding error.

Table 2. Results for Incremental Variance associated with ATI

Study	Sample Size	Change in R squared	Eta Squared	Incremental Variance
Cicchelli (1982)	64		.02	
Gettinger (1983)	129	.06		
Janicki & Peterson (1981)	117	.002		
Kallison (1986)	67		.02	
Koran & Koran (1980)	84		.08	
Kozma (1982)	182	.01		
Lehman et al (1984)	160		.10	
McKinney et al (1983)	85		.01	
Peterson (1977)	98	.001		
Peterson (1979)	143	.02		
Peterson & Janicki (1979)	100	.01		

Table 2 (cont'd)

Study	Sample Size	Change in R squared	Eta Squared	Incremental Variance
Peterson et al (1981)	93	.003		
Smith (1985)	144		.03	
Total	1466			.03[1]
Average	113			

[1] Average incremental variance associated with the interaction is weighted by sample size. Differences are due to rounding error.

computational or typographical errors, measurement contamination or deficiency and different factor structures of measures. Meta-analysis techniques typically can correct for unreliability and range restriction; however, in this case, such quantitative corrections are not feasible due to the complexity of the relationships among the variables and/or reporting deficiencies.

The effect on the variance among effect sizes due to unreliability cannot be corrected because the relationships among the dependent, independent and treatment variables are too complex. Yet a qualitative assessment of its impact is possible. The impact of unreliability in the criterion, as noted by Cohen and Cohen (1983), is straightforward, although not simple to estimate, as it always serves to attenuate effect sizes.

As indicated in Table 3, the average reliability of criterion measures (r_{yy}) is .77 with a standard deviation of .11 and a range from .50 to .93. The wide range and large standard deviation suggest that much of the variability in the standardized regression weights accumulated in the meta-analysis could be due to unreliability in the criterion. As noted above, unreliability in the criterion only attenuates effect sizes suggesting that the average weighted regression weight is underestimated.

The impact of unreliability in the independent variables, aptitude and treatments, is also too complex to estimate. Unreliability in this case, however, may either increase or decrease effect sizes. Cohen and Cohen (1983) suggested that the way to deal with unreliability in such situations is to use independent

Table 3. Data and Summary for Error of Measurement

Study	r_{xx}	r_{yy}
Adams & McLeod (1979)	.78	.50
Cicchelli (1982)	.76	
Elawar & Corno (1985)	.79	.82
Gettinger (1983)		.77
Janicki & Peterson (1981)	.87	.93
Kallison (1986)	.73	.83
Koran & Koran (1980)	.79	.75
Kozma (1982)		.92
Lehman et al (1984)		.72
McKinney et al (1983)		.70
McLeod & Adams (1980)	.82	.61
Pascarella (1978)	.84	.75
Peterson (1977)	.94	.78
Peterson (1979)	.88	.51
Peterson & Janicki (1979)	.92	
Peterson et al (1980)		.93
Peterson et al (1981)	.87	.82
Smith (1985)		.82
Stinard & Dolphin (1981)		.92
Threadgill (1979)	.72	.86

Table 3. (cont'd)

Study	r_{xx}	r_{yy}
Winn (1980)		.80
Winn (1981)	.88	.83
Average	.84	.77
Variance	.004	.013
Standard Deviation	.061	.114

variables that have perfect or at least high reliabilities. Although they may not be high enough to dispel all threat to the results, the reliabilities of the independent variables in these ATI studies, treatments and aptitudes, are fairly high. The reliability of the treatment variable can be said to be perfect or nearly perfect since treatments are experimental variables which are manipulated by the experimenter who has a vested interest in preserving the integrity of the treatments (Pedhazur, 1982). In addition, the reliability of the measures of aptitude can be said to be fairly high since, as shown in Table 3, the average reliability coefficient for ability was .84 with a range from .73 to .94 and a standard deviation of .06. As with unreliability in the criterion, the variance in the standardized regression weights accumulated for this meta-analysis could be due, in part, to unreliability in the measurement of aptitude.

Effect sizes are also attenuated when the ranges of variables are restricted by the sampling procedure (Cohen & Cohen, 1983). The average weighted effect size, therefore, could be an underestimate if range restriction has occurred. In addition, range restriction can increase the variability among the effect sizes of the individual studies, if the sampling procedures used in different studies result in different ranges. For example, (given a true relationship between aptitude and learning) the effect size in a study that assesses the relationship between aptitude and learning for gifted students will be smaller than the effect size in a study which uses average and gifted children. The variability in the

estimates of the true relationship would be due to the restriction in range on the aptitude variable.

In this case, range restriction may account for some of the variance in effect sizes because the different samples reflect different ranges on aptitudes. The studies in the meta-analysis utilized samples that could be classified into three groups: primary school students, secondary school students, and college students. Compulsory school attendance laws require that primary aged children attend school. This almost insures that, at the primary school level, a wide range of aptitude levels are sampled. However, older students are not required to attend school. For whatever reasons, it appears that those students who leave school when they are old enough, tend to have lower aptitude levels which restricts the range on samples drawn from secondary schools. Aptitude ranges are further restricted in college samples since aptitude tests are used in selecting students to enroll in college programs. Those students with low aptitude test scores are not even accepted into college. Therefore, aptitude ranges may decrease as the samples are drawn from higher educational levels.

Range restriction could also occur in the treatment manipulations if investigators used different strengths or levels of treatments in their experiments. All the investigators may have manipulated amount of support given to students but they may have used different levels of support. The variability in the strength of treatment could also increase the variability in the effect sizes among the studies.

Corrections for range restriction in aptitudes and treatments are not possible in this case since much of the necessary information is not reported in the original studies. The evidence suggests, however, that ranges are not severely restricted since, as will be discussed in more detail in the next section, the average effect size is actually larger as educational level increases, rather than smaller as would be expected from restricted ranges.

Unreliability in the independent variables, unreliability in the criteria, and range restriction, therefore, could account for some of the remaining variance in effect sizes. The critical question is whether these statistical artifacts (and the others that can never be corrected) account for all the variance. If they do, then all of the variance is explained and a search for moderators is unnecessary. However, if they do not, then moderators probably account for the remaining variance and a search for relevant moderators is appropriate. This question cannot be easily answered, since, in this meta-analysis, corrections were not possible for any of the statistical artifacts besides sampling error. Thus, while we cannot refute the possibility that all the remaining variance is due to statistical artifacts, the evidence taken together is not so overwhelming to preclude a search for moderators.

Potential Moderators

Two potential moderators were investigated and the results of this analysis are reported in Table 4. First, ability was divided into specific (mathematical and verbal) and general abilities to investigate empirically Cronbach and Snow's (1977) judgment that the

Table 4. Results of Moderator Analyses

Study	Sample	Ability Measure
Adams & McLeod (1979)	college	specific-quantitative
Elawar & Corno (1985)	elementary	general
Greene (1980)	elementary	general
McLeod & Adams (1980)	college	specific-quantitative
McLeod & Briggs (1980)	college	specific-quantitative
Pascarella (1978) Sample 1	college	specific-quantitative
Pascarella (1978) Sample 2	college	specific-quantitative
Peterson et al (1980) Sample 1	high school	specific-verbal
Peterson et al (1980) Sample 2	high school	specific-verbal
Ross & Rakow (1981)	college	specific-quantitative
Stinard & Dolphin (1981)	college	specific-science
Threadgill (1979)	high school	specific-verbal
Winn (1980)	high school	general
Winn (1981)	high school	specific-verbal

Table 4. (cont'd)

Group	Number of Studies	Sample Size	Weighted Avg. Reg. Coeff.	95% Conf. Interval
Education Moderator Analysis				
College	7	831	.28	.12 - .44
High School (Grades 9-12)	5	760	.20	.03 - .37
Elementary (Grades 1-8)	2	501	.02	-.08 - .12
Ability Measure Moderator Analysis				
General Ability	3	623	.03	-.11 - .17
Specific Abilities	11	1469	.26	.10 - .42
Quantitative	6	631	.26	.10 - .42
Verbal	4	638	.24	.08 - .40
Science	1	200		

interaction between general aptitudes and treatments would be stronger than between specific aptitudes and treatments. The findings are not consistent with this hypothesis in two ways. First, the confidence interval for general ability overlaps the confidence interval for specific ability. Taking a very conservative stance, this suggests that the regression weights for general and specific ability are not significantly different from each other. However, the weighted average regression weight for general ability does not fall within the confidence interval for the specific tests and the weighted average regression weight for the specific tests does not fall within the confidence interval for the general tests. Taking a less conservative stance, this suggests they are significantly different from each other. Second, the findings are inconsistent with the hypothesis because, differences across tests were also in the wrong direction. Where Cronbach and Snow predicted the interaction of treatment with general ability would be greater than the interaction of treatment with specific abilities, the opposite result was obtained.

Second, the subject pool was subgrouped by education level to test for the impact of range restriction by looking at whether interaction effects become weaker as subjects are older and have a higher educational level. The findings are inconsistent with this hypothesis in several ways. First, as indicated by the nonoverlapping confidence intervals, the weighted average regression weight for primary school students is significantly <u>smaller</u> than the average regression weight for college students, just opposite what

would be predicted from range restriction. In addition, the confidence interval for secondary school students overlaps with the confidence intervals of both college and primary students. On the other hand, the weighted average regression coefficient for college students (.28) is within the confidence interval for secondary school students, but outside the confidence interval for primary school students. This can be interpreted as suggestive of the possibility of an educational level by ATI interaction effect, particularly since the middle group falls between the extremes.

Discussion

The purpose of this meta-analysis was to test for the presence and magnitude of ATIs in educational settings. The evidence suggests that such ATIs were present and that the average slope difference was .19 and ranged between .04 to .34. In addition the evidence suggests that some of the inconsistent findings and lack of replication among single studies can be explained by sampling error. It is possible that the remaining unexplained variance in achievement could be due to error of measurement or other statistical artifacts, but this could not be directly confirmed. It is also possible that the unexplained variance in achievement could be due to moderators such as ability or education. The supplemental analysis indicated that the sample size weighted average incremental change in explained variance was .03.

The results of this meta-analysis are consistent with the ATI hypothesis that instructional support would interact with ability to affect learning. Most of the ability measures were some assessment

of prior achievement--often using mathematics, science, vocabulary or reading achievement tests. The treatments manipulated instructional support. For example, in highly supportive treatments, subjects had extra learning aides in the form of adjunct questions or diagrams, or teacher or small-group assistance and, in low support treatments, subjects worked on their own or heard lectures.

Generally, the rationale for this interaction is that providing instructional support will increase the performance of low aptitude individuals with little effect on high aptitude individuals. It would be expected, then, that the relationship between aptitude and performance for individuals in the high instructional support treatment would be weaker than that for individuals in the low instructional support treatment. Although this cannot be directly tested through the meta-analysis, it is interesting to note, as shown in Table 1, that the standardized regression coefficient for low instructional support treatments is greater than the standardized regression coefficient for high instructional support treatments in 8 of the 14 cases. This observation is consistent with the idea that instructional support compensates for low aptitude.

The moderator analyses were weakly consistent with the notion that the strength of ATIs varies due to other factors. The finding that the interactions of treatments with specific abilities may have stronger effects on achievement than the interactions of treatments with general abilities challenges Cronbach and Snow's (1977)

conclusion that the most consistent significant findings of ATIs were associated with general ability.

The data were inconsistent with the hypothesis that educational level would be a moderator due to increasingly restricted ranges in ability because the slope differences become stronger, not weaker, as educational level increased. Snow's (1978) hypothesis that achievement is enhanced when educational treatments match the underlying cognitive processes suggests two possible explanations for this finding. First, primary school-aged children may not have developed the stable, refined cognitive processes that college-aged students have developed. Since their cognitive processes would be changing and developing, primary school students are less likely to reflect a match between treatment and cognitive processes. On the other hand, college students should have fairly stable and mature cognitive processes which can more easily either match or not match with a specific treatment. This would result in small slope differences among treatments for students at lower educational levels and larger slope differences for students at higher educational levels. Second, if they develop and mature at early ages, then cognitive processes become a stable individual difference from lower to higher educational levels. Differences in slopes then may be due to the increased complexity of material that must be learned as educational level increases. A student who had a fairly simple cognitive process could handle material at lower levels but could not handle material at higher educational levels. Assuming that students with complex cognitive processes can handle simple and

complex material or treatments, then many students' processes would match treatments at lower levels resulting in small slope differences. However, at higher levels, fewer students would match with the more complex treatments. This would result in larger differences in relationships between ability and achievement.

This is clearly a post hoc explanation for an issue that is not resolved since plausible theories exist to generate competing hypotheses. One hypothesis predicts that if there is restriction in range then the strength of an ATI effect should decrease as educational level increases. An alternative hypothesis predicts that if differences among students increase over time, then the strength of an ATI effect should increase with educational level. Future research can explore these competing hypotheses.

Several words of caution may be necessary to prevent inappropriate application of these results. The confidence interval around the weighted average regression coefficient revealed that the coefficient for the interaction was significantly different from zero, but the lower end point of the confidence interval was .04 which may not be large enough to justify redesigning instruction. The benefits associated with such a small difference between treatment effects for the different ability levels may be unlikely to outweigh the costs of adapting the treatments to individual needs. In contrast, the upper end point is not trivial (.34) and suggests that adapting treatments to individuals would yield significant benefits. In addition, it is not possible to establish whether the interaction is ordinal or disordinal without more

information about the measures. However, according to Cronbach and Snow (1977) both types of interactions can be applied to instructional design. Disordinal interactions more obviously suggest that different treatments should be used for different individuals, depending on their aptitudes. Ordinal interactions are particularly useful if one treatment is more costly than another. Instructional costs could be reduced if the more costly treatment is given only to those who are likely to benefit from it.

It is also important to consider whether the results may be due to some bias in the meta-analysis itself, such as publication bias. Publication bias occurs when an analysis employs only published studies which may not be representative of the total set of conducted studies, in that studies with nonsignificant results are underrepresented. There are two pieces of evidence that publication bias is not a problem. First, a simple count of the outcomes of the single study findings shows that almost half of the studies included in the meta-analysis published nonsignificant results. Of the 14 independent samples providing regression coefficients, 6 provided nonsignificant and 8 provided significant results. Second, the calculation of the number of unpublished studies specifically on this particular ATI which would bring the results to nonsignificance is large. Orwin (1983) adapted Rosenthal's fail-safe formula for Z scores to use with the effect size, d. Since correlations can be converted to ds, Orwin's formula (listed in Appendix B) was used to estimate the fail safe number for this meta-analysis. The results indicate that over 200 studies with a small average effect size

would have to exist to reverse these results. Therefore, it is unlikely that publication bias is a serious threat since an unlikely number of studies are needed to reverse these results.

In addition, it is important to determine whether reporting bias could have affected the results. Reporting bias occurs if the subset of studies which did not report sufficient statistical information to be included in the meta-analysis is methodologically defective or systematically nonsignificant. Reporting bias does not appear to be a substantial problem in this case for two reasons. First, the incremental change in explained variance associated with the interaction was estimated using a number of the studies that did not provide sufficient information to calculate a regression coefficient. The incremental change was .03 which is small, but not trivial, since the interaction is entered as the third variable in the regression equation. Second, a comparison of the distribution of the significance of results for those multiple regression studies which could not be included in the meta-analysis (6 significant, 7 nonsignificant) is nearly identical to the distribution of the significance of results for those studies which were included (8 significant, 6 nonsignificant).

Although it does not compromise the results of this meta-analysis, deficient reporting is obviously a serious problem in recent ATI research (e.g., Berliner & Cahen, 1973; Cronbach & Snow, 1977). Half of the relevant studies had to be excluded due to omitted statistics. The omission of basic descriptive statistics for subgroups was the reason effect sizes could not be calculated in

most cases. A number of studies did provide some basic statistics (means and sample sizes), but they often failed to include standard deviations. Standard deviations are necessary for the calculation of the standardized regression coefficient. In addition, standard deviations may account for an interaction (Cronbach & Snow, 1977). Researchers who used multiple regression techniques often gave F statistics and change in R squared, but failed to give parameters for the equations. Another all too common problem was that studies only included information for statistically significant effects and passed off nonsignificant effects in a short summary sentence such as "no other interactions were statistically significant." For others to gather and summarize information across studies with any degree of confidence, researchers and journal editors must insist on providing sufficient statistical evidence for all tests such as means, sample sizes and standard deviations for all treatment groups, complete regression equations or ANOVA tables, and nonsignificant, as well as significant findings.

In addition, it is important not to over-interpret the results of the moderator analyses since several of the subgroups are based on a small number of studies. Schmitt et al (1984) warned that the results may not be stable when the number of studies (k) in a meta-analysis is small, as the number of studies influences estimates of variance, the primary focus of moderator analysis. When k is small, the variance estimates are based on only a few sample points and could be changed significantly by the addition of only a few studies. On the other hand, the estimates of the average effect

size are not affected drastically by the number of studies since they are derived from the total sample size. Average effect sizes are unlikely to be changed significantly by the addition of a few studies.

Future research needs to continue to look for moderating effects, particularly from different treatment characteristics. In addition, the differential costs and benefits of instructional programs should also be considered to determine the efficacy of using one or more treatments for individuals.

Finally, these results support the use of meta-analysis for other ATI investigations with other aptitudes and different treatment manipulations, e.g., as inductive reasoning and programmed instruction versus regular text treatments. Future research could rely on meta-analytic techniques to assess the effect of the interaction of cognitive style or personality variables and treatments on different dependent variables such as achievement or attitudes. It could also investigate ATIs in different environments. Most of the research has been conducted in education, however, other instructional situations exist which could also benefit from ATIs, such as adult education centers or industrial training centers. The findings here should encourage continued investigation of aptitude-treatment interactions.

CHAPTER 3: LITERATURE REVIEW OF ADAPTATION WITH TRAINING

Overview

Baldwin and Ford (1988) argued that many studies in which outcomes of training are evaluated ignore aptitudes and individual differences. They identified only 25 studies which assessed any type of trainee characteristics (ability, motivation and personality). Based on their review of these studies, they concluded that "...empirical investigations of ability, personality and motivational effects on training and transfer outcomes are quite limited" (p. 68).

The few that do assess aptitudes typically validate them as predictors of training success in a fixed-training treatment. These studies, although not direct tests of adaptive treatment, may provide some information about the relationship between aptitude and payoff for a given treatment, and thus, some information about how to adapt treatments for a more complete test. The focus of this review is on the surprisingly few published studies of training in which aptitudes have been assessed in one fashion or another. First, the studies which use aptitudes to predict training success will be briefly described and evaluated for the information they contribute for adaptive treatment. These studies typically involve variables such as personality or ability. A second set of studies tests the validity of predictors of training success and performance that are based on aspects of training itself. The third set of studies actually tests ATIs in training situations.

Finally, the empirical literature is evaluated. The results of these sets of studies, together with findings in the educational literature, suggest that adaptive treatment may increase payoffs to the organization; however, these tests have been inadequate in many ways. The theory expounded by Cronbach and Gleser (1965) and Cronbach and Snow (1977) and the empirical evidence in education and training encourage future tests of adaptive treatment in industry and suggest how future tests can be conducted more effectively.

Fixed-treatment Selection in Training

An extensive search was conducted to find studies which focused on using aptitudes as predictors of training success. Studies were gathered from several sources: references in Cronbach and Snow (1977) which used industrial or military settings and in Wexley (1984) and Baldwin and Ford (1988) which focused on trainee characteristics; a computer search of Psychological Abstracts using key words such as training (personnel, human relations, military and on-the-job) and individual differences, personality traits, and ability; and the reference lists of all studies identified in the other sources. Studies which gathered demographic information or manipulated motivational variables were eliminated from the list that resulted from this search, leaving studies which assessed personality or ability constructs.

An examination of these studies, summarized in Table 5, reveals that the research in this area has tested aptitudes under four general training treatments: military, management development, skills training and pilot training. In each of these, the

Table 5. Empirical Studies of Aptitudes Predicting Training Success

Study (Year)	Aptitude Variables	Training Course	Criterion Measure	Significant Predictors
Baumgartel et al (1978)	Personality	Management Development	Self-report Transfer	High nAch, LOC and energy level
Buch (1984)	Personality	Pilot Judgment	Learning, Performance	N/A
Love & O'Hara (1987)	Personality	Job Training (Miscellaneous)	Performance	Vocational Maturity
Miles (1965)	Personality	Management Development	Learning, Behavior	none
Neel & Dunn (1960)	Personality	Supervisory Training	Learning	Supervisory Understanding
Noe & Schmitt (1986)	Personality	Management Development	Reaction, Learning, Behavior, Results	N/A
Ross & Houtz (1979)	Personality	Job Training (Miscellaneous)	Reaction	none
	Ability (Self-report)			Post-training perception of ability
Ryman & Biersma (1975)	Personality	Job Training (Diving)	Learning	Training Confidence

Table 5 (cont'd)

Study (Year)	Aptitude Variables	Training Course	Criterion Measure	Significant Predictors
Taylor & Tajen (1948)	Ability	Job Training (IBM Tabulating Equipment)	Learning	Clerical Speed, Figure Classification
Tubiana & Ben-Shakhar (1982)	Personality	Basic Military Training	Learning	Specific Personality Not Specified
	Ability			Intelligence
Williams et al (1982)	Ability	Basic Military	Learning	Intelligence, Physical

researchers typically tested personality variables as predictors of a learning measure of training success. Unfortunately no one personality characteristic was tested consistently enough to make any generalizations about its impact on training success across training studies. A few studies tested the validity of ability measures as predictors of training success. Although caution should be exercised since the number of studies is small, the results do indicate that in every instance, ability was predictive of training success: intelligence was predictive of learning criteria for basic military training (Tubiana & Ben-Shakhar, 1982; Williams et al, 1982) and clerical abilities were predictive of learning for military recruits in training to learn to use and maintain IBM Tabulating equipment (E. K. Taylor & Tajen, 1948). In addition, a self-report measure of ability taken after training was predictive of satisfaction with training for IBM managers who took a variety of training courses (Ross & Houtz, 1979). This finding, however, could have resulted because self-reported ability and satisfaction were both measured right after training. It could also be possible that the self-report of ability is less a measure of ability and more a measure of confidence or self-efficacy, which could be considered a relatively unstable personality characteristic or even an attitude.

These studies, at best, suggest very little about payoff functions for treatments. The most reassuring, but hardly startling, finding is that ability is related to success in training for basic military training such that those with high ability are more likely to be successful. This finding certainly does not

suggest any new and earthshattering ideas; however, it does imply how future research can be conducted. First, Table 5 contains only 3 studies (out of the 11) which assessed post-training performance. Since the primary goal of training is to improve job performance, studies of aptitudes in different treatments need to employ better criteria, going beyond simple learning and reaction measures assessed right after training, to measuring various aspects of job performance (Baldwin & Ford, 1988). Second, the table indicates that few aptitudes have been tested under different training programs. Although the strength of a relationship between the aptitude and an outcome may have been estimated in one treatment, it needs to be assessed in a different treatment to establish the consistent or differential nature of the relationship. For example, it seems hypothetically appropriate to determine whether locus of control consistently predicts performance in management development training. After establishing this relationship, it would be appropriate to test for the effect of locus of control in a different type of training treatment, such as in job skill training. If internal locus of control predicts success in one treatment and external locus of control predicts success in another, then locus of control could be used to assign individuals to treatments. However, testing for payoff functions in this fashion is highly inefficient and can be conducted in more efficient ways, using more than one training program per study, which will be discussed below.

Selection based on samples of training tasks

The procedures associated with trainability testing and with early training success appear, on the surface, to be situations in which training is adapted to the individuals' aptitudes. However, an important, but subtle aspect of the procedures suggests that they are not instances of adaptive treatment, but are examples of sequential testing. This aspect refers back to the decision process in which decisions are either terminal or investigative. Measures of trainability and early training success are used as selection tests, tests of information which can be used to make terminal decisions. However, when these measures are collected, the decision process is still in the investigatory stage. Presumably, some earlier tests have indicated that the applicants have the potential to be successful employees but the tests have not been used to make a terminal decision. The terminal decision is yet to come, based on the performance of the applicants on the trainability tests or in early training modules. Adaptive treatment refers to terminal decisions. Since these are not yet terminal decisions, they are not instances of adaptive treatment. However, results based on tests of trainability and early training success, as predictors of performance with different treatments may provide information on payoff functions for these tests and treatments.

Trainability

Robertson and Downs (1979) defined trainability as the capacity of the applicant to learn to do the job, i.e., the abilities and capabilities the individual has that will increase the probability

of successful training and adequate performance on the job after training. The concept stems from Wernimont and Campbell's (1968) proposal of behavioral consistency that the best predictor of future performance is past performance. This is applied by looking at samples of behavior or performance which include tasks that have to be learned or trained. For example, Robertson and Mindel (1980) conducted an experiment to assess whether a trainability test would be predictive of success in craft courses. Potential trainees were taught how to use a few tools and construct a certain model. They then constructed a model on their own. This model was a test piece to predict whether the trainees should pursue that craft. Therefore, trainability tests show that potential trainees have the capability to learn a sample of the skills that are needed on the job. This is used as evidence that they can learn the population of skills and will successfully complete the training course and, ultimately, will perform the job adequately.

Previous research on trainability has primarily been conducted in manual and clerical jobs (Robertson and Downs, 1979), so it has most frequently been operationalized as a motor skill (e.g., Downs, 1970; Gordon & Kleiman, 1976; Robertson & Downs, 1979; Robertson & Mindel, 1980; Siegel, 1983; Siegel & Bergman, 1975; Smith & Downs, 1975). However, Gill (1982) suggested that in-basket technology and training procedures in identifying priorities and making decisions might be used to assess trainability for management positions.

Early Training Success/Time in Training

Several researchers have suggested that time in training is a legitimate focus for evaluating and predicting training success (e.g., Gordon, 1955; Gordon & Cohen, 1973; McGehee, 1948; Zink, 1982). The logic is also based on behavioral consistency in that the time it takes to learn a task or to complete training is indicative of subsequent performance levels. The hypothesis is that time to completion on early training modules is predictive of total time to completion of the training program. The empirical results have been encouraging. In all four studies listed above, time to completion (or a similar measure) was an excellent predictor of training performance.

Although trainability and early training success are not presently considered aspects of adaptive treatment, they certainly contain elements that could be used to adapt treatments. Trainees in a certain treatment who perform poorly or slowly on early training modules in that treatment may perform significantly better in a different training treatment. Rather than being terminated from training altogether, they could be placed, for example, in a remedial or compensatory program and evaluated for success (using trainability or early training success measures) in the different program. Such adaptation could increase the trainees' performance in several ways. It may be that these trainees need a little more practice with the basic tools or procedures. Remedial training or practice with the basics can increase competence so that more complex tasks encountered later in training can be performed

satisfactorily. It may be that these trainees need a little more confidence in their ability to do the tasks. Ryman and Biersner (1975) found that training confidence was associated with training success. Slower trainees may just need a little more time and practice to build their confidence. There is a point, however, where the time a trainee needs to reach a satisfactory level of performance is excessive. McGehee (1948) noted that time spent in training can be expensive. Somewhere in the sequence of testing and training, a cutoff would need to be established at which trainees who have not reached a reasonable standard of competency would have to be terminated or put into another job or training program.

In summary, measures of trainability or early training success could be used in the usual way as predictors of training success and sources of information to the decision maker who then places trainees in terminal treatments. On the other hand, they could be used for selection and to help adapt treatments.

Aptitude-treatment interactions in training

An extensive computer search of Psychological Abstracts and cross-referencing of a number of references yielded eight published studies on adaptive treatment or ATI in training settings which met Cronbach and Snow's (1977) criteria for adequate ATI research, four of which are described briefly in Cronbach and Snow (Beswick & Tallmadge, 1971; James, 1962; Tallmadge and Shearer, 1969; 1971). These studies were, interestingly, all conducted using military recruits for subjects in a variety of training situations: a task module on intra-muscular injections for trainees learning to be

paramedics (Brooks, Ebener, Manning & Balson, 1985), a computer-driven course in inventory management (McCombs & McDaniel, 1981), a course in automotive mechanics (C. W. Taylor, 1952), a laboratory-level training module in an industrial search/quality control skill (Duncan, 1971), a short module on information in the field of military aviation (James, 1962) and a series of modules in Aircraft Recognition, Celestial Navigation and Transportation Techniques (Tallmadge & Shearer, 1969; 1971). Only McCombs and McDaniel (1981) and Tallmadge and Shearer (1971) relate their studies to research on aptitude-treatment interactions and adaptive training; however, the others certainly fit into this research paradigm. Each of these studies will be reviewed briefly and evaluated for their contribution to adaptive-training treatments. Those studies that looked at the interaction of measures of ability or achievement with training are reviewed first, followed by studies that looked at other aptitudes.

Interactions between Ability and Training

Prior achievement. Three studies looked at prior achievement as aptitude: C. W. Taylor (1952), Duncan (1971) and Brooks et al (1985). C. W. Taylor (1952) validated a test battery of some paper and pencil tests and work sample tests that measured prior knowledge and skill in automotive mechanics for selecting men who had prior automotive experience and could skip the first third (4 weeks) of a 12 week automotive mechanics course and still complete it successfully. The battery was tested by letting 60 high performers on the test battery and 60 low performers skip the first four weeks

of training and go straight into the last phases of training. After training, the 60 high scoring performers had significantly higher grades on training tests than the 60 lower performers. (However, despite this difference, almost all the low scorers passed the training course.)

These results were used to sort trainees into two groups. Trainees who performed above the cut-off on each test in the battery skipped the first four weeks of the course. Those who scored below the cut-off on any one of the tests took the whole 12 week training course. A follow-up test six months after the test battery compared the final course grades of those trainees who skipped the first part of the course with the final course grades of those who took the whole course. This comparison indicated that the 178 trainees who skipped part of the course performed significantly better (average final grade = 85.01 with a standard deviation of 2.01) than the 629 trainees who received the whole course (average final grade = 80.97 with a standard deviation of 3.77).

This is an example of adaptive treatment in which aptitudes are used to provide preferential treatment (Salomon, 1972) for high scoring applicants. These applicants performed better on the learning tests, despite receiving a shorter training course than most of the trainees. A utility analysis based on calculations of saved training hours indicated that almost 14 man-years of training time was saved during the first six months using this procedure since four weeks of training was saved for each man who skipped the first phase.

Duncan (1971) conducted a short-term, low task involvement, laboratory experiment in which Army recruits learned how to locate faults. Prior achievement was measured as initial unaided performance on a fault-location task. The recruits were given a specific decision-tree strategy to search for faults on a training task and were told to try to generate a general search strategy. Duncan assessed several aspects of the process, including (a) whether the trainees transferred the training to a similar but different task, (b) whether the training was retained, (c) whether ability affected transfer and retention and (d) which strategy was used to locate faults (general or specific). Duncan found that retention declined over time for low ability trainees, but transfer did not decline much for either low or high ability trainees. Duncan concluded that stability of training effects over time depends more on whether the trainee learns and remembers a general search strategy.

Training was not adapted in this study, but the results suggest how adaptation could help low ability trainees. Since low ability trainees had difficulties remembering a strategy, they apparently needed compensatory training in generating a general search strategy and in learning that strategy. High ability trainees, on the other hand, apparently could both learn the specific routines and generate a general solution which subsequently leads to better performance.

Duncan's study did not manipulate the training but rather relied on self-reports of which strategy was used. This lack of

control leaves open the possibility that other unmeasured and uncontrolled factors could account for the results.

Lastly, Brooks et al (1985) measured prior achievement with standardized English and mathematics achievement tests. They manipulated a characteristic of the training environment rather than the training itself: homogeneous or heterogeneous achievement or ability pairing. Trainees often practice tasks in pairs. The composition of the pairs, whether the individuals are at the same or different level of achievement, were hypothesized to affect performance. Brooks et al tested this hypothesis with 232 male military trainees in a paramedic training program on a task module in which trainees learned to give intramuscular injections. The trainees were grouped into pairs for practice so that ability levels either matched (homogeneous pairs--High/High or Low/Low) or did not match (heterogeneous pairs--High/Low). Subjects' scores were posted so ability levels of partners could have been known. The trainees all had the same training program with a lecture, video and practice of injections on each other. The criterion was passing or failing a mastery test within two days.

The results indicated that there were no differences in performance within homogenous pairings. High/High pairs and Low/Low pairs performed the same. However, within heterogeneous pairs, low ability trainees did better at the expense of high ability trainees.

This study also did not manipulate training per se, but it did manipulate one aspect of training. Studies like this are needed to help identify which training design characteristics can be modified

in adapting treatments. Some, seemingly insignificant, characteristics of training may be able to compensate for low achievement by some trainees in such a way that their performance is significantly improved.

General ability. James (1962) used the Armed Forces Qualifying Test to assess general ability. The intended focus of the study was to assess the impact of the interaction between training presentation (reading or lecture) and subjects' preference for a reading or lecture presentation on learning. Over 360 basic airmen took a training module on information in the field of military aviation. Their general ability level and preference was assessed prior to assignment to a training treatment so that the treatments could be counterbalanced with equal numbers of high and low ability and match, mismatch or no preference subjects. Performance was measured with a 30 item learning test. No preference effect was found; however, a significant ordinal interaction, shown in Figure 3, was found such that the difference in performance between high and low ability recruits was greater in the reading treatment than in the lecture treatment.

The usefulness of this ordinal interaction may be limited, depending on the costs associated with the reading and lecture presentations. Mean performance was higher using the reading treatments; however, only high ability recruits performed substantially better under the reading rather than the lecture presentation. If the cost of reading presentations is significantly greater than the cost of lecture presentations, then this

Figure 3. Ordinal Interaction (James, 1962).

interaction can be used to increase payoffs: the low ability recruits would be given the lecture and the high ability recruits would be given the reading presentation since they are the ones most likely to benefit from it. On the other hand, if costs are comparable or low, all recruits should receive the reading presentation.

Interactions between Other Aptitudes and Training

McCombs and McDaniel (1981), Tallmadge and Shearer (1969; 1971) and Beswick & Tallmadge (1971) assessed interactions with a variety of aptitudes. The main purpose of the study by McCombs and McDaniel (1981) was to describe a procedure designed to use the ATI or adaptive treatment approach to ascertain what aptitudes suggest the need for adaptive training treatment and to determine when such adaptation would yield the largest payoff. The study was part of a project in which a computer-based methodology was designed to identify aptitudes which could be used for adaptation and to develop effective adaptation procedures. The general procedure started with an assessment of the relationships between individual difference variables--memory ability, anxiety and curiosity--and performance on the original (a priori) training procedure. Those individuals who performed poorly would be identified. Then a new training treatment would be designed to provide them with the support that they need (e.g., attention-getting, embedded questions, practice problems). With the two treatments, a test for ATI should be conducted. The goal was to raise the performance of the originally poor performers by providing compensatory or remedial training for them.

McCombs and McDaniels provided two examples of Air Force technical training courses on inventory management which were designed using this procedure. They found significant interactions after modifying treatments which suggested that the adapted treatments were effective in meeting the needs of the students for whom they were designed. In addition, they noted a practical benefit of adapting the treatments was a reduction in training time.

This is obviously the most extensive and thorough test of adaptive treatment. This study provides an excellent example of the application of the theories presented by Cronbach and Gleser (1965) and the techniques suggested by Cronbach and Snow (1977). Although they only presented the results of two uses of the procedures, McCombs and McDaniel noted that these procedures have been developed and used successfully for thirteen different technical training programs. Such a program seems to have immediate utility in the military where large numbers of recruits with varying aptitudes need to be processed and trained expediently. However, such techniques appear to be applicable to many technical training situations in industry, particularly in mass re-training programs.

Tallmadge and colleagues (Tallmadge & Shearer, 1969; 1971; Beswick & Tallmadge, 1971) designed a series of studies to identify which aptitudes predicted success in different types of learning tasks with different learning methods. They measured 28-39 different aptitudes (e.g., mathematical, mechanical and clerical achievement, personality variables, general ability) of Navy enlisted men awaiting assignment to Basic Electricity and

Electronics School. The study appears to be a laboratory study since these subjects were not enrolled in a training program at the time of the study and the training tasks were not a part of a formal training program. However, the three tasks, Aircraft Recognition, Celestial Navigation and Transportation Techniques were relevant to their jobs in the Navy.

The three tasks were taught either using an inductive or a deductive style (which were not well described in any of the studies). The interaction of these two styles with each of the 28-39 aptitudes was tested in several ways: differences in correlations across treatments, ANOVA and multiple regression. The results of the first study (Tallmadge and Shearer, 1969) showed that the interaction between Kuder Scientific Interest and training was significant such that for the Transportation Technique training task, learning was greater with the inductive style for high scientific interest subjects only. In addition, a significant interaction of training with a composite score (consisting of scientific interest, ascendancy and musical interest) indicated that those with high composite scores performed better using inductive methods and those with low composite scores performed better using deductive methods in Transportation Techniques.

The second study (Tallmadge & Shearer, 1971) was conducted somewhat differently in that Celestial Navigation was substituted for Transportation Techniques and more aptitudes were assessed. The results revealed that an interaction of a composite score hypothesized to reflect anxiety and treatment style was significant

such that high anxiety recruits performed better in the inductive treatment and low anxiety recruits performed better in the deductive treatment.

Beswick and Tallmadge (1971) reevaluated the two previous studies, recognizing that the results did not replicate. They reanalyzed the studies using Beswick's theory of cognitive curiosity in an effort to reconcile the differences.

These studies should not be taken too seriously. A major flaw was that the results most likely occurred by chance. Throughout the studies the composite scores were derived solely on an empirical basis, creating composites that had little or no meaning. In addition, the first study employed over 100 statistical tests, with 231 subjects and over 30 variables. Although the second study performed fewer statistical tests, it had the same basic problem--too many variables and tests and not enough data points. Therefore, these studies were empirically driven; the first time theory was brought into the research program was in the post hoc analyses in Beswick and Tallmadge (1971).

Despite these problems, the work performed by Tallmadge and his colleagues does make a contribution to the ATI literature. They are excellent examples of what to avoid in future research on ATIs. In their defense, Tallmadge and Shearer conducted these studies before Cronbach and Snow's original report on ATI was available in 1969. They were conducting exploratory research in an area which was just beginning to get systematic and well-considered research attention.

The individual contributions of each of these studies have been discussed. Taken together, they suggest that the theory and procedures may have utility in industry, but do not provide conclusive evidence of the character or strength of adaptive treatment. Further tests appear warranted, however, they need to be more systematic, controlled and theory-driven than some of these previous tests.

Summary

The theory behind adaptive selection and aptitude treatment interactions and the empirical evidence on ATIs in educational settings and on payoff functions and ATIs in training settings provide the foundation for the research proposed here. The major contributions from each of these areas appear to be as follows. First, Cronbach and Gleser (1965) developed the arguments for the utility of adaptive selection using decision theory. They showed mathematically that the payoff function for adaptive selection in which expected payoff is a function of aptitude and treatment can lead to increased utility to the organization. This payoff function also shows how part of the procedure in assessing the utility of adaptive selection includes testing for aptitude-treatment interactions. Second, Cronbach and Snow (1977) expounded on how to test for aptitude-treatment interactions and provided an extensive summary of research on ATIs. Future tests of ATI should be designed along the lines they suggested.

The results of the meta-analysis suggest that the effects of training treatments can be modified by providing different levels of

instructional support. The results show that achievement of students in educational settings is enhanced by matching their abilities to amount of instructional support. It seems appropriate to test whether trainees' performance may be enhanced in the same way. More specifically, the meta-analysis indicated that the slopes of the lines for regression equations of ability on achievement for high and low instructional support treatments are significantly different, but it could not identify whether the regression lines intersected (indicating a disordinal interaction). Future research should identify the nature of the interaction (whether ordinal or disordinal) as well as assessing the impact of costs on the interaction.

The review of the training/industrial literature on aptitudes or individual differences mainly suggests that success in training is related to ability. Different treatments may have differently-sloped payoff functions with ability as an aptitude. The findings imply that such ATIs should be pursued in a more rigorous and efficient fashion.

The few studies in the training literature on adaptive treatment provide encouragement to researchers to continue studying this area. For example, C. W. Taylor (1952) found evidence that adaptive treatment using a preferential approach (Saloman, 1972) was effective in increasing efficiency (reducing time spent in training). McCombs and McDaniel (1981) found across a number of different training modules that adapting training using a compensatory approach (Saloman, 1972) was also effective.

The main import of these lines of research and inquiry suggests that future research would be effective if focused on ability in conjunction with different treatments which manipulate instructional support. As discussed above, the effectiveness of this interaction is reflected in the magnitude of standardized regression coefficient associated with the interaction. Based on the literature and the meta-analytic results, it is hypothesized that the standardized regression coefficient associated with the interaction of ability and treatment is significantly greater than zero, such that:

$$H1: Beta_{AT} > 0$$

However, this past research has not clearly specified the nature of the interaction, i.e., the processes by which instructional support may interact with ability to influence achievement. The approaches proposed by Salomon (1972) model two different possible processes. The first process, suggested by both the remedial and compensatory approaches, suggests that high instructional support will increase the achievement or performance of low aptitude individuals but will not influence the achievement or performance of high aptitude individuals. In particular, the remedial approach suggests that high instructional support may consist of <u>background information</u> that would compensate for low aptitude individuals' <u>information deficiencies</u>. High aptitude individuals would not benefit from the high instructional support because they would have already mastered that background information. On the other hand, low instructional support would omit this background information which would limit the achievement

of low aptitude individuals but would not hinder the achievement of high aptitude individuals.

On the other hand, the compensatory approach suggests that high instructional support may consist of learning assistance or aides. The achievement of low aptitude individuals would increase as the learning assistance compensated for their learning process deficiencies. High aptitude individuals would not benefit from the learning assistance because they have adequate learning skills. Low instructional support would lead to low achievement by low aptitude individuals since they lack those basic learning skills; whereas, high aptitude individuals would perform well since they possess those learning skills.

Thus, both the remedial and compensatory approaches suggest the interaction of ability and treatment would be disordinal with the intersection of the regression lines in the upper right quadrant as shown in Figure 4a. The approaches differ in the content of the instructional support: the remedial approach focuses on providing needed background information while the compensatory approach focuses on providing needed learning support.

The preferential approach reflects a somewhat different process by suggesting that high instructional support treatments may capitalize on learners' aptitudes. Instructional support would increase the achievement of low aptitude individuals some, but would increase the achievement of high aptitude individuals even more. Low instructional support would lack instructional aides, so high aptitude individuals would perform better than low aptitude

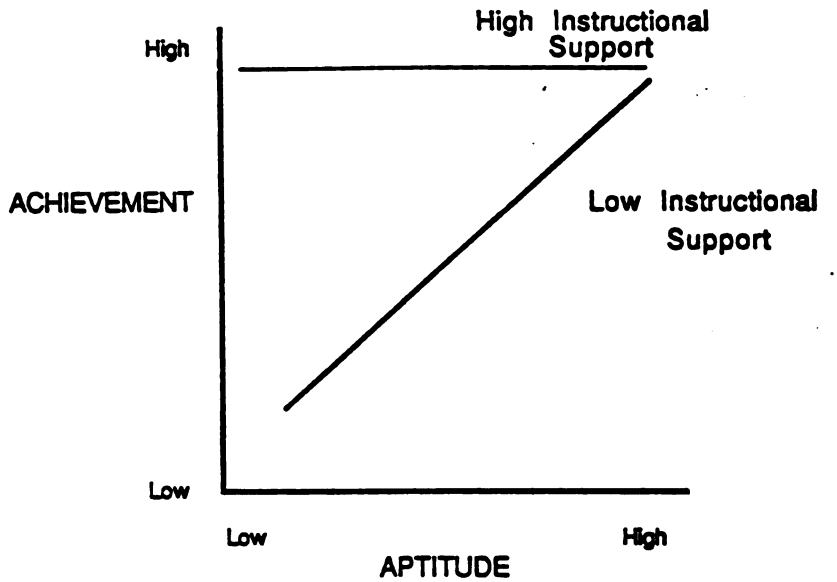


Figure 4(a). Expected Disordinal Interaction for Remedial and Compensatory Approaches

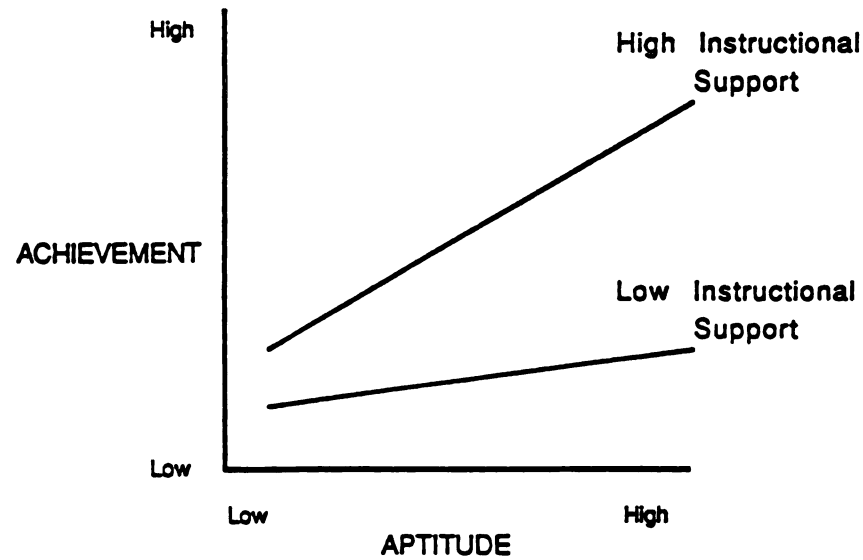


Figure 4(b). Expected Ordinal Interaction for Preferential Approach

individuals. As shown in Figure 4b, the preferential approach suggests that an ordinal interaction would model the effect of ability and treatment on achievement.

In this test, instructional support will provide instructional aide, rather than background information. Therefore, the compensatory and preferential approaches suggest the following competing hypotheses:

H1a: Based on the compensatory approach, the slope of the regression line for high instructional support should not differ significantly from zero while the slope of the regression line for low instructional support should be greater than zero. The regression lines converge as the level of aptitude increases. The interaction of aptitude with treatment on achievement is disordinal. This hypothesis can be stated as follows:

$$H1a: \quad Beta_H = 0 \text{ and } Beta_L > Beta_H$$

H1b: Based on the preferential approach, the slope of the regression line of ability on achievement for a high instructional support treatment is greater than the slope of the regression line of a low instructional support treatment. The regression lines would not intersect so the interaction of aptitude with treatment on achievement is ordinal. This hypothesis can be stated:

$$H1b: \quad Beta_H > Beta_L$$

Previous tests of ATIs in training have been conducted using military tasks; therefore, for greater generalizability, it is appropriate to investigate adaptive treatment in industrial training tasks in accordance with the guidelines suggested by Cronbach and

Snow (1977). The current project tests the competing ATI hypotheses listed above in an experimental business setting.

In addition, the training program explores other hypotheses. ATI researchers have disagreed over whether general ability or specific abilities will interact more strongly with treatment (Cronbach and Snow, 1977). In fact, the same group of researchers has argued both sides. Cronbach and Gleser (1965) originally argued that specific aptitudes, such as verbal or quantitative ability, would have more utility because they would match the context of the treatment more closely. Cronbach apparently was forced to change his mind, however, when confronted with the evidence summarized in Cronbach and Snow (1977). The evidence suggested that when interactions occurred, they were most likely to occur with general ability. However, Snow (1978) modified this perspective by suggesting that the match with treatment determines whether general or specific abilities are more salient. This latest idea seems to suggest that the nature of the task might influence the focus of ability construct. General ability may predict adequately for many knowledge learning tasks; however, specific aptitudes may be appropriate for very specific tasks. For example, it seems that eye-hand coordination and spatial ability may be more predictive of success in learning and playing baseball than general cognitive ability.

The following two hypotheses are based on the deduction by Cronbach and Snow (1977); however a finding of no difference may be attributable to the nature of the current task.

H_2: As suggested by Cronbach and Snow, general ability (G) will be more predictive of performance after training than specific abilities (S) such as verbal comprehension and numerical ability.

$$R^2_G > R^2_S$$

H_3: As suggested by Cronbach and Snow, the interaction of general aptitude with training will be stronger than the interaction of specific aptitudes with training.

$$Beta_{GT} > Beta_{ST}$$

Finally, the issue of cost should be considered. Cronbach and Gleser (1965) suggested that adaptive treatment would be more beneficial to the organization than fixed treatment in which every individual is treated the same. Their arguments suggested the following hypothesis:

H4: The utility of adaptive treatment will be greater than the utility of fixed treatment in the experimental setting.

$$U_A > U_F$$

CHAPTER 4: METHOD

Overview

Teachers in educational settings often adapt their instructional methods to meet the needs or match the abilities of their students. As shown in the meta-analysis, students' learning or performance levels can be increased when teachers accurately identify how to adapt their methods to meet the individuals' needs. The effect of adapting treatments to individuals' aptitudes has not been effectively tested in training situations. The way to do this is (1) to measure participants' aptitudes, (2) to allocate participants into two training treatment groups, and (3) to examine criterion test scores after training to see how participants' aptitudes interact with treatment to affect performance.

In this research project, a miniature training course was patterned after the 13-week Basic Tax Preparation Course offered by a national tax preparation service. Participants were assessed for general ability, numerical ability and verbal comprehension. Half of the participants were randomly assigned to a self-paced treatment using the training materials, while the other half were assigned to an instructor-supported treatment using the same training materials. Performance was measured using an examination based on the tests used by the tax preparation service. This general research design is described in detail below.

Subjects and Tasks

Subjects

Subjects were recruited through advertising in the free campus newspaper, The State News, and through announcements in management courses seeking volunteers to participate in a miniature tax preparation training course. To guard against attrition and to increase motivation, all subjects who completed the entire project and scored an 80% or above on the exam shared in a $1200 cash prize. Subjects were recruited in this fashion in order to reproduce some of the conditions found in the basic training course that was being simulated (Fromkin & Streufert, 1976). In particular, the cash prize induced a similar level of externally-provided motivation in that it reflected the reward opportunity for the tax service's real trainees. Those trainees who take the Basic Course and perform 80% or above on the final test are eligible to be considered for a tax preparation job. In addition, subjects who were recruited from management courses received extra credit points if they completed the entire project. The extra credit points were offered to encourage participation. This could represent a departure from the conditions found in the criterion setting, in that the personal interest of the subjects in the tax training could be less than the personal interest of Basic Course trainees. However, most subjects had several different opportunities to earn extra credit points but chose this project, presumably because of their interest in the topic area.

Cronbach and Snow (1977) lamented the traditionally inadequate sample sizes in ATI research and suggested that 100 subjects per treatment would be a good rule of thumb for an adequate test of ATI. However, a more precise estimate of the power of the test to detect significance can be found using techniques suggested by Cohen and Cohen (1983). The appropriate analysis for ATI requires the use of moderated or hierarchical regression (Cronbach and Snow, 1977). In this analysis, the statistic of interest is the one-tailed t-test of the regression coefficient associated with the interaction term in the equation.

Assuming, as suggested by the results of the meta-analysis, the standardized regression coefficient for the interaction equals .19, in a complete regression equation that explains a moderate amount of variance (e.g., $R^2 = .30$), 140 subjects would provide a power of .80 at an alpha level of .05. To account for attrition, 160 subjects were recruited.

Task

Cronbach and Snow (1977) noted that the best way to test for adaptive treatment or ATI was to use a controlled experiment format. This project followed that recommendation by creating a test largely under the experimenter's control. However, as Cronbach and Snow also pointed out, the goal of ATI research is ultimately practical, so such experiments should also be designed to increase external validity.

The training course was modeled after the Basic Tax Preparation Course offered by the tax service. The Basic Course was chosen

because it can be taught within the context of a controlled experiment and because the tax service uses it to train its own tax preparers. The Basic Course materials included an instruction manual, homework exercises, quizzes and tests. These materials were supplemented by lectures and interaction with a trained instructor who provides instructional support and extra information beyond the basic learning material. Because of time limitations and the tax needs of the subjects, only part of the material (Chapters 1-6) was presented.

This training format provided an opportunity to test the three hypotheses that treatments that vary the amount of instructional support interact with aptitudes in determining performance. Two treatments were devised which differed in the amount of instructional support. The low instructional support treatment consisted of the instruction manual. These materials inherently provide little instructional support. The high instructional support treatment consisted of the instruction manual, the workbook and a trained instructor who conducted a highly structured class covering the material.

It is important to note that the integrity of the high instructional support treatment is highly dependent on the behavior of the instructor. The instructor was instructed to follow the format typically used by the tax service. This format included several activities that were repeated at each class session. First, the instructor reviewed exercises in the workbook and answered any questions on material presented in previous class sessions and

chapters. Second, the instructor explained the material presented in the textbook and sometimes provided supplemental examples or illustrations. Lastly, the instructor answered questions on the new material and suggested some exercises for practice.

The behavior of the instructor was evaluated in more detail through the Teacher Behaviors Inventory (TBI) which was completed by 72 of the subjects who attended the tax seminar. The TBI (Murray, 1983) consists of 60 items which describe specific, observable behaviors with a 5 point frequency scale that ranges from almost never to almost always. Murray (1983) found a median inter-rater reliability of .76 across the items. The survey items and descriptive statistics for the TBI are shown in Table 6.

Murray (1983) conducted a principal axis, varimax rotation factor analysis on the items of the TBI and found 9 factors. These factors and the percent of variance explained included Clarity (20.6%), Enthusiasm (12.3%), Interaction (9.2%), Task Orientation (5.9%), Rapport (5.0%), Organization (4.6%), Use of Media (4.3%), Pacing (3.8%) and Speech (3.4%). Based on average ratings within the factors, the instructor was described as being clear, providing organization and having a task orientation. She had a high amount of enthusiasm and good rapport with the trainees as a group but had little personal interaction with the trainees.

Table 6. Descriptive Statistics for Teacher Behaviors Inventory (TBI)

Teacher Behavior: Scale runs from 1 - almost never to 5 - almost always		Mean	S. D.	Factor[1]
1.	Speaks slowly	2.8	.97	
2.	Moves about while lecturing	2.8	.91	E
3.	Uses concrete examples	4.0	.68	C
4.	Uses headings and subheadings	3.0	1.01	T
5.	States own viewpoint on issues	3.6	.94	
6.	States teaching objectives	3.3	.90	O
7.	Shows concern for students	3.1	.91	C
8.	Encourages questions and comments	3.3	1.06	I
9.	Speaks in monotone	2.4	1.06	I
10.	Gestures with hands and arms	3.0	.97	E
11.	Repeats difficult ideas	3.3	.89	C
12.	Explains how each topic fits in	3.4	.92	O
13.	Uses humor	3.5	.98	E
14.	Advises students about tests	2.3	.92	T
15.	Tolerant of other viewpoints	3.3	1.02	R
16.	Asks questions of individual students	1.6	.85	I
17.	Stutters, mumbles or slurs words	1.4	.57	S
18.	Avoids eye contact with students	1.6	.80	E
19.	Uses graphs and diagrams	2.8	1.18	C
20.	Puts outline of lecture on board	1.8	.97	O
21.	Suggests practical applications	3.5	.90	U,C
22.	Provides sample exam questions	2.0	1.11	
23.	Addresses students by name	1.4	.85	I
24.	Asks questions of class as whole	2.2	1.00	I
25.	Speaks expressively or emphatically	3.5	.71	E
26.	Shows distracting mannerisms	1.7	.77	R
27.	Stresses important points	3.8	.77	C
28.	Signals transition to new topic	3.8	.91	T
29.	Shows strong interest in subject	4.4	.66	C
30.	Digresses from topic of lectures	2.0	.92	T
31.	Offers help with problems	3.0	1.10	
32.	Praises students for good ideas	2.4	1.08	I
33.	Speaks softly	1.9	.87	E
34.	Smiles or laughs	4.1	.83	E
35.	Suggests mnemonic aids	3.0	.96	C
36.	Summarizes periodically	3.4	.85	P
37.	Uses variety of media	2.4	.98	U
38.	Dwells on obvious points	2.4	.89	T
39.	Talks with students after class	3.2	1.08	I
40.	Fails to take initiative in class	1.9	.99	C
41.	Speaks loudly	3.9	.77	S
42.	Shows energy and excitement	4.0	.81	E

Table 6 (cont'd)

Teacher Behavior: Scale runs from 1 - almost never to 5 - almost always	Mean	S. D.	Factor[1]
43. Writes key terms on board	3.4	1.02	P
44. Gives preliminary overview of lecture	3.1	.99	O
45. Reads lecture verbatim from notes	2.4	1.20	E
46. Proceeds at rapid pace	3.3	1.04	T
47. Sensitive to student needs	3.0	.84	R
48. Provides opportunity for participation	2.9	.94	I
49. Speaks clearly	4.3	.66	S
50. Relaxed and confident	4.1	.80	
51. Uses audiovisual aids	3.4	1.30	U
52. Presents thought-provoking ideas	2.8	1.06	I
53. Sticks to point in answering	3.7	.84	R
54. Friendly, easy to talk to	3.9	.77	R
55. Shows facial expressions	4.1	.72	E
56. Gives multiple examples	3.5	1.00	C
57. Relates subject to current events	3.9	.93	U
58. Criticizes students for errors	1.5	.91	
59. Speaks rapidly	3.0	1.17	
60. Uses big words	2.0	.78	

[1] Abbreviations are as follows:

C - Clarity
E - Enthusiasm
I - Interaction
T - Task Orientation
R - Rapportability
O - Organization
U - Use of Media
P - Pacing
S - Speech

Variables

Tests

Copies of the measures for all variables are found in Appendix C.

Verbal comprehension and numerical ability. The contribution of two specific aptitudes, verbal comprehension and numerical ability, to achievement were assessed because of their prevalence in prior research and because of their relevance to the training and job performance of a tax preparer. The tests used were from the Employee Aptitude Survey (Ruch & Ruch, 1980). The Employee Aptitude Survey (EAS) which consists of ten tests measuring ten different dimensions from verbal comprehension to manual speed and accuracy was designed for practical application in business settings. The EAS can be administered to a group, has a short testing time, and is easy to administer, score and interpret (Schneider and Schmitt, 1986).

The tests for verbal comprehension and numerical ability were used to assess these specific aptitudes. The verbal comprehension test was designed to assess individuals' ability to use words in written and oral communication (Ruch & Ruch, 1980). The test consists of 30 items in which examinees essentially identify synonyms. Ruch and Ruch (1980) reported that the verbal comprehension test has an equivalent forms reliability of .85. They also provided evidence of the construct validity of the verbal comprehension test. This test correlated .85 with the verbal subtest of the Primary Mental Abilities Test and .75 with the verbal

subtest of the Cooperative School and College Ability Test. In addition, the verbal comprehension test and nonverbal tests such as mechanical comprehension and quantitative subtests were not highly correlated. Finally, Ruch and Ruch (1980) demonstrated the criterion-related validity of the verbal comprehension test by providing the validity coefficients for over 30 samples for a total sample size of 3089. Across a variety of criteria (e.g., supervisory ratings, training success and ranking), the sample-size weighted average correlation was .28. The verbal comprehension test has also been used in several published validity studies. Guion (1965) found that the verbal comprehension test was marginally predictive of leadership. Grimsley and Jarrett (1973; 1975) found it was useful in discriminating between top and middle managers where top managers had higher verbal comprehension scores than middle managers. Tenopyr (1969) reported that it correlated .29 with salary (corrected for age and seniority) as a measure of managerial success. In the current sample, the coefficient alpha reliability was .75.

The numerical ability test consists of three parts with 25 items each. It was designed to measure skill in addition, subtraction, multiplication and division (Ruch & Ruch, 1980). The equivalent forms reliability is .87. The construct validity is demonstrated by a correlation of .51 with the Number test of the Primary Mental Abilities Test and .53 with the Bennett Test of Mechanical Comprehension. Discriminant validity is suggested by a low (.10) correlation with the verbal subtest of the Cooperative

School and College Ability Test. In addition, Lunneborg and Lunneborg (1967) demonstrated its convergent validity as correlations between the numerical ability test and several math and numerical tests ranged from .65 to .75. They also demonstrated its discriminant validity as correlations between the numerical ability test and various verbal tests ranged from .15 to .49. Finally, the sample-size weighted correlation coefficient of the numerical ability test with a variety of criteria based on the over 35 samples (for a total sample size of 3343) described by Ruch and Ruch (1980) is .36. In addition, Grimsley and Jarrett (1973; 1975) found point biserial correlations between numerical ability and level of management (top versus middle) of .42 and .44. In the current sample, the coefficient alpha reliability was .92.

The criterion-related validities for verbal comprehension and numerical ability can be compared to those reported in the meta-analysis by Schmitt et al (1984). Across all criteria they found a sample-size weighted correlation coefficient for special aptitudes of .27. They found a weighted average correlation of .16 with performance ratings and of .28 with achievement or grades. Therefore, the correlations reported by Ruch and Ruch (1980) and others are similar to those reported by Schmitt et al (1984) for a large number of special aptitudes and measures.

General intelligence

The general intelligence test used was the Wonderlic Personnel Test (Wonderlic, 1983). The Wonderlic Personnel Test (WPT) is a 12 minute, 50 item test of adult intelligence which has been used as a

selection test by a large number of firms for many years. (Schneider & Schmitt, 1986).

The psychometric properties of the WPT are within generally acceptable ranges. The test-retest reliabilities of the WPT have ranged from .82 to .94 (Wonderlic, 1983); and, in a recent assessment of the reliability, Dodrill (1983) found a test-retest reliability =.94. Dodrill (1981) assessed the construct validity of the WPT as a measure of general intelligence by correlating the WPT with the Wechsler Adult Intelligence Scale (WAIS) for 120 normal adults in principal and cross-validation samples. The two tests correlated .93 for the principal sample and .91 for the cross-validation sample. In addition, criterion-related validation studies show that the WPT is a significant predictor of job performance. The Wonderlic Personnel Test Manual cites 22 published and unpublished criterion-related validity studies which can be summarized using meta-analytic techniques (Hunter, Schmidt & Jackson, 1982). The sample size weighted average correlation between WPT scores and a variety of performance criteria (primarily supervisory ratings) is equal to .39 with a confidence interval from .34 to .44. In comparison, Schmitt, Gooding, Noe and Kirsch (1984) found an average correlation of .25 when they used meta-analytic techniques to calculate the sample size weighted average validity for a variety of general mental ability tests across different criteria. In addition, Jensen (1977) conducted item analyses and found no evidence of culture bias between blacks and whites in large, representative samples.

There is one major limitation associated with the Wonderlic--it is inappropriate to calculate verbal or numerical subscores despite the broad range of problem types (Wonderlic, 1983). The WPT is designed with the range of problem types and structured so the problems become increasingly difficult. To attain a high score on the WPT, an individual must have a high level of several abilities and must be able to manage them well in problem-solving. The general intelligence construct, therefore, is understood as this ability to manage many abilities (Wonderlic, 1983).

Aptitudes

The measures of general, verbal and numerical ability were confirmed by a factor analysis of the verbal comprehension test, the numerical ability test and the general ability test. As shown in Table 7, the first factor of a principal components factor analysis with orthogonal rotation consisted primarily of the Wonderlic, the general ability test, and only partially of the specific tests. The second factor reflects the EAS-2, the numerical test; while, the third factor reflects the EAS-1, the verbal test.

Table 7. Results of Principal Components Factor Analysis of Aptitude Tests

	Factor Loadings for Rotated Factor Matrix		
Factors Aptitude Test	I	II	III
Wonderlic	.91417	.27628	.29659
EAS-2 (Numerical)	.23440	.97193	.02005
EAS-1 (Verbal)	.25550	.01941	.96661
Variance Explained	58.9%	30.1%	11.0%

The SPSS-X factor analysis program was used to generate standardized factor scores for each individual (SPSS Inc, 1986). Each individual then had three ability scores derived from the factor analysis: a general ability factor score, a numerical ability factor score and a verbal ability factor score. Since they were standardized, the ability scores had a mean of 0 and a standard deviation of 1 across the sample. This had the advantage that each individual's score was a deviation score, as suggested by Cronbach and Snow (1977). These standardized factor scores were used for all analyses.

Performance

A learning test consisting of 40 items was given to a control group (n=31) that did not encounter the tax material in any way to estimate a baseline. The average score was 15 correct out of 40 (37.5%) with a standard deviation of 4.00. The coefficient alpha of this test was .37. The performance of this group who all appeared to be guessing indicated that training was needed in order to perform adequately on the test and that additional items might increase the reliability of the test. As a result, 10 items were added to the final test. The coefficient alpha reliability of the 50 item test for the experimental treatment groups was .87.

The 50-item measure of performance or learning was based on the material presented in the textbook and consisted of questions generated by three sources. The first set (14 items) was taken directly from the standardized exam the tax service gives training participants after completion of the first ten chapters of the Basic

Course. The second set (22 items) consisted of true-false, quiz and final exam questions that were re-written into a multiple choice format. The third set (14 items) were new, multiple choice questions written by the researcher in order to maximize the reliability of the test. Those items were evaluated and approved by the instructor of the course. An individual's score on this test consisted of the number of questions answered correctly.

Reaction to Training

Attitudes toward and satisfaction with the instructional materials were measured using a questionnaire, shown in Table 8 with basic descriptive statistics. The coefficient alpha reliability of this measure was .70. This measure consisted of 12 items which were derived from the evaluation form used by the tax preparation service. A thirteenth item measured subjects' overall evaluation of the training materials. This item was not related to the 12 item measure ($r=-.04$, $p=.36$), possibly due to the ambiguity of the new scale or to the change of scale.

Instruction time and time spent studying

Cronbach and Snow (1977) criticized previous research for failing to control for differences in instructional time or time spent studying under various treatments. They recommended controlling the experiment so that instructional time was kept constant across treatments; however, they noted that in training research this was rarely feasible. Rather, they suggested that instructional time should be measured and used as a control

Table 8. Basic Demographics of Evaluation of Instructional Materials

Evaluation Items: Scale runs from 1 - Strongly disagree to 5 - Strongly agree	Mean	S. D.
1. I learned what I expected to learn.	3.9	.58
2. The material presented was relevant to me.	4.1	.54
3. The material presented was valuable to me.	4.2	.60
4. The text was written clearly.	4.1	.68
5. The text was well-organized.	4.1	.52
6. The text presented the tax material in a logical sequence.	4.0	.62
7. The examples were helpful illustrations of the material.	4.1	.60
8. The text is written in a comprehensible fashion.	4.0	.59
9. The text is written in an interesting fashion.	3.7	.64
10. The text presented the tax material in sufficient detail.	3.9	.51
11. The tables enhance the overall presentation of the material.	3.8	.59
12. I have mastered the subject matter of this course.	3.8	.66
13. Considering all your previous answers, use the scale below to rate the instructional materials overall.		
1 - very poor to 5 - excellent	3.6	.82

Coefficient alpha for all 13 items = .65
Coefficient alpha for first 12 items = .70.
Correlation between scale of first 12 items with summary item (13) = -.04 (p=.36).

variable. In order to assess instructional time a record was kept of class attendance for those individuals in the instructional support treatment. Time spent studying was measured using a grid provided by Entwistle and Entwistle (1970). Subjects were asked to keep track of the time they spent studying over the three weeks of the course. As it is conceivable that some subjects completed the grid only at the end of the course, a convenient subsample was contacted to report their previous day's studying time. On November 12, 1987, 50 subjects indicated how much time they had spent studying the previous day, November 11. Of these 50 subjects, 41 completed the experiment and turned in a tax diary two weeks later. Of these 41, 88% provided the same information about November 11 at both times. Klein (1987) used this grid, although he tested it somewhat differently. He found that the subsample's responses were not significantly different from the responses for that day from the full sample.

The assessment of instruction time and time spent studying also served as a manipulation check. Individuals who had been assigned to the instructional support treatment were expected to attend class; however, a few individuals did not attend regularly or at all. The integrity of the treatment might have been jeopardized if highly irregular attenders or non-attenders were analyzed with others in the instructional support treatment. Four individuals who were assigned to the instructional support treatment and completed all the measures did not attend any of the instructional sessions. The experience of these individuals was more similar to the self-

paced treatment than to the instructional support treatment. It seemed appropriate to reassign them to the self-paced treatment. Analyses were conducted both ways--first, with these four individuals in the self-paced treatment and second, in the instructional support treatment. The differences were negligible; therefore, these individuals were reassigned to the self-paced treatment. Only results based on this reassignment are reported.

Procedure

Subjects were recruited through the campus newspaper during October, 1987. They registered for the training class in advance. An insufficient number of individuals responded to the advertisements, so subjects were also recruited from two management classes. Individuals who responded to the advertising and students in one of the management classes were assigned to the instructional support treatment. Individuals in the other management class were assigned to the self-paced treatment. Initially 160 subjects were assigned to one of the treatments with 58 subjects in the self-paced treatment and 102 subjects in the instructional support treatment. It was expected that more individuals would drop out of the instructional support treatment because of schedule conflicts.

The training course was conducted the first three weeks of November, 1987. It consisted of an introductory session, five 2-hour instructional sessions (for those in the instructional support treatment) and a final session. In the introductory sessions (conducted separately for the two treatments), the experimenter explained the training course and research project, obtained

subjects' informed consent, gathered aptitude data, and distributed training materials. The instructional sessions were conducted by a paid and experienced instructor provided by the tax service who explained text information, reviewed workbook problems and answered questions as described above. At the final session, all subjects took the exam, completed the final questionnaire and turned in time spent studying grids. Within a week, the exams were graded and the cash prizes were awarded to those who completed the project and scored an 80% or above on the exam.

Analysis

Descriptive statistics including means, standard deviations, internal consistencies and intercorrelations among all variables were calculated and are shown in Table 9. The basic model for analysis of the major issues was moderated regression (Cronbach & Snow, 1977; Stone & Hollenbeck, 1984; Zedeck, 1971). Aptitude was entered into the regression equation first. Cronbach and Snow (1977) recommended that individuals' aptitude scores be entered as deviations from the mean score on the aptitude across all subjects. This was satisfied since the scores were the standardized factor scores derived from the factor analysis. A dummy variable for treatment (+1 for instructional support and 0 for self-paced) was entered second. In the third step, the cross-product of the aptitude score and the treatment was entered into the equation. Using this procedure with unstandardized regression coefficients provided information about the independent effects of aptitude, treatment and the interaction. Cronbach and Snow (1977) strongly

Table 9. Intercorrelations for Scales and Variables

Variable	General Ability Factor	Numerical Ability Factor	Verbal Ability Factor
General Ability	1.00		
Numerical Ability	.00	1.00	
Verbal Ability	.00	.00	1.00
Treatment	.19**	.08	.03
Hours	.16*	.12	-.04
Seminar Attendance	.20**	.07	.02
Reaction	.01	.15	.01
Learning	.37**	.31**	.26**
Wonderlic	.91**	.26**	.26**
EAS-1 (Verbal)	.22**	.12	.96**
EAS-2 (Numerical)	.22**	.97**	.10

Table 9 (cont'd)

	Treatment	Hours	Seminar Attendance	Reaction	Learning
Treatment	1.00				
Hours	.61**	1.00			
Seminar Attendance	.93**	.64**	1.00		
Reaction	.13	.14	.06	1.00	
Learning	.58**	.44**	.54**	.11	1.00
Wonderlic	.20**	.17*	.27**	.06	.51**
EAS-1	.08	.003	.18*	.01	.36**
EAS-2	.12	.16*	.12	.15	.40**

	Wonderlic	EAS-1	EAS-2
Wonderlic	1.00		
EAS-1	.51**	1.00	
EAS-2	.49**	.18*	1.00

N = 118

* $p < .05$
** $p < .01$

recommended that analysis should focus on unstandardized regression coefficients in order to see the influence of ATI in standard deviations. The regression equation in which measures are standardized, however, provided information about the strength of the effects in terms that are comparable across units of measure. This procedure was used to test the first hypothesis--that the absolute value of the standardized regression coefficient for the interaction term is greater than zero. The t test for that regression coefficient (or the F test for the incremental change in R^2) will test whether the interaction is significantly greater than zero.

The unstandardized regression equations were graphed in order to identify the nature of the interaction, whether it is ordinal or disordinal, and to assess which process suggested by Salomon (1972) might explain the interaction, testing hypotheses 1a and 1b.

It is important to note that cost considerations influence the practicality of the interaction. However, the practicality or utility of adaptive treatment in this situation cannot be directly assessed because of the complexity of the payoff function. Such estimates can be made only after more levels of instructional support (i.e., more treatments manipulating instructional support) are tested for their payoffs.

Separate regression equations were generated for general ability (factor scores based on the first factor that emerges from a principal components factor analysis as described earlier) and each of the specific aptitudes (the second and third factors) separately.

To test H2, confidence intervals were constructed around the cross-validated (shrunken) coefficients of determination ($\tilde{R}^2$) to indicate whether general or specific aptitudes predict performance differently. The confidence intervals were constructed as follows:

$$\tilde{R}^2 - 1.96\ (sd) < \tilde{R}^2 < \tilde{R}^2 + 1.96\ (sd)$$

$$\text{and } sd = \{[4\tilde{R}^2\ (1 - \tilde{R}^2)]^2/n\}^{1/2}$$

If the confidence intervals overlap then they do not measure different amounts of variance in performance; if the confidence intervals do not overlap and if $\tilde{R}^2$ for performance using general ability is greater than $\tilde{R}^2$ for performance using specific abilities, then general ability explains more variance. It is important to note that there is not a generally accepted test for testing the difference between two $\tilde{R}^2$s. The use of confidence intervals is one of several approximate tests that can be used.

To test H3, the standardized regression coefficient for the interaction between general ability and training was compared to the standardized regression coefficient for the interaction between verbal comprehension and training and between numerical ability and training using confidence intervals in the same manner as above.

Testing H4, that the utility of adaptive treatment is greater than that of fixed treatment, was more complicated and exploratory. The procedure that would have been followed is described below. However, as discussed below, this hypothesis was not tested because of the results.

Cronbach and Snow (1977) suggested a method to compare the outcome of adaptive treatment to putting everyone in one treatment

when only two treatments have been investigated. They proposed that utility depends on the strength of the interaction and the distance of the cross-over point on the interaction (where the outcome for a particular level of ability is the same regardless of which treatment is experienced) from mean ability. They assumed that the standard deviation of the outcome (performance) equalled 1.0 and that the adaptive treatment offered preferential treatment for those with high ability, so in their case, the regression slope of the preferential treatment (A) was greater than the regression slope of the original treatment (B). They expressed the change in utility per person (ignoring cost) if treatments are adapted to individual aptitudes by the following equation:

$$\text{Change in Utility per person} = (B_A - B_B)\,\xi$$

where

B_A is the regression coefficient for the preferential treatment

B_B is the regression coefficient for the original treatment

ξ is the ordinate of the normal distribution at the crossover between the regression equations for A and B

This equation needs to be modified in two ways to assess the change in utility per person for the current situation which would provide self-pacing as a type of preferential training for higher aptitude trainees. First, an estimate of the value of performance is needed. This estimate is provided by the standard deviation of performance in dollars. In this case the standard deviation can be estimated using procedures devised by Schmidt, Hunter, McKenzie and Muldrow (1979) and amended by Bobko, Karren & Parkington (1983)

where supervisors provide estimates of the value of job performance for tax preparers at the 15th, 50th and 85th percentiles. As suggested by Bobko et al, supervisors will first estimate the 50th percentile. These estimates will be averaged and the average will be fed back to the supervisors who then estimate the 15th and 85th percentiles. The differences between the 50th and 15th percentiles and between the 85th and 50th percentiles yield estimates of the standard deviation of performance. This estimate may be verifiable through archival records since tax preparers at the tax service are paid by commission. Second, the relationship between training success and performance must be added in order to make the linkage between performance improvements in the training course and performance on the job. The relationship between training success and performance would be expressed best in the correlation coefficient estimated from a study of the effect of the tax preparer training on actual job performance. Unfortunately such a study has not been performed at the tax service. However, an estimate of the correlation coefficient derived from a meta-analytic summary of training can be used. In their meta-analysis, Hunter and Hunter (1984) found a correlation of .28 between training success and overall ratings of performance for a large number of studies.

$$\text{Change in utility per person} = (B_A - B_B)(r)(\xi)(\sigma_e)$$

where

B_A is the regression coefficient for the preferential treatment

B_B is the regression coefficient for the original treatment

ξ is the ordinate of the normal distribution at the crossover

r is the correlation between training success and performance

σ_e is the standard deviation of performance in dollars.

This formulation does ignore costs; however, the use of the preferential treatment actually decreases the cost per person since instructors will not have to be paid to help the higher aptitude trainees.

CHAPTER 5: RESULTS

Description of Sample

Of the 160 subjects who responded to the recruiting, 118 completed the training and the learning test. The instructional support treatment in which subjects received materials and a training class had 69 individuals complete the training and learning test (68% of those recruited); whereas, in the self-paced treatment, 49 individuals completed the training and learning test (84% of those recruited). These figures indicate that there was more attrition than expected. However, as shown in Table 10, the ability test scores of those who completed the training were not significantly different from the scores of those who did not complete the training. Thus, the attrition did not alter the critical characteristics of the sample.

Table 10. Comparison of Subjects for Differential Attrition

	Failed to Complete Training		Completed Training		
Variable	Mean	SD	Mean	SD	
Self-paced Instruction Treatment					
Score on EAS-1 (Verbal)	16.9	5.7	17.1	3.5	
Score on EAS-2 (Numerical)	46.8	12.5	46.6	9.9	
Score on Wonderlic	25.0	4.0	24.1	4.3	
Instructional Support Treatment					
Score on EAS-1 (Verbal)	19.7	4.0	18.6	4.2	
Score on EAS-2 (Numerical)	44.3	12.8	49.2	11.2	(p=.06)
Score on Wonderlic	25.7	5.2	27.0	4.8	

The demographic characteristics of the whole sample (n=118) indicated that most (93%) were enrolled in a degree program and 58% were juniors in college while 36% were seniors. As would be expected, then, 33% were 20 years old; 36% were 21; while 23% fell between the ages of 22 and 30. There were a few more females (56%) than males (44%). Most (74%) were business majors. These characteristics indicate that those subjects who completed the training had primarily been recruited from the management classes.

Subjects also provided information about their family income and tax experience. Fifteen percent indicated they had family income below $10,000 annually; whereas, 39% indicated their family income exceeded $50,000. These figures may be misleading, however, since it was not ascertained whether students included their parents' income or just their own. Forty-four percent of the subjects indicated that they had never completed their own tax return or that other individuals had completed their returns for them. Almost as many subjects (43%) had filed their own returns at least twice or had helped other people complete their returns. Most people (59%) filed the simplest tax form--1040EZ.

Descriptive statistics on the whole sample for the major variables are shown in Table 11. The near-zero means and near-one standard deviations on the ability factors reflect the standardization procedure utilized by the factor analysis in generating factor scores for the subjects. Individual scores therefore, deviate from zero and reflect deviations from the mean as suggested by Cronbach and Snow (1977).

Table 11. Descriptive Statistics for Whole Sample

Variable	Mean	St. Dev.
Score on EAS-1 (Verbal)	18.00	4.00
Score on EAS-2 (Numerical)	48.22	10.79
Score on Wonderlic	25.92	4.84
General Ability Factor[1]	.02	1.03
Numerical Ability Factor[1]	-.06	.96
Verbal Ability Factor[1]	.07	.96
Hours Spent on Course	10.43	5.41
Learning Score	29.37	8.27
Evaluation of Materials	3.97	.28

[1]Ability factors are standarized such that means are zero and standard deviations are one.

n = 118

As shown in Table 12, the breakdown of the major variables by treatment reflects several anticipated results. First, mean attendance for subjects who were in the instructional treatment was four out of five sessions. In addition, most subjects attended at least three of the five. This is reflected in hours spent on the course in that individuals who attended the seminars spent almost twice as much time (13.2 hours) on the course than those who studied on their own (6.5 hours). The tendency for individuals in the instructional support treatment to spend more hours on the course was also suggested by the high correlation (.61) shown in Table 10 between treatment and hours. Finally, on the average, individuals who completed the instructional treatment scored significantly higher answering 33 of 50 questions correctly on the learning test compared to those who studied on their own who answered 24 of 50 questions correctly.

One result shown in Table 12 was not anticipated: there was a significant difference between the treatments on the average score on the general ability factor. This appears to be primarily a function of small, but significant differences on the average performance of the EAS-1 (Verbal) and the Wonderlic tests in which the average score for the instructional support treatment was greater than for the self-paced treatment.

As a follow-up to this unexpected result, comparisons were made between the groups on their motivation to learn. In a questionnaire administered prior to training, subjects completed a set of 6

Table 12. Descriptive Statistics by Treatment[1]

Variable	Self-paced Mean	SD	Instructional Mean	SD
Score on EAS-1 (Verbal)*	17.0	3.4	18.7	4.3
Score on EAS-2 (Numerical)	46.3	10.6	49.6	10.8
Score on Wonderlic*	24.3	4.5	27.1	4.7
General Ability Factor*	-.27	1.0	.23	1.0
Numerical Ability Factor	-.03	1.0	.14	1.0
Verbal Ability Factor	-.21	.9	.05	1.0
Hours Spent on Course*	6.5	4.3	13.2	4.3
Seminar Attendance*	0.0	0.0	4.0	.8
Learning Score*	23.7	5.9	33.4	7.3
Evaluation of Materials	3.9	.3	4.0	.3
Motivation to Learn*	3.6	.6	4.0	.4

* Significant Difference between Treatments (alpha < .05)

[1] Self-paced treatment (n=49); Instructional support treatment (n=69)

questions derived from a scale by Ryman and Biersma (1975) assessing motivation to learn which had a coefficient alpha reliability of .76. On a scale of 1 (strongly disagree) to 5 (strongly agree), subjects in the low instructional support treatment had a mean of 3.6 (sd = .57) while subjects in the high instructional support treatment had a mean of 4.0 (sd = .44), which was significantly different ($p<.01$). This indicated that as a group the low instructional support subjects were not as motivated to learn the tax information.

These differences reflect the failure of randomization to minimize subject differences between treatments on general ability and motivation to learn. Randomization had been attempted by randomly assigning management classes to treatments, assuming that students were randomly placed in the classes through the registration process. However, the evidence is inconsistent with this assumption.

Hypotheses

Moderated Regression

The results of two regression equations generated for each ability factor are shown in Tables 13 to 19. Several patterns emerge across all the equations. First, the main effects for hours, ability factor and treatment are consistently significant for all ability factors and account collectively for a moderate percentage of the variance (36-39%). Second, hours and treatment appear to be redundant predictors of learning as suggested by the high (.61) correlation between them. This is apparent in that the omission of

Table 13. Moderated Regression on Learning for General Ability Factor Controlling for Hours

Step	Variable Entered	Multiple Correlation	Shrunken R^2	Increment in R^2	p for Increment in R^2
1	Hours	.44	.19	.19	< .001
2	General Ability	.54	.28	.10	< .001
3	Treatment	.64	.39	.12	< .001
4	Ability X Treatment	.65	.40	.01	.053

Table 14. Moderated Regression on Learning for General Ability Factor not Controlling for Hours

Step	Variable Entered	Multiple Correlation	Shrunken R^2	Increment in R^2	p for Increment in R^2
1	General Ability	.37	.13	.14	< .001
2	Treatment	.62	.38	.25	< .001
3	Ability X Treatment	.64	.39	.02	.043

Table 15. Moderated Regression on Learning for Numerical Ability Factor Controlling for Hours

Step	Variable Entered	Multiple Correlation	Shrunken R^2	Increment in R^2	p for Increment in R^2
1	Hours	.44	.19	.19	< .001
2	Numerical Ability	.51	.25	.07	< .01
3	Treatment	.64	.39	.15	< .001
4	Ability X Treatment	.64	.39	.004	.176

Table 16. Moderated Regression on Learning for Numerical Ability Factor not Controlling for Hours

Step	Variable Entered	Multiple Correlation	Shrunken R^2	Increment in R^2	p for Increment in R^2
1	Numerical Ability	.31	.09	.09	< .001
2	Treatment	.63	.39	.30	< .001
3	Ability X Treatment	.64	.39	.006	.141

Table 17. Moderated Regression on Learning for Verbal Ability Factor Controlling for Hours

Step	Variable Entered	Multiple Correlation	Shrunken R^2	Increment in R^2	p for Increment in R^2
1	Hours	.44	.19	.19	< .001
2	Verbal Ability	.52	.26	.08	< .001
3	Treatment	.62	.37	.12	< .001
4	Ability X Treatment	.62	.37	.002	.266

Table 18. Moderated Regression on Learning for Verbal Ability Factor not Controlling for Hours

Step	Variable Entered	Multiple Correlation	Shrunken R^2	Increment in R^2	p for Increment in R^2
1	Verbal Ability	.26	.06	.06	< .01
2	Treatment	.61	.36	.30	< .001
3	Ability X Treatment	.61	.36	.001	.302

Table 19. Full Regression Equations for Learning with Ability Factors

Equation	Ability	Treatment	Interaction	Constant
		Regression Coefficients		
		General Ability Factor		
Unstandardized	.86	8.7	2.07	23.95
Standardized	.11	.53	.19	

R = .64. Cross-validated R^2 = .39. Change in R^2 associated with interaction = .02 (p=.043). Standard Error for Beta on Interaction = .11

		Verbal Ability Factor		
Unstandardized	1.07	9.35	.72	23.94
Standardized	.12	.56	.07	

R = .61. Cross-validated R^2 = .35. Change in R^2 associated with interaction = .001 (p=.3). Standard Error for Beta on Interaction = .13

		Numerical Ability Factor		
Unstandardized	1.44	9.23	1.37	23.76
Standardized	.17	.55	.12	

R = .64. Cross-validated R^2 = .39. Change in R^2 associated with interaction = .006 (p = .2). Standard Error for Beta on Interaction = .11

hours from the equation did not significantly reduce the predictive value of the regression. Third, treatment appeared to have a greater influence on learning than ability. This is reflected in the incremental changes in R^2 associated with entering ability and treatment. The incremental change in R^2 associated with entering ability ranged from .06 to .14 while the incremental change in R^2 associated with entering treatment ranged from .25 to .30 (in equations not controlling for hours). Lastly, the interaction of ability and treatment had the weakest effect on learning across ability levels.

Hypothesis 1

The standardized regression coefficients for the interaction terms are shown in Table 20. The largest coefficients are those associated with the general ability factor. These coefficients (.18 and .19) are significantly different from zero as indicated by the one-tailed test's significance levels (.053 and .043, respectively) associated with the increment in R^2 given in Tables 13 and 14. The standardized regression coefficients of the comparable interaction terms for the numerical and verbal ability factors are substantially smaller and not significantly different from zero.

Hypotheses 1a and 1b

Hypotheses 1a and 1b predicted the interaction of ability and treatment would be due to different processes. Hypothesis 1a predicted a disordinal interaction due to the compensatory nature of instructional support; while, hypothesis 1b predicted an ordinal

Table 20. Standardized Regression Coefficients for Interaction Terms

Equation Controlling for Hours			Equation Not Controlling for Hours	
Ability	Beta	S.E.	Beta	S.E.
General	.18	.11	.19	.11
Numerical	.11	.11	.12	.11
Verbal	.08	.13	.07	.13

interaction due to the preferential nature of instructional support. These hypotheses could only be evaluated for the general ability factor since interaction terms on the other factors were not significantly different from zero. The unstandardized regression equations for each treatment for general ability (not controlling for hours) were graphed, as shown in Figure 5. The unstandardized regression weight for ability on performance in the high instructional support treatment is significantly different from zero at the .05 level; whereas, the unstandardized regression weight for ability on performance in the low instructional support treatment is not significantly different from zero at the .05 level. The results are consistent with hypothesis 1b in that the slope of the line for high instructional support is greater than the slope of the line for low instructional support and the lines do not intersect.

Hypothesis 2

Hypothesis 2 predicted that general ability would be more predictive of performance after training than specific abilities. This hypothesis was tested by comparing the percent of variance

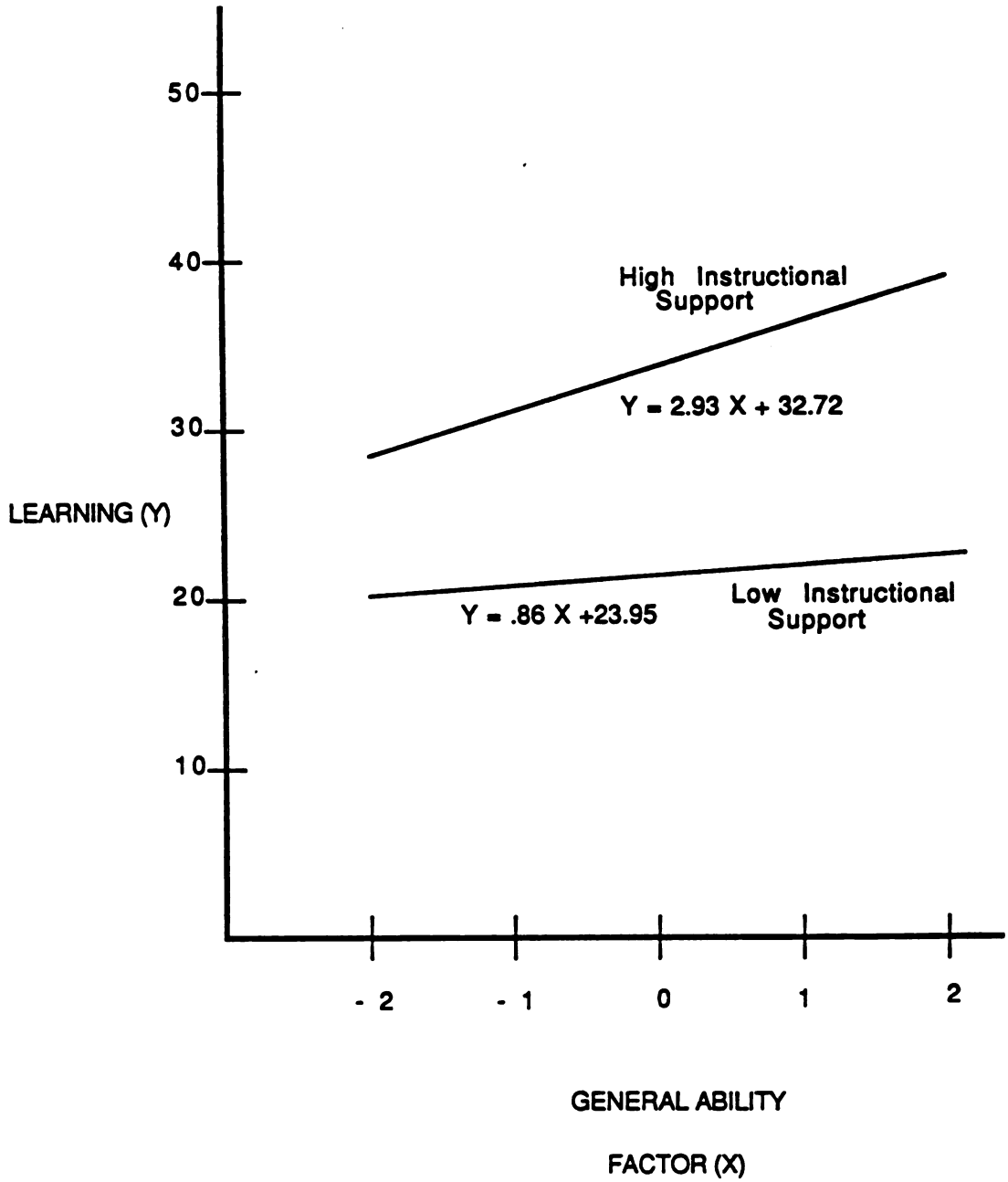


Figure 5. Graph of Ordinal Interaction of General Ability Factor and Treatment on Learning

explained (shrunken R^2) by the total regression equations for general versus specific abilities. The percent of variance explained and the coefficients of determination are given in Table 21. Using the equations given in Chapter 4, the standard deviation for shrunken R^2 is .088. It is apparent from looking at Table 21 that the confidence intervals around the shrunken R^2s would all overlap, indicating that they are not significantly different from each other. This finding is inconsistent with the hypothesis, in that both specific and general abilities predict learning to the same extent.

Table 21. Coefficients of Determination for Regression Equations

	Equation Controlling for Hours		Equation Not Controlling for Hours	
Ability	Multiple R	Shrunken R^2	Multiple R	Shrunken R^2
General	.64	.394	.65	.399
Numerical	.64	.392	.64	.393
Verbal	.61	.350	.62	.368

Hypothesis 3

Hypothesis 3 predicted that the interaction of general ability with training would be stronger than the interactions of specific abilities with training. As is apparent from looking at Table 20, confidence intervals using the standard errors around the standardized regression coefficients would all overlap. This finding is also inconsistent with the hypothesis in that the

strength of the interactions are not significantly different from each other.

Hypothesis 4

Hypothesis 4 was based on Cronbach and Gleser's (1965) assertion that the utility of adaptive treatment would be greater than the utility of fixed treatment. As discussed above, a number of treatments at different levels of instructional support are needed to test this hypothesis appropriately, although Cronbach and Snow (1977) suggested a method to estimate the difference in utilities with only two treatments. Their method estimates the difference in utilities from two bits of information: the strength of the interaction and the distance of the cross-over point for the interaction from mean ability. Apparently, this method is dependent on the assumption that the interaction is disordinal. Unfortunately, the interaction found in this training situation is ordinal; therefore, this method cannot be used.

These analyses are difficult to apply because ability was derived from factor scores; thus, the analyses were conducted a second time using actual test scores on the Wonderlic, EAS-1 (Verbal), and EAS-2 (Numerical). The full moderated regression equations based on the tests themselves, shown in Table 22, are totally consistent with the results based on the factor scores. For example, the strength of the effect on learning of the interaction of treatment and the Wonderlic (.21) is almost identical to the strength of the effect on learning of the interaction of treatment and the general ability factor (.19).

Table 22. Full Regression Equations for Learning with Ability Tests

Equation	Ability[1]	Treatment	Interaction[1]	Constant
		Regression Coefficients		
		Wonderlic		
Unstandardized	-.34	8.15	-.48	24.27
Standardized	-.20	.49	-.21	

R = .69. Cross-validated R^2 = .46. Change in R^2 associated with interaction = .02 ($p=.028$)[2]. Standard Error for Beta on Interaction = .11

		EAS-2 (Numerical Ability)		
Unstandardized	-.15	8.96	-.15	24.01
Standardized	-.20	.54	-.15	

R = .68. Cross-validated R^2 = .43. Change in R^2 associated with interaction = .009 ($p=.08$)[2]. Standard Error for Beta on Interaction = .11

		EAS-1 (Verbal Ability)		
Unstandardized	-.36	8.92	-.23	24.06
Standardized	-.18	.53	-.09	

R = .63. Cross-validated R^2 = .38. Change in R^2 associated with interaction = .003 ($p = .25$)[2]. Standard Error for Beta on Interaction = .13

[1] Sign of Ability and Interaction terms is negative because scores were entered as deviations from the mean.

[2] Significance test is based on a one-tailed test.

Summary

The results indicate that for these two training treatments there is a significant ordinal interaction between general ability and treatment such that the relationship between general ability and learning is stronger in the high instructional support treatment than in the low instructional support treatment. The interactions of treatment with specific abilities were not significant. In addition, in this setting, the overall predictability of general ability was not significantly greater than for specific abilities; and, the strength of the effect of interaction of general ability with treatment on learning was not significantly different from the strength of the effects of the interactions for specific abilities.

The dependent variable in the stated hypotheses was learning; however, learning is only one outcome of training (Kirkpatrick, 1967). Subjects' reaction to the training is another relevant outcome that was assessed by surveying subjects' evaluation of the instructional materials and of the amount of learning they acquired. Reaction may have been influenced by the same variables as learning, ability, treatment and the ATI, as a number of ATIs have been found to influence individuals' attitudes (Cronbach & Snow, 1977). However, regressions assessing the influence of these variables indicated that none of them was a significant predictor of reaction. This is probably due to the finding, as shown in Tables 11 and 12, that there was little variance in reaction to predict since the mean

and standard deviation for the self-paced treatment were 3.9 and .3 while the mean and standard deviation for the instructional treatment were 4.0 and .3.

CHAPTER 6: DISCUSSION

The results and implications of this investigation warrant further consideration. First, this consideration summarizes and compares the results to the major goals of the study, including estimation of the practical implications of adapting treatments and evaluation of the contribution of the results to the ATI literature. The results are used to propose a taxonomy of instructional support that could guide hypothesis generation in future ATI investigations. Second, the study needs to be evaluated for its internal validity by considering alternative explanations for the results and for its external validity by considering issues related to the generalizability of the results. Finally, several issues to be investigated in future research are discussed.

Summary and Implications of Results

The primary goal of this study was to assess the utility of adaptive treatment by testing whether aptitudes and treatments interact in determining performance. Two strategies were pursued to meet this goal. The first was to estimate the utility of adaptive treatment in the preparation of tax preparers by determining the strength of an ATI. The second was to estimate the costs and benefits of different treatment strategies. If the utility of adaptive treatment favors this approach, the application of the results could enhance the productivity of workers, increase attention paid to the individual employee and ease the impact of the impending labor shortage of skilled or professional workers.

The first strategy pursued the strength of an aptitude-treatment interaction. Interactions were estimated for three different abilities--general, numerical and verbal--in two different treatments manipulating instructional support. Only one interaction, general ability by treatment, was significant (p=.043). This interaction reflected a standardized slope difference of .19 between the regression equations of ability on achievement for the two instructional support treatments. This slope difference is not trivial; however, as further investigation revealed, it is not just the strength of the interaction that impacts utility, but also the nature of the interaction.

The nature of the interaction was tested by investigating two competitive approaches concerning how ability and treatment might interact to influence performance. The first, the compensatory approach, predicted that instructional support would interact with ability such that high instructional support would aide low aptitude individuals and raise them to the same level of performance as high aptitude individuals. The nature of this interaction is disordinal with the intersection of the separate regression lines for the treatments at high aptitude, high performance. The second, the preferential approach, predicted that both low and high aptitude individuals would respond to instructional support, but that high aptitude individuals would respond more. The nature of this interaction is ordinal.

The results were consistent with the preferential approach, indicating that the interaction of general ability and treatment was

ordinal. This finding appears to imply that the high instructional support treatment is better for everyone, regardless of aptitude, and that adapting treatments to individuals' aptitudes is unnecessary.

However, Cronbach and Gleser (1965) and Cronbach and Snow (1977) argued that ordinal interactions may still imply that providing adaptive training treatments could benefit the organization. The determination of the utility in such a situation depends on the balance of selection (in terms of the minimum cutoff), of the costs of treatments and testing, and of the differential gains in performance. As shown below, the results of the current study suggest performance may be maximized if the adaptive treatment strategy involves selecting high aptitude individuals and putting them into the high instructional support treatment.

The evidence for this strategy consists of several points derived from Figure 5, the graph of the ordinal interaction of general ability and treatment on learning. First, the positive slope of the regression lines indicating that ability is a significant predictor of performance suggests that high ability individuals should be selected. Average performance will increase as the ability levels of selected individuals increases. For example, if high ability individuals (average ability = +2 standard deviations) are selected and experience the high instructional support treatment, then their expected performance on the learning measure is, on average, 39 correct answers out of 50. On the other

hand, if low ability individuals (average ability = -2 standard deviations) are selected and experience the high instructional support treatment, then their expected average performance on the learning measure is 27 correct answers out of 50, a reduction of 24%.

Second, the graph indicates that high ability individuals would benefit more than low ability individuals from experiencing the high instructional support treatment. For example, high ability individuals (average ability = +2 standard deviations) would get an average of 13 extra answers correct (a gain of 26%) if they experience the high instructional support treatment instead of the low instructional support treatment. In contrast, low ability individuals (average ability = -2 standard deviations) would only get an average of 5 extra answers correct (a gain of 10%) if they experience the high, instead of the low, instructional support treatment. Similarly, high ability individuals who are +1 standard deviations from the mean would experience a 22% increase in correct answers (11 extra) ; whereas, low ability individuals who are -1 standard deviations from the mean would experience only a 14% increase in correct answers (7 extra).

These estimates can be used to highlight how organizations may experience a gain in utility from adaptive treatment, despite an ordinal interaction. The gain depends in its simplest form on whether the net gain in performance as reflected in the increases in correct answers has a greater value to the organization than the costs of testing. For example, the costs of tests used to screen

out low aptitude individuals must be less than the gain in performance for training only high aptitude individuals. In this situation, high aptitude individuals perform 12 questions better than low aptitude individuals in the same (high instructional support) treatment. The value of this 12 question gain per person should be greater than the cost of testing each person.

Utility also depends on the determination of the cutoff for selection and the differential gain in performance from the different treatments for selected subgroups, balanced against the costs of testing and of treating subgroups differently. For example, if the cutoff for selection is -1.1 standard deviations for the mean, then the gain in performance for low aptitude individuals (-1 standard deviations) who experience different treatments can be compared to the gain in performance for high aptitude individuals (+2 standard deviations) who experience different treatments. Low aptitude individuals only gain 7 correct answers if they experience the high instructional support treatment; whereas, high aptitude individuals gain 13 correct answers from the high instructional support treatment. The 13 point gain for high aptitude individuals may be worth the cost of the treatment and testing; however, the 7 point gain for low aptitude individuals may not be worth the cost. If this is the case, then the organization would place high aptitude individuals into the high instructional support treatment and either reject low aptitude individuals or search for another treatment to increase their performance.

Specific recommendations for adaptive treatment cannot be made in this, or any, situation without information on the costs of testing and treatment. Regardless, the evidence argues strongly for selection since ability is predictive of performance and since the high instructional support treatment has a greater payoff in performance for high ability individuals. Whether the situation calls for adaptive treatment depends on more information than is currently available. However, the possibility is not remote that utility may be enhanced by adapting treatments.

Hypothetical Utilities

The findings regarding the strength and nature of the interaction are only suggestive that there may be utility in adapting treatment in the current situation by offering two treatments, one for each ability group, since the interaction was ordinal. However, it is interesting to speculate what the utility of adaptive treatment would be if the strength of the interaction was the same, while the nature of the interaction was disordinal.

Table 23 shows what the standard deviation of performance would have to be for different increases in utility given (1) the interaction found in this test (.19), (2) the average strength of the relationship between training success (learning test) and performance (.28) based on Hunter and Hunter's (1984) meta-analysis and (3) various cross-over points between treatment equations. For example, in order for the increase in utility per person for adapting treatments to be $100 when the cross-over is at mean ability, the standard deviation of performance would have to be

Table 23. Hypothetical Differences in Utility

Standardized Slope Difference	Correlation b/n Training & Performance	Difference in Utility Per Person	Standard Deviation of Performance
	Cross-over of Interaction at Mean: ξ = .399		
.19	.28	$ 100	$ 4,711
.19	.28	$ 200	$ 9,422
.19	.28	$ 500	$23,555
.19	.50	$ 100	$ 2,638
.19	.50	$ 200	$ 5,276
.19	.50	$ 500	$13,190
	Cross-over of Interaction at +1 Standard Deviation: ξ = .242		
.19	.28	$ 100	$ 7,767
.19	.28	$ 200	$15,534
.19	.28	$ 500	$38,836
.19	.50	$ 100	$ 4,350
.19	.50	$ 200	$ 8,699
.19	.50	$ 500	$21,749
	Cross-over of Interaction at +2 Standard Deviation: ξ = .054		
.19	.28	$ 100	$34,809
.19	.28	$ 200	$69,618
.19	.28	$ 500	$174,046
.19	.50	$ 100	$19,493
.19	.50	$ 200	$38,986
.19	.50	$ 500	$97,465

$4711. But for the increase in utility per person to be $500 in the same situation, the standard deviation of performance would have to be $23,555. Given this situation, it is much less likely that adapting treatments would lead to a $500 gain in utility per person. As the cross-over deviates from the mean, the required standard deviation in performance increases to totally unrealistic levels (e.g., $174,046).

However, it should be noted that part of the reason for the limited benefit of adapting treatments given the current information derives from the fact that the dependent variable was training success and not job performance and that the relationship between training success and job performance was only moderately strong (r=.28). In other situations, that relationship may be stronger. Table 23 also contains estimates of the required standard deviation in performance for situations in which the correlation between training success and job performance is stronger (.50). The results indicate that with this stronger relationship, the required standard deviation in performance is almost half as large as with the weaker relationship (.28).

Application of Results

As discussed, the data were consistent with an ordinal interaction predicted by the preferential approach. Although the positive utility of adapting treatments was only suggested by the results, it is interesting to speculate on the effects of the findings.

It has been proposed that adapting treatments would have several positive outcomes: enhancing productivity, focusing attention on individuals and easing the labor shortage. The results of this study suggest that productivity is enhanced by treating all trainees similarly, i.e., putting them in a highly supportive instructional environment; although, high ability trainees respond more to this supportive environment than low ability trainees. Therefore, the best strategy is to select high aptitude individuals for the high instructional support treatment since their gain in performance is greater than the gain in performance for low aptitude individuals.

The finding that high ability trainees responded more to the supportive instruction than low ability trainees implies that the impact of the impending labor shortage will not be lessened by hiring more low ability employees. For example, if organizations try to increase the performance of low ability employees by providing extra instructional support, they may achieve only partial success. Organizations will fill vacancies, but will not fill them with high performers. Average performance under the labor shortage will be less than under current, adequate supply conditions.

Application of the findings indicates that performance may be enhanced somewhat, in that low ability individuals, if selected and trained, will increase their performance to some extent. However, treating all trainees similarly will not increase attention on individuals or significantly reduce the impact of the labor shortage.

Comparison of Meta-analytic Results and Empirical Results

The training study was designed to compensate for and to extend findings in the previous literature as represented by the meta-analytic results. This invites a comparison of their results.

Effect sizes. The standardized regression coefficients associated with the interactions not controlling for hours in the training study reflect the same interaction calculated in the meta-analysis (.19). A comparison of these effect sizes indicates that they are consistent with each other. The average weighted regression coefficient across the studies in the meta-analysis was .19 with a confidence interval from .04 to .34. The standardized regression coefficients found in the training study ranged from .07 for verbal ability to .19 for general ability, fully within the range found in the meta-analysis. These similarities reflect consistency in the estimates of the strength of ATIs' effects on achievement.

Nature of instructional support. Although their standardized regression coefficients indicate the same strength for the interaction of ability and treatment, the meta-analysis and the training study appear to reflect different types of interactions. As discussed earlier, in 9 of the 14 studies in the meta-analysis, the regression coefficient associated with the low instructional support treatment is greater than that associated with the high instructional support treatment. In contrast, in the training study, the result was just the opposite. The regression coefficient

associated with high instructional support was greater than that associated with low instructional support.

These conflicting results primarily suggest that the type of instructional support may influence the nature of the interaction. Two different ways to provide instructional support--remedial information and teaching aides--were discussed earlier. However, the results suggest that there may be many more ways to provide instructional support. For example, some manipulations of high instructional support involved a student-centered approach in which individuals progressed through material at their own pace (e.g., Ross & Rakow, 1981). Other manipulations involved the addition of learning aides to reading material, such as attention directing devices or supplementary questions (e.g., Threadgill, 1979; Winn, 1981).

The possibility that there are a number of ways to provide instructional support which may in turn affect the type of ability and treatment interaction suggests that the nature of instructional support needs to be investigated more thoroughly. A taxonomy of instructional support treatments would provide a structure that would help in predicting interactions. The absence of a taxonomy limits the interpretation of the results of the meta-analysis somewhat since types of instructional support might be an important moderator that cannot be tested at present.

The taxonomy would also help to distinguish between high and low instructional support, a difference which is sometimes difficult to discern from descriptions of studies. For example, in one study,

such as the present empirical study, an instructor provides high instructional support because he or she supplements reading material by providing examples, reiterating important points and answering questions. In another situation, an instructor provides low instructional support as he or she lectures without interacting with trainees or students.

The difficulty in distinguishing between instructional support manipulations highlights how critical it is to identify what constitutes instructional support and how important it is to describe the instructor's behaviors carefully. It also provides support for the procedure used in the meta-analysis to estimate the strength of the interaction of accumulating the absolute values of the slope differences. These issues illustrate how the nature of the interaction is highly dependent on the types and levels of instructional support treatments. Because of its practical and theoretical contributions, a proposed taxonomy based on the results of Murray's (1983) factor analysis of the TBI is described below.

General and specific ability. The meta-analysis and the training study yielded essentially similar results comparing the effects of the interactions of treatment with general and specific abilities on achievement. In both, the confidence intervals around the standardized regression coefficients for the interaction of general, verbal and numerical abilities with treatment overlapped, indicating that the coefficients were not significantly different from each other. On the other hand, the strength of the interactions was significantly greater than zero (except for general

ability) in the meta-analysis, but not in the training study. As discussed above, this may be due to the differences in the nature of the instructional support manipulations. The most glaring difference is that the weighted average regression coefficient for general ability by treatment in the meta-analysis is almost zero (Beta = .03, ns), while the parallel standardized regression coefficient in the training study is significantly different from zero (Beta = .19, p = .043). This discrepancy may be due to the low number of effect sizes in the general ability moderator analysis (k=3). On the other hand, it could be due to sample differences in educational level. The results of the educational moderator analysis in the meta-analysis were consistent with the proposition that the strength of the interaction effect increased as the educational level increased. The subjects in the training study were college students, who represent a highly educated population and would be expected to reflect a fairly strong interaction effect. However, the subjects in the meta-analysis who were included in the general ability moderator analysis were younger and less well-educated, since 80% (501) were elementary students and the remaining 20% (122) were high school students.

These comparisons suggest the nature of the tasks may influence the relative predictive power of general versus specific abilities. In the meta-analysis and the training study, the tasks may have required both general and specific abilities; so both were predictive of performance. In other tasks with more specific requirements, the specific ability might be most predictive while

other specific abilities and general ability are not. For example, performance in a task with high verbal requirements and no quantitative requirements (such as editing) might be predicted by the interaction of verbal ability and treatment.

Summary. The results of the meta-analysis and the training study reflect some similarities and differences. Although the basic findings regarding the strength of ATIs were consistent, they were inconsistent regarding the nature of the interactions. Some of these inconsistencies might be explained by different conditions discussed above such as nature of instructional support. However, there are other differences between educational settings and training settings that could explain the results.

First, the educational level and age of individuals undergoing training is typically higher than individuals still in school. Much of the research in education has focused on elementary and high school students; whereas, industrial training is conducted for individuals who are older and have higher educational backgrounds (high school degree and above). The finding that the strength of the interaction increases as educational level increases suggests that results in educational settings with younger subjects may not be replicated in training settings with older subjects. In addition, the possibility of a strong interaction effect in training settings due to a relatively more educated subject population increases the potential utility from adapting treatments on the basis of that interaction.

Second, training is typically conducted for a specific task or job; whereas education is focused on general tasks. The identification of an ATI for a specific task can have high and immediate utility to the organization as adapting treatment on the basis of that ATI can directly enhance performance on that task. The finding of a specific ATI in education may have less utility because its application is limited and its contribution to the overall education of the individual is small.

Third, the salience and intensity of a training experience for individuals is greater than for an educational experience, in that they know they will need to apply the knowledge they are gaining in training. It is more difficult for students to identify when and where they will apply the knowledge they are gaining in school. This relatively vague sense of relevance may introduce more random noise into the educational setting than into the training setting. This noise would enhance random error in education settings which would be reduced in training settings.

These reasons may partially explain the differences in the educational (meta-analytic) results and the training results. In addition, they provide a strong rationale for further assessment of ATIs in training settings; since, despite a great deal of research in educational settings, there may be important differences that restrict the applicability of educational findings to training situations.

A Taxonomy of Instructional Support

As discussed above, the conflicting results between the meta-analysis and the empirical study may be due to differences in the nature of instructional support. It is difficult, however, to delineate what these differences might be since no taxonomy of instructional support exists to guide classification of aspects of support. Such a taxonomy is proposed here based on the Teacher Behaviors Inventory (Murray, 1983). Future research could examine this taxonomy and use it to guide hypothesis generation.

As discussed above, the Teacher Behaviors Inventory (TBI) consists of 60 behavioral items rated in terms of their frequency of occurence. Murray (1983) conducted a principal-axis factor analysis with varimax rotation and identified 9 orthogonal factors he named: Clarity, Enthusiasm, Interaction, Task Orientation, Organization, Rapport, Use of Media, Pacing and Speech.

Although these factors are mathematically uncorrelated due to the orthogonal rotation, they appear to be conceptually related. For example, task orientation and organization may be dependent since individuals who are conscientious about sticking to the task are likely to be highly organized. In addition, people tend to attribute good rapport to others who have positive interactions with them. Based on these conceptual relationships, there may be more parsimonious factors underlying the 9 factor revealed by Murray's analysis.

First, three of the primary factors (Clarity, Interaction and Use of Media) appear to reflect a factor that could be called

Personal Support. Personal Support appears to reflect how the leader supports and interacts with the individual trainee or student by answering questions, giving examples, asking questions of individuals and calling individuals by name. Second, Task Orientation, Pacing and Organization appear to reflect Initiating Structure. Initiating structure could be described as the extent to which an instructor supports trainees or students through behaviors such as using headings or subheadings, signalling transitions, writing key terms on the board and summarizing periodically. Third, Enthusiasm, Speech and Rapport may reflect Group Dynamism. Group Dynamism reflects how the leader generally relates to the students or trainees as a class or group. This factor would be reflected in items such as using humor, speaking clearly, or showing energy and excitement.

These three factors suggest that instructional support may have three dimensions: Personal Support, Initiating Structure and Group Dynamism. As shown in Figure 6, the processes by which each of these dimensions interact with ability may differ. First, the interaction of ability with personal support may be an ordinal interaction in which the personal attention and support given by the instructor compensates for individual deficiencies. Low ability individuals are likely to gain a great deal from the personal attention as their particular learning or motivational needs and deficiencies are met; however, high ability individuals would not have as much room for improvement from the individual attention.

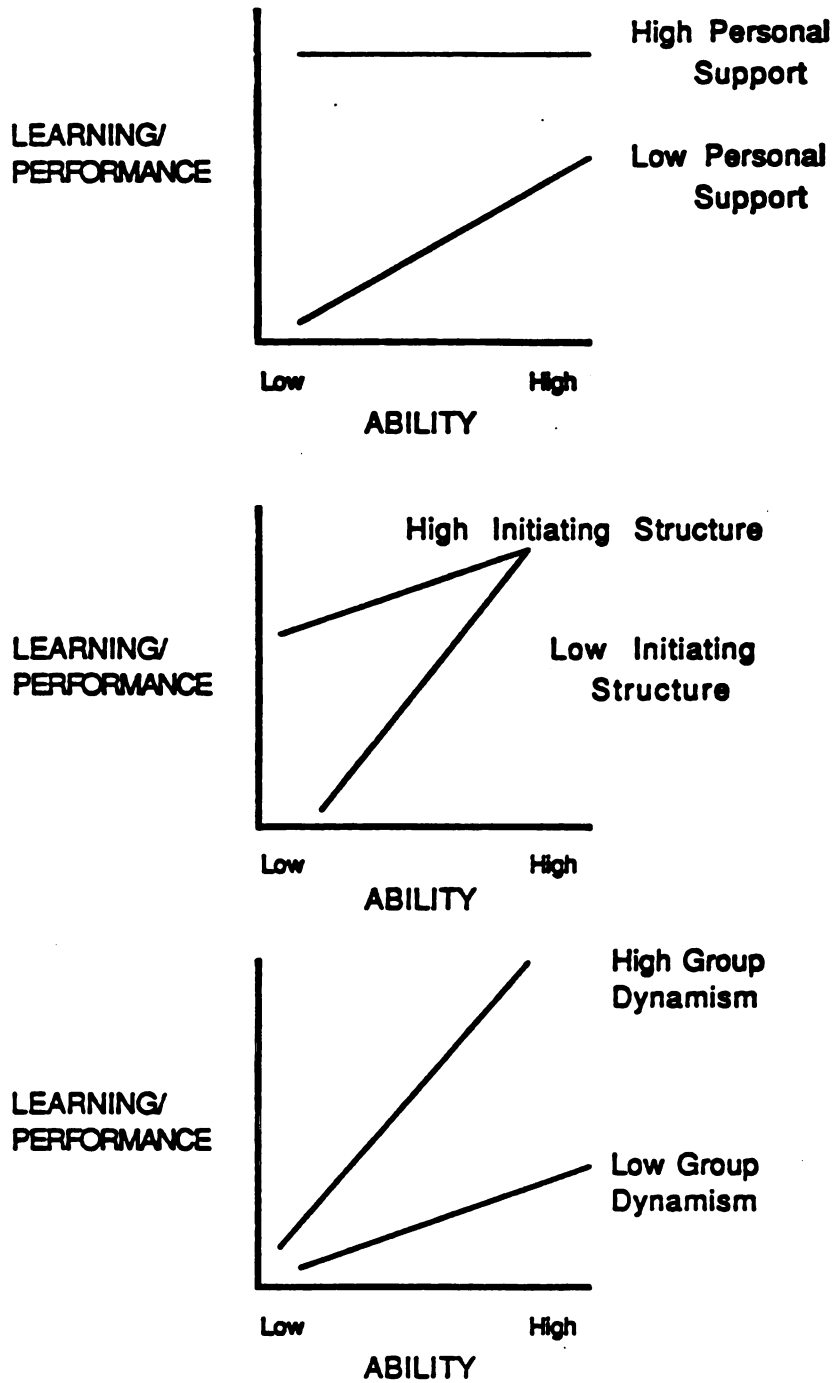


Figure 6. Hypothesized Interactions of Ability with Different Dimensions of Instructional Support

Second, the interaction of ability with initiating structure may be a disordinal interation in which the organization and structure provided increases the performance of the low ability trainees or students who cannot provide it for themselves. However, high ability individuals who can organize for themselves do not benefit from the instructional support.

Third, the interaction of ability with group dynamism may be an ordinal interaction in which the enthusiasm and group dynamics encourage and motivate high ability individuals to try harder and perform better. However, low ability individuals may try harder, too, but are not provided with enough structural support to compensate for their ability problems. In this case, both high and low ability individuals increase their performance because they are trying harder, but high ability individuals, because of their greater capabilities, experience greater performance gains.

The taxonomy can be used to analyze the results of this study. The tax preparation instructor appears to have been low on initiating structure with a mean rating on the scale of 3.3 and on personal support with a mean rating of 3.13. However, she appears to have been high on group dynamism with a mean rating of 3.8. For example, as indicated in Table 6, she used numerous examples but did not provide much organization for the trainees which on balance suggests relatively low initiating structure. In addition, she had a great deal of enthusiasm and spoke clearly, reflecting high group dynamism. However, she did not get to know the students personally,

ask them questions or get to know their particular training needs, indicating a lack of personal support.

The finding of an ordinal interaction in which high ability trainees gain more than low ability trainees is consistent with the idea that instructional support in this situation may have been due to increased group dynamism. Individuals in the low instructional support treatment did not have an instructor; therefore they had no one to encourage them to try harder. On the other hand, individuals in the high instructional support condition had an instructor who had a high degree of group dynamism. The structure was provided by the textbook for both treatments; and, neither treatment received personal support.

This taxonomy is obviously based on post hoc analyses of the results. However, it does appear to have utility in guiding future research on ATIs. Future research should test these hypotheses as well as evaluate the taxonomy.

Alternative Explanations for Results

There are methodological arguments for the failure to support the hypotheses besides the substantive ones given above. Some of these arguments are discussed below.

Power

Despite attempts to recruit enough subjects to account for attrition, only 118 of the 180 original subjects completed the training program. Given the results for general ability without controlling for hours (R^2=.39, Beta=.19) and alpha = .05, the power to detect differences with 118 subjects was between .70 and .75.

Selection

The comparisons of the two treatment groups indicated that there were significant differences between the treatments on general ability in that mean ability for the high instructional support was greater than mean ability for the low instructional support. This reflects the failure of the randomization procedure used.

Restriction of Range

Restriction of range is often a probable explanation for results. The typical problem is that the sample has a smaller range of test scores than the population under consideration. The impact of this smaller range is to restrict variance and attenuate the strength of relationships. Although the relationships in the current study may be attenuated, there are two reasons why range restriction in the independent variables (aptitude) and one reason why range restriction in the dependent variable (learning) are probably not problems.

First, Cohen and Cohen (1983) argued that regression coefficients are less affected than correlations by range restriction. Regression coefficients are more stable because the decrease in the simple correlation is offset by an increase in the ratios of standard deviations. As Cohen and Cohen pointed out, "the fact that [the regression coefficient] tends to remain constant over changes in the variability of [independent variables] is an important property of the regression coefficient [and]...this makes them more useful as measures of relationships than correlation coefficients in some analytic contexts" (1983, p. 70). The focus of

this investigation is on regression coefficients which, as noted by Cohen and Cohen, are stable despite some restriction in range on the predictor.

Second, restriction in range does not appear to be a problem in the current study because the critical characteristics of the sample are fairly similar to assumed characteristics of the population. Restriction in range calls for a consideration of the relative size of the standard deviations in the sample and the population. The population under consideration in this investigation is applicants for tax preparation positions. Unfortunately norms for the test scores are not available for this particular position; however, norms are available in the test manuals for similar positions. For example, the standard deviation for a claims adjuster in an insurance firm for the EAS-1 (Verbal) is 4.6 (Ruch & Ruch, 1980). In comparison, the standard deviation in the current sample for the EAS-1 was 4.00. These standard deviations can be compared using Hartley's test for homogeneity of variance (Winer, 1971). The Fmax statistic is calculated as a ratio of the larger variance to the smaller variance. For the comparison for the EAS-1, this ratio equals 1.32 (ns). Similarly, the standard deviation for a claims adjuster for the EAS-2 (Numerical) is 12.1, compared to the 10.79 in the current sample. These standard deviations are also not significantly different from each other (Fmax=1.26, ns). Therefore, the ranges do not seem to be restricted in the cases where the regression coefficients may be attenuated. On the other hand, the range is restricted for the Wonderlic. The standard deviation for a

claims adjuster is 7.1, while the standard deviation for the current sample on the Wonderlic was 4.84. The Fmax statistic=2.15 which is significant at the .01 level. However, significant results were attained for general ability, and would, therefore, only be enhanced by consideration of range restriction.

Lastly, range restriction in the criterion, learning, could attenuate the relationships. However, this is assuredly not the case in the current situation, in that range was enhanced, not restricted, in this investigation. The tax preparation service allows its trainees to use their notes and books for part of their tests; whereas, the subjects in this study were not allowed to refer to any material during the learning measure. Therefore, the range is somewhat wider in the study than in the tax preparation setting.

In summary, although range restriction could be an alternative explanation for the results, it is possible to rule it out. The major reason for this is that the focal statistics in the current investigation, regression coefficients, are highly stable despite range restriction.

Manipulation Failure

The most probable explanation for the results seems to lie in problems in the low instructional support treatment. As shown by the flatness of the slope for low instructional support in Figure 5, there is little change in performance due to ability. This finding probably reflects the low motivation of subjects in that treatment. There are several pieces of evidence for this statement. First, subjects in the low instructional support treatment spent little

time on the material (average of 6 hours over 3 weeks). Second, none of the subjects in this treatment made 80% or more on the test to earn the bonus money; however, all of the subjects were from management classes and received extra credit for participating, regardless of score. Third, as discussed above, this group revealed lower motivation to learn in a questionnaire administered prior to training. This also reflects the failure of the randomization procedure.

The impact of this problem with the low instructional support treatment is hard to predict. Greater motivation to learn might increase or decrease the strength of the relationship between ability and performance. Thus the difference in slopes between high and low instructional support might increase or decrease. In addition, the nature of the interaction might be influenced as a shift in the low instructional support regression equation. If the impact of increased motivation was the same across ability levels, then the intercept would increase but not the slope and the interaction would become disordinal rather than ordinal.

Generalizability of Results

The external validity of any study needs to be evaluated (Dipboye & Flanagan, 1979). This evaluation should be conducted by looking at the boundary conditions, those critical differences between the experiment and the criterion setting, and by considering the relevance of external validity for the study.

Bracht and Glass (1968) suggested that a two-stage process is involved in external validity. First, a target population of

persons, settings or time is defined; and then, samples are drawn to represent that population. Cook and Campbell (1979) pointed out that occasionally these samples are drawn from the population with known probabilities, but that usually accidental or convenient samples are drawn which need to be evaluated for their representativeness.

The target population or criterion setting in this study involved training in an organizational setting. Because this research was essentially a laboratory study, it was designed to model the criterion setting in several critical ways: training content, instructional procedures, intrinsic interest and motivation. The training materials were the actual instructional books used by the tax service; the instructor teaches the basic tax course for the tax service. Subjects were volunteers who participated because they were interested in learning about tax procedures and because they could earn a reward for sufficient performance. This reward was contingent upon meeting the same level of performance as the rewards set by the tax service.

It appears, therefore, that this experiment exhibits some mundane realism--the degree to which the experiment reflects things that occur in the criterion setting; however, experimental realism--the degree to which the experiment captures the essence of the theoretical variables--is more important (Berkowitz & Donnerstein, 1982). The theoretical variables in this study involved aptitudes and instructional treatments while the goal of the study was to test for their interaction. Except for theoretical reasons discussed

above, such as the different types of instructional support or the influence of moderators such as educational level, it seems likely that the variables examined would operate similarly in other circumstances.

It is important to note that no experiment is totally generalizable or totally lacking in generalizability. As noted by Cook and Campbell (1979), external validity is ultimately a matter of replication. ATI studies have been replicated in educational settings, but not in training settings. As discussed above, there are critical differences between the two--educational level, task specificity and relevance--which suggests that replication and extension of test of the ATI paradigm should be conducted in training settings.

Future Research

Several issues should be investigated in future research. First, the influence of the interaction of aptitudes and treatment on other variables such as job performance needs to be investigated. This study focused on training success by evaluating the effect of ATIs on learning; however, the ultimate focus is on performance on the job after training. It is possible that differences in learning may be enhanced in job performance. Thus, the utility of adapting treatments to individuals' differences may be even greater than implied by current findings.

Second, the processes by which aptitudes interact with treatments need to be identified. This study focused on instructional support as providing teaching assistance. This focus

limited consideration to two processes--compensatory and preferential. Future studies should test the remedial process by providing instructional assistance in the form of needed background information.

Third, as suggested above, there are many ways to think about instructional support. A taxonomy of instructional support has been proposed. This taxonomy should be used to test whether the processes by which aptitudes interact with different types of instructional support are different. This taxonomy would also help in identifying which types of instructional support provide the greatest utility to an organization.

Fourth, other aptitudes need to be evaluated for the effects of their interaction with treatment on performance. These aptitudes should be chosen to match the results of a job analysis which has described the nature of the task and identified the job specifications.

Finally, Cronbach and Gleser (1965) proposed that adaptive treatments were being used more frequently and by more organizations than had been recognized. However, this proposal has never been evaluated. Organizations should be surveyed to assess the current use, success and problems of adaptive treatments. Such a survey could drive further investigations since the primary goal of adaptive treatment is to increase performance in organizations.

Conclusion

This study extended previous ATI research by investigating whether aptitudes and treatments might interact in a training setting, by identifying the processes by which ability and instructional support might influence learning, and by being conducted as recommended by Cronbach and Gleser (1965) and Cronbach and Gleser (1977) to test for the interactions. Given that support was found for some of the hypotheses, the next step is to extend these findings in various ways by taking into account theoretical issues and boundary conditions. Future research should investigate the processes underlying interactions, identify different instructional support techniques, and evaluate the utility of adaptive treatment on job performance.

APPENDICES

Appendix A

Decision-Theoretic Foundations of Adaptive Treatment

Decision theory and personnel decisions

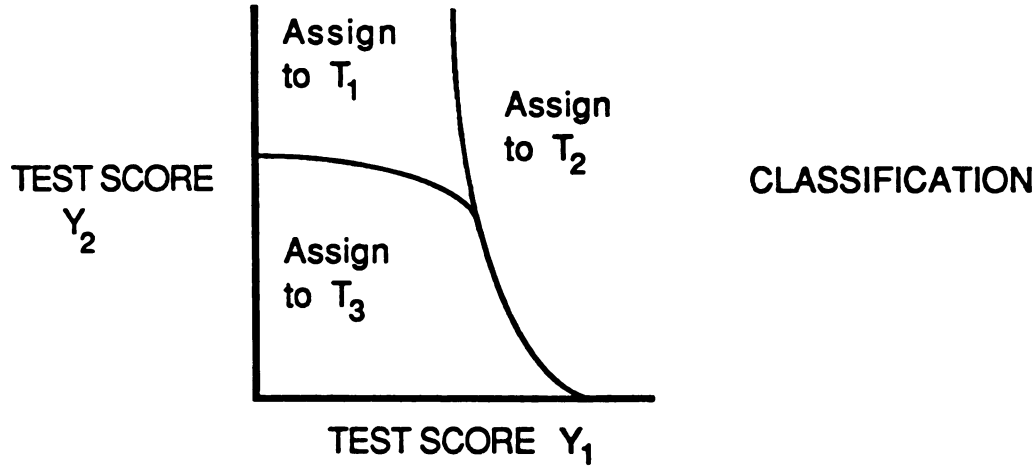
The call for consideration of adaptive treatment was issued originally by Cronbach and Gleser (1965) as part of their challenge to personnel researchers and practitioners to use decision theory to evaluate personnel decisions. To understand the reasoning behind adaptive treatment, therefore, it is important to look at it in the context of this greater challenge. This challenge was based primarily on the argument that, although many personnel procedures were based on assumptions that the person interacted with the environment or the treatment, few evaluations of those procedures were based on those assumptions. Cronbach and Gleser identified the types of personnel decisions and the processes underlying evaluation of those procedures to derive formulas to evaluate the decisions more appropriately.

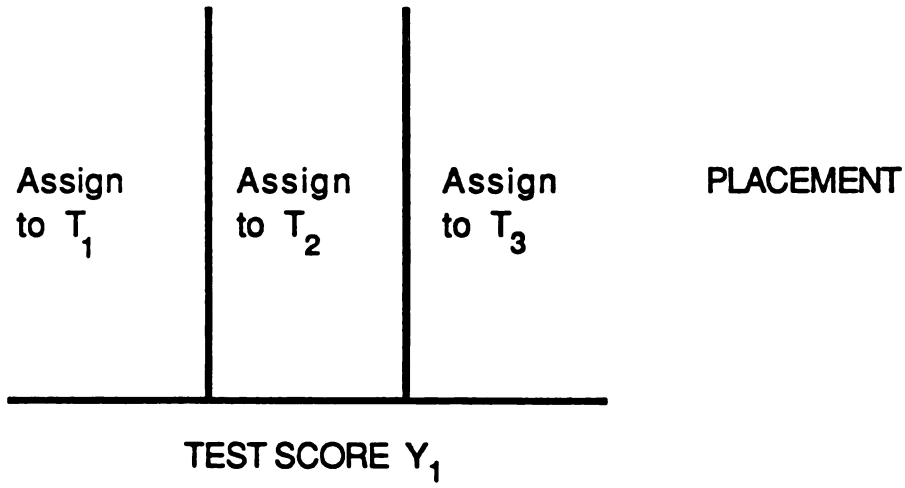
Types of personnel decisions. Cronbach and Gleser focused on institutional decisions in which people, as representatives of an institution, make a large number of comparable decisions using the same value system or a constant set of decision rules to maximize the average gain from a whole set of decisions. There is a wide variety of institutional decisions, but those that are institutional personnel decisions are attempts to decide what to do with individuals, singly or in groups. Or, to use Cronbach and Gleser's terminology, institutional personnel decisions are attempts to decide to what "treatment" each individual should be assigned, where

treatment is defined broadly and can refer to such experiences as being hired or rejected, receiving training or working on the job.

The general decision problem in personnel is one of classification, in which a decision maker must use several pieces of information to classify individuals into different treatments. This is a very complex process requiring assessment of numerous variables. There are two more specific and simple cases of classification. The first is the placement problem, in which placement of persons into treatments is based on information that consists of only one score or one dimension. The second occurs when a list of possible treatments includes rejection or elimination from the institution, then the decision problem is one of selection. Figure 7 illustrates schematically the differences among these personnel decision problems.

Characteristics of decision processes. As indicated above, all personnel decisions place an individual into one of two or more treatments on the basis of some information about the individual. However, the decision process is more complex than this. First, information about individuals must be interpreted through the use of a decision rule or strategy consisting of conditional probabilities for any possible contingency. The use of the strategy leads either to a terminal decision, in which the individual is assigned to a treatment, or to an investigatory decision, in which more information is gathered. After more information is gathered, then the strategy is used again to make a decision. The evaluation of decisions is based on the outcomes of the treatments or all the





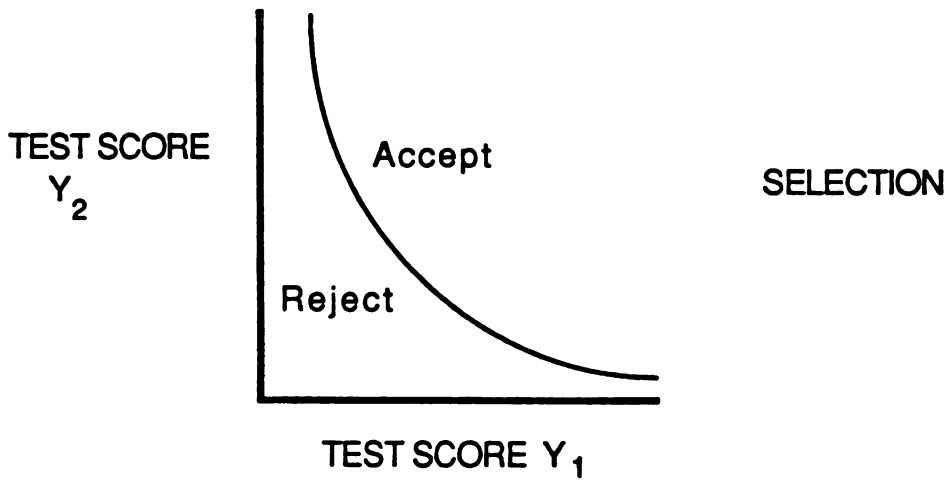
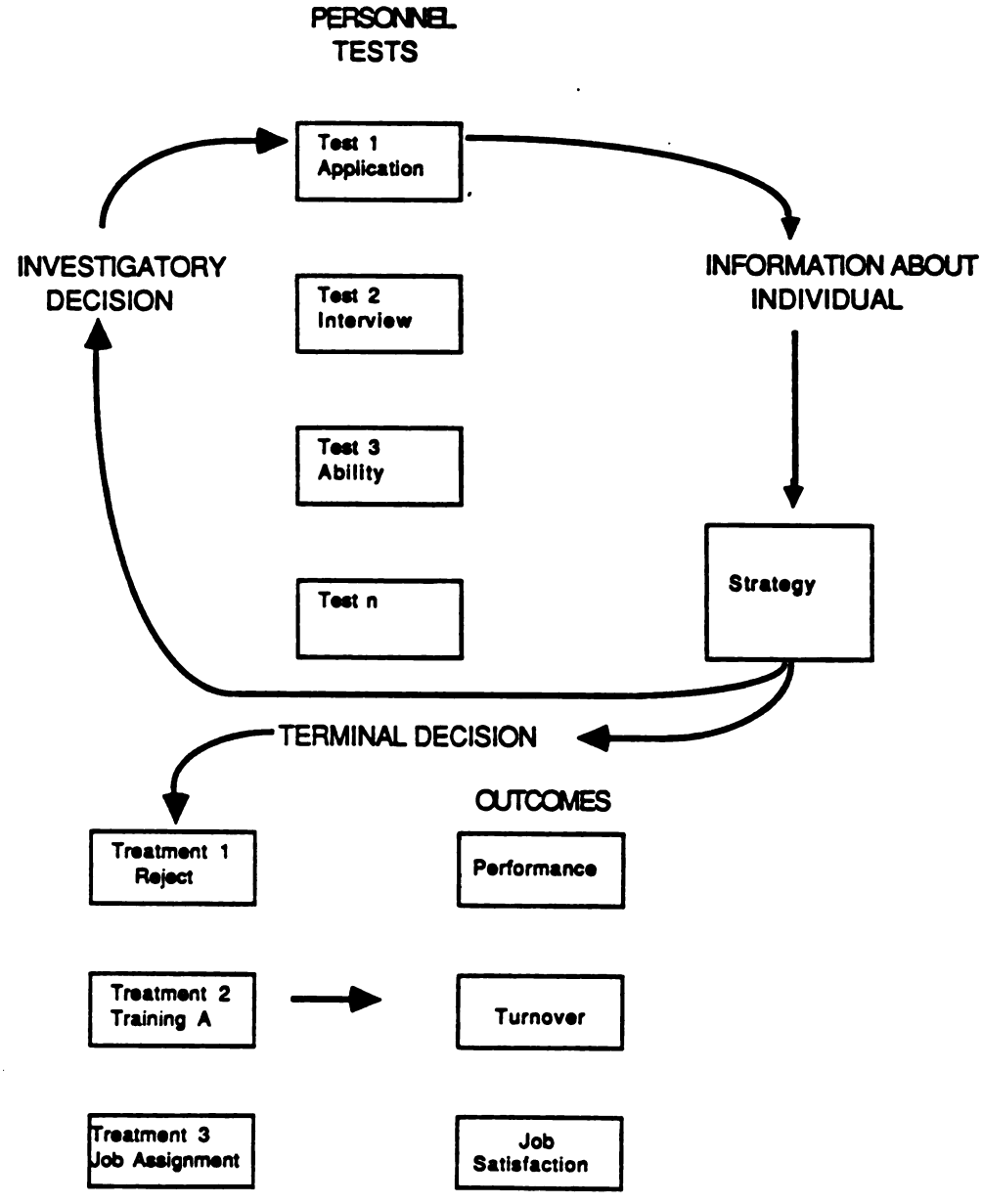


Figure 7. Personnel Decision Problems.

consequences that concern the decision maker. The focus is generally on the performance of the individual after assignment to a treatment where this performance is a function of the characteristics of the individual and the situation.

A flowchart of the decision process adapted to the specific case of personnel decisions is presented in Figure 8. In this case, the possible treatments listed include rejection, different training programs and different job assignments; and possible outcomes include job performance, turnover and job satisfaction. Information is gathered through the use of personnel tests which are defined broadly to include application blanks, interviews, references or work history as well as any cognitive ability or personality tests. Using this flowchart as an example, the process for a selection decision might proceed as follows. A decision maker, e.g., a personnel manager, evaluates test scores from applicants and estimates what performance levels and other outcomes the test results suggest about the different applicants. The manager may make terminal decisions about some of the applicants--extending job offers to individuals whose scores are sufficiently high and rejecting individuals whose scores are low. The manager may decide to seek more information on others, getting more references or setting up an extra interview, before making a terminal decision, i.e., assigning them to a treatment.

This example follows a typical selection decision process in which one group of applicants is rejected and no longer interacts with the organization and the other group of applicants is accepted



Note: Different personnel tests would be administered after the decision to investigate the applicant further.

Figure 8. Schematic View of Decision Process for Personnel Decisions.

and assigned to the same treatment. The use of only one treatment for accepted individuals is common in selection, but, as Cronbach and Gleser argue, perhaps more frequently a variety of treatments might achieve the same ends:

> Assigning men to fixed categories, or predicting their scores under a single treatment, is all that the industrial and military psychologist has attempted. But...adaptation may be possible. One may vary such important conditions of the job as amount of on-the-job instruction, amount of supervision, and pacing of work. Introducing radical changes in degree of responsibility or amount of automatic control makes changes in payoff [or outcomes] even more likely. So long as one can expect to employ men of a given quality, one should set the treatment so as to maximize their payoff. Within the limits of practicality, a change in quality of men, calls for adaptation of treatment (Cronbach & Gleser, 1965, p. 28).

Cronbach and Gleser suggest, then, that assignment to a fixed treatment is appropriate as long as there is low variance in the aptitudes of accepted applicants and as long as that particular aptitude level is the one most predictive of success in that treatment. However, if there is "a change in quality of men," or in other words, if there is variance in the aptitude levels of accepted applicants, then the expected outcome across the group will be higher if treatments are adapted to match the different aptitude levels of the individuals.

If it is appropriate to assume that there is variance in aptitude levels of applicants who would be accepted, then the original selection decision scenario needs to be revised. As Figure 8 indicates, individuals can be placed in one of several treatments--different training programs or job assignments with modified responsibilities. The decision process then becomes one of using

the test information, not just to choose employees, but also to choose treatments.

This is the situation of adaptive treatment, as opposed to fixed treatment, defined above. The process of evaluating the usefulness and accuracy of a personnel decision process, including the validation of the selection tests, is different depending on which treatment strategy, fixed or adaptive, is in operation.

Evaluation of personnel decisions. In general the decision maker bases the decision either on predictions of the probability distribution of possible outcomes or on the expected outcome over many similar decisions. These predictions or expectations are based on the results of previous cases in which the relationship between information and outcomes can be determined. This relationship is displayed in a validity matrix of conditional probabilities for a treatment with different information dimensions on one axis of the matrix and different outcomes on the other. Each outcome has a certain value (for example, the value of high job performance, X, may be greater than the value of high job satisfaction, Y) which can be combined with the conditional probabilities for each information category to estimate the expected payoff of assigning that individual to that treatment. However, expected payoff for an individual only becomes helpful to decision making when actual payoff can be averaged across individuals in the same information category and used to provide information about the usefulness of the particular decision strategy. This procedure leads to a general utility equation, assuming that the distribution of test scores for

each information category in the population tested is known, which determines the utility of the set of decisions for a treatment by adding the expected payoff for each test score weighted by the probability of that score. This equation can be written:

$$U = N \sum p_y \sum p_{t/y} \sum (p_{c/yt})(e_c) - N \sum p_y C_y$$ where:

U = the utility of the set of decisions,

N = the number of persons about whom decisions are made,

y = the information category,

t = the treatment,

c = outcome,

e_c = the value of the outcome,

C_y = the cost of gathering information,

p_y = the assumed test score distribution,

$p_{t/y}$ = conditional probabilities in a strategy matrix (the probability of being assigned to a treatment given the information from the test score), and

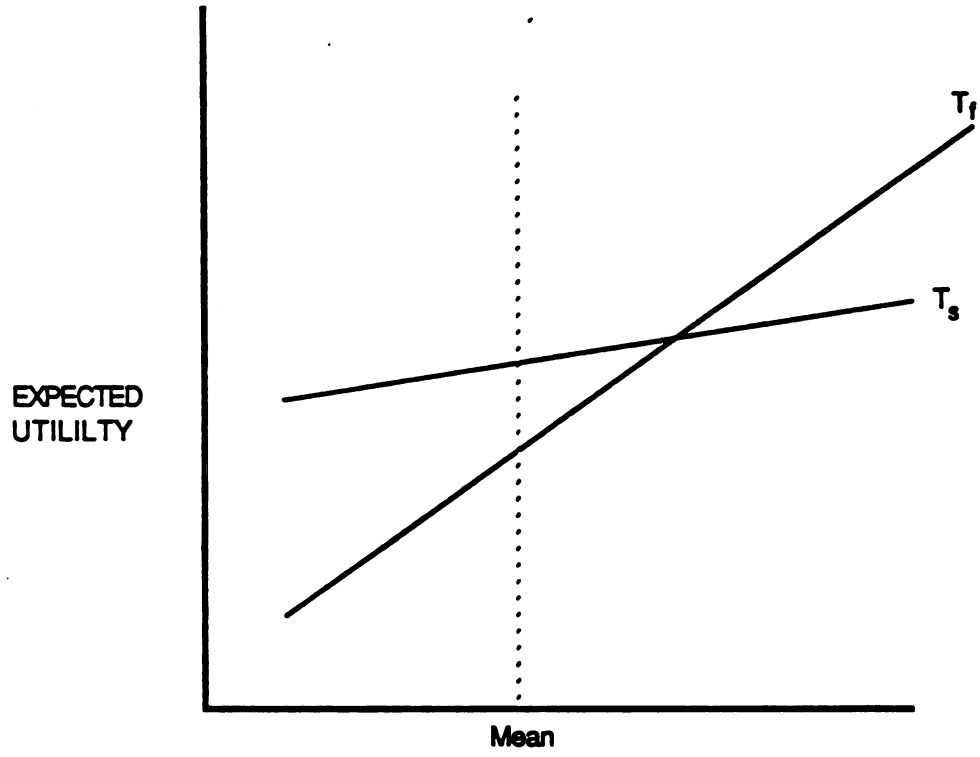
$p_{c/yt}$ = conditional probabilities in a validity matrix (the probability of achieving that outcome given the test score and the treatment).

This basic equation is modified depending on the type of personnel decision (classification, selection or placement) and type of treatment situation (fixed or adaptive). The details of calculating utility for some of these combinations are given below.

Evaluation of predictive utility of test scores. It is also important to look at the payoff function relating expected payoff

(e) to an information category or test score (y) for a particular treatment. This function provides information as to how much benefit a person with a given score is expected to contribute in that treatment. Often a number of different treatments, different training programs or modified job responsibilities, may yield the same benefits. If there are several treatments, then each treatment will have its own payoff function. These payoff functions, relating a test score to an expected outcome, may be very different from each other. For example, as illustrated by Cronbach and Gleser, students with a strong math background and technical vocabulary may advance quickly in a fast-paced technical training program, whereas, students with weak backgrounds would flounder. On the other hand, those strong students would not progress any more rapidly than the weak students in a slow-paced program. Graphs depicting different payoff functions for these different treatments are presented in Figure 9. The fast-paced program (T_f) would have a steeply sloped payoff function associated with it and the slow-paced program (T_s) would have a relatively flat payoff function associated with it.

These different payoff functions suggest that treatments should be adapted to the aptitudes of the individuals. As noted before, test information would be used both to select individuals and to select treatments for those individuals that would maximize expected payoff. These different payoff functions also indicate that there are significant interactions between aptitudes and treatments, which is just another way to point out that, in such a situation, the use of adaptive treatments may increase the payoff to the organization.



T_f - Treatment in which pace of teaching is fast

T_s - Treatment in which pace of teaching is slow

Figure 9. Payoff Functions for Different Treatments

One of the primary personnel problems in the organization which could benefit the most from adaptive treatments is selection. Cronbach and Gleser developed theory and formulae for evaluating whether adaptive treatments will increase the payoff to the organization in selection. The theory and formulae relating to adaptive treatment selection, discussed below, provide the basis for the proposed test of adaptive treatment.

Adapting Training Treatments in Selection Decisions

Fixed- and adaptive-treatment in selection decisions

As discussed above, selection decisions are those decisions in which one of the possible treatments is rejection. The selection decision process closely follows that depicted in Figure 10, modeled after the general personnel decision process shown in Figure 8. Information is gathered through the use of selection tests which will generally be considered aptitude tests. This information is used to predict the outcomes associated with the different treatments. Besides rejection, there is typically one other available treatment in selection, more specific than just acceptance, as new employees generally either start their new jobs directly or enter a training program. This paper focuses on training as the primary treatment experienced after selection. The outcomes of training programs will be measured using Kirkpatrick's (1967) criteria for evaluation of training: reaction, learning, behavior and results. The decision maker uses the expected outcomes to determine whether to make a terminal decision, rejection or assignment to the training program, or an investigatory decision in

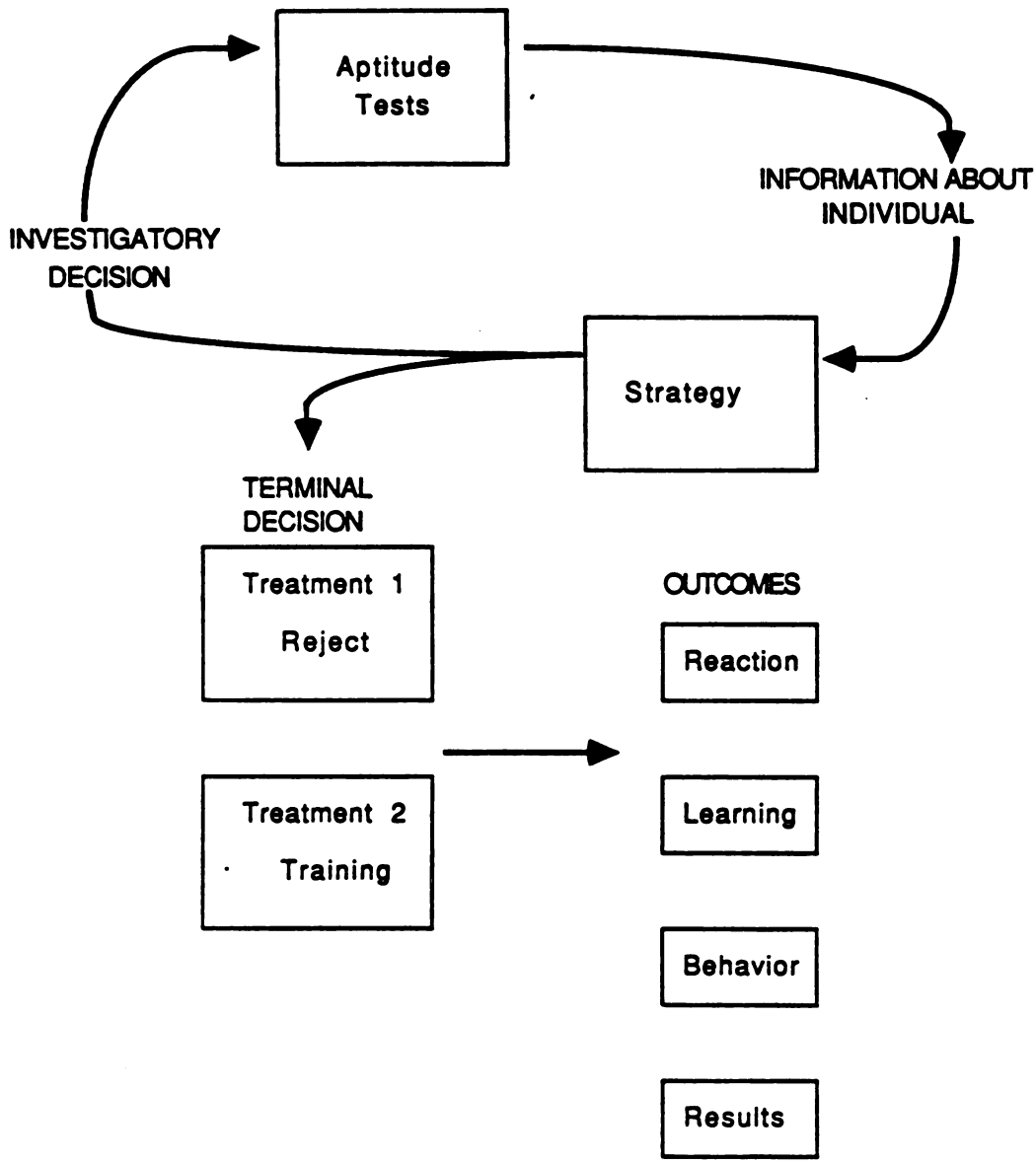


Figure 10. Selection Decision Process.

which more information is sought. The strategy for making the selection decisions, using the tests to reject and accept applicants, is eventually evaluated by looking at the payoff that accrues to the organization.

Fixed-treatment selection. This decision process is one of fixed-treatment selection if there is only one treatment, e.g., one training program, that cannot be modified, which all selected applicants enter. Since it is assumed that rejected applicants no longer have any contact with the organization, then the expected payoff or utility of fixed treatment selection is estimated by looking at the payoff function for the training treatment. This function is a more specific case of the one presented earlier to estimate the utility of a strategy.

The utility can be assessed by looking at the net gain in utility per individual tested from selection for a fixed treatment. The formula presented by Cronbach and Gleser (1965) is presented below:

$$U = \sigma_e \, r_{ye} \, \xi(y') - C_y \quad \text{where:}$$

U = net gain in utility per person

C_y = average cost of testing one person,

r_{ye} = the correlation of the test with the criterion in in the *a priori* population

σ_e = the standard deviation of the outcome

y' = the cutting score on the test

$\xi(y')$ = the ordinate of the normal curve at that point.

This equation reflects several points. First, the net gain in utility per person is linearly related to the validity of the test.

Therefore, $\sigma_e r_{ye}$ represents the slope of the payoff function, relating utility to the test score. However, more than just validity affects the slope of the payoff function. The standard deviation of the outcome which reflects the range of outcome or criterion score and reflects the value associated with one unit on the scale measuring outcome also affects the slope.

The standard deviation of the outcome gives the utility equation its potential for practical significance. "A large σ_e is an indication that individual differences on the criterion [outcome] in question have large practical importance. Tests for important decisions [with large σ_es] which fall short of the ideal predictor [which has a large validity] may be much more worth using (and improving) than tests which give excellent guidance in making small decisions" (Cronbach and Gleser, 1965, p. 39).

However, the standard deviation of the outcome has also precluded the use of the utility equation for fixed treatment selection until recently. The problem was estimating σ_e. However, lately several methods for estimating $_e$ have been devised by researchers (e.g., Cascio & Silbey, 1979; Schmidt, Hunter, McKenzie and Muldrow, 1979). These methods have been applied in the equation to estimate the utility of a given test battery/selection strategy given a fixed treatment. For example, Schmidt, Mack and Hunter (1984) used a modified version of the formula given by Cronbach and Gleser to estimate the utility of using a valid test versus a

structured interview in selecting park rangers. In a related test, Hunter and Hunter (1984) used another version of the formula with an estimate of σ_e of $13,598 to estimate that the United States government could gain $15.61 billion in productivity for one year if it hired federal workers on the basis of cognitive ability tests rather than randomly.

It is important to differentiate the utility formula for fixed-treatment selection from the utility formula for evaluating the impact of intervention programs (or training programs). The selection utility equation was modified by Schmidt, Hunter and Pearlman (1983) to evaluate the effect of training a group of workers. Where the fixed-treatment selection utility formula ignores the impact of training on performance, this modified training utility formula ignores selection. It simply assesses the difference in utility (based on performance differences) from training or not training a cohort of employees. Neither formula, therefore, can be used to assess the joint impact of selection and training on performance.

In summary, the utility formula for fixed-training treatment selection provides a way to estimate the value of a selection decision process based on getting information on individuals' aptitudes and predicting the expected outcomes associated with those aptitudes in a fixed training situation. The major limitation is that this formulation does not allow for modifications of the treatment situation based on that same information. As Cronbach and Gleser (1965) suggest, the expected payoff for a selection strategy

or set of selection tests, should be higher if the treatments are adapted to accommodate or match individuals' aptitudes.

Adaptive-treatment selection. Even in its simplest form, adaptive-treatment selection is actually a combination of fixed and adaptive-treatment placement in which there are two treatments: the first treatment consists of those individuals who are rejected on the basis of their test scores, while the second treatment is the one best suited to the average aptitude of the selected individuals. However, it is more obviously a combination of fixed-treatment and adaptive-treatment placement in more advanced forms in which there are several treatments for selected individuals. This situation is depicted in Figure 11.

There are essentially two cutoff scores in this form of adaptive treatment selection. The first cutoff score (y') indicates the minimum score required for selection. Individuals with scores falling below y' are rejected. The second cutoff score (y") separates selected individuals into treatments. Those individuals whose test score falls between y" and y' receive one treatment that has been determined to yield the best payoff for this aptitude range, whereas, those individuals with test scores above y" receive a different treatment which has been determined to yield the best payoff for that aptitude range. The goals are to understand how treatments yield different payoffs for different levels of aptitude and to identify what affects the utility of an adaptive-training treatment selection decision process.

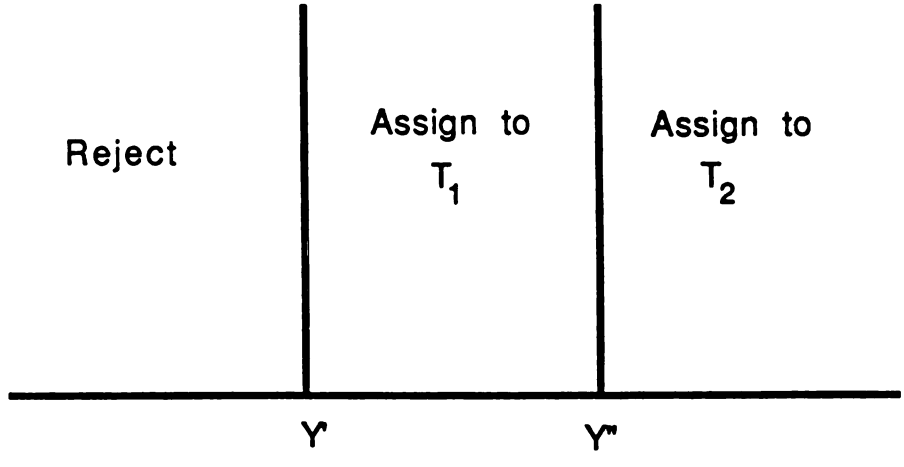


Figure 11. Adaptive Treatment Selection Decision Problem.

Expected payoff under adaptive selection is a function of treatment and aptitude. In estimating payoff for adaptive selection it is advantageous to include the concept of an aptitude (s) intervening between the test score and payoff such that $r_{ye} = r_{ys}r_{se}$. This helps separate the aspects of the decision that are associated with treatments from those associated with the test. It is also assumed that there is only one aptitude that intervenes between the test scores and the payoff, although there can be multiple measures of that aptitude (i.e., several different tests). This makes it possible to invoke the same payoff function for all tests for each treatment.

Cronbach and Gleser (1965) derived the equation relating aptitude and treatment to payoff and provided a graph of the expected payoff surface. The equation for expected payoff is:

$$e_{st} = m_{st}s + c + bm_{st} - am_{st}^2 \text{ where:}$$

e_{st} = the expected payoff for a given level of aptitude and a given treatment

m_{st} = slope of the treatment payoff function (and is equal to $\sigma_{et}r_{set}$ or product of the standard deviation of the payoff for that treatment and the validity of that test-outcome combination for that treatment

a,b,c = parameters

The graph of this function is given in Figure 12(a). The graph is in three-dimensional space with payoff (e) on the vertical axis as the dependent variable and treatment (represented by the slope of

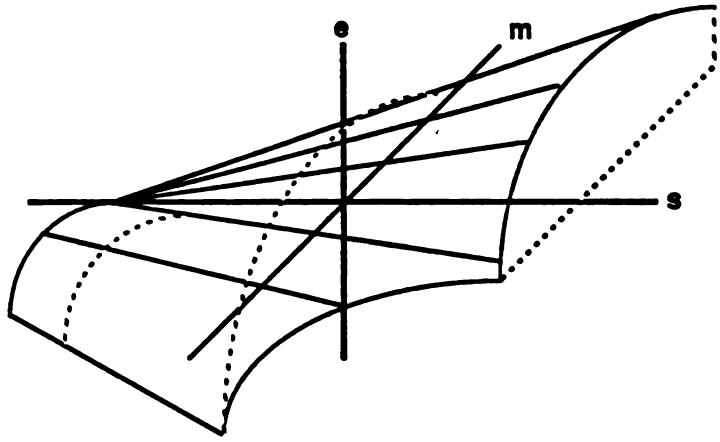


Figure 12(a). Expected Payoff as a Function of Aptitude and Treatment.

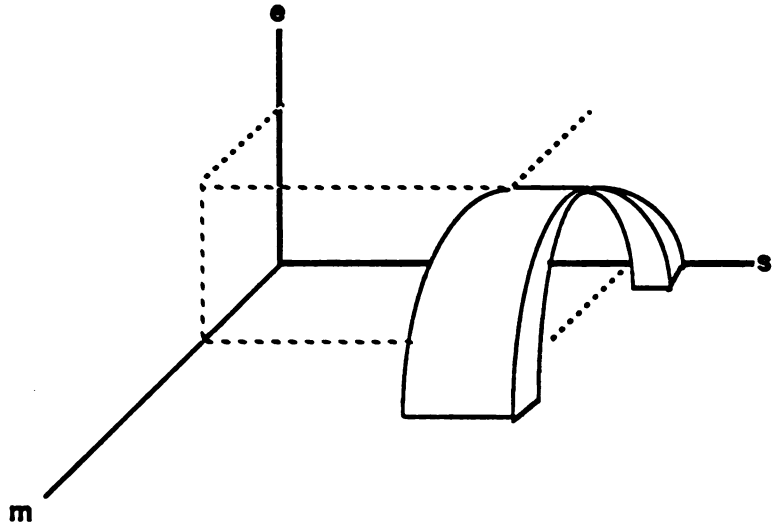


Figure 12(b). Expected Payoff and Treatment for Aptitude Level S_1

its payoff function, m) and aptitude (s) on the perpendicular horizontal axes. The equation and graph depict several points. First, all cross sections where the treatment is a constant are straight lines. This indicates that the relationship between different levels of the aptitude and payoff for a given treatment is linear. The derivations of graphs of aptitude-treatment interactions are based on these linear relationships. Second, all cross sections where s is constant (same level of aptitude) are parabolas. This implies that for each level of aptitude, payoff is maximized by assignment to a specific treatment.

Figure 12(b) shows a simplified way to see this. This figure shows a corner of a room where the vertical corner between the two walls represents payoff (e). Aptitude (s) is represented where the east wall hits the floor and treatment (t) is represented where the south wall hits the floor. A child's "London Bridge" has been pushed up flush against the east wall (s) and is sticking out into the room parallel to the south wall. The bridge represents a slice of the payoff surface where aptitude level is constant.

For each level of aptitude (s) there are an infinite number of possible treatments, each having a different payoff function. The continuous line of the side of the bridge represents these infinite number of treatments. The height of the bridge from the floor represents the level of payoff. Since the bridge has a maximum point at the top, there is a maximum payoff possible for this level of aptitude. The treatment (t_1) that provides this maximum payoff is the one which should be used for this aptitude level.

In Figure 12(c) there is an identifical drawing of the room, except that the "London Bridge" is set at a different aptitude level and is constructed for smaller children. The maximum payoff for this aptitude is smaller and provided by a different treatment (t_2).

The graph of the payoff functions for these two treatments (t_1 and t_2) in two dimensional space with aptitude on the x axis and payoff on the y axis, as shown in Figure 12(d), shows an aptitude-treatment interaction. This indicates that one of the primary ways to assess whether adaptive selection will be beneficial to the organization is to look for ATIs.

As with fixed treatment, utility formulas provide an estimate of the extent to which adaptive treatment will be beneficial to the organization. The formula for gain in utility for adapting treatments to the aptitudes of selected individuals is quite complex:

$$U = [r_{ys}^2/4a][\xi^2(y')/\phi(y')] + [br_{ys}/2a][\xi(y')] - C_y, \text{ where}$$

U = the average gain in utility

r_{ys}^2 = correlation between the test and the aptitude

y' = cutoff score

C_y = cost of testing

a, b are parameters from the function for the payoff surface

$\xi(y')$ = ordinate at the cutoff score

$\phi(y')$ = selection ratio with the cutoff score.

This formula essentially represents the difference in average utility between the best <u>a priori</u> treatment in which treatment is fixed to the level best suited for the average individual and the

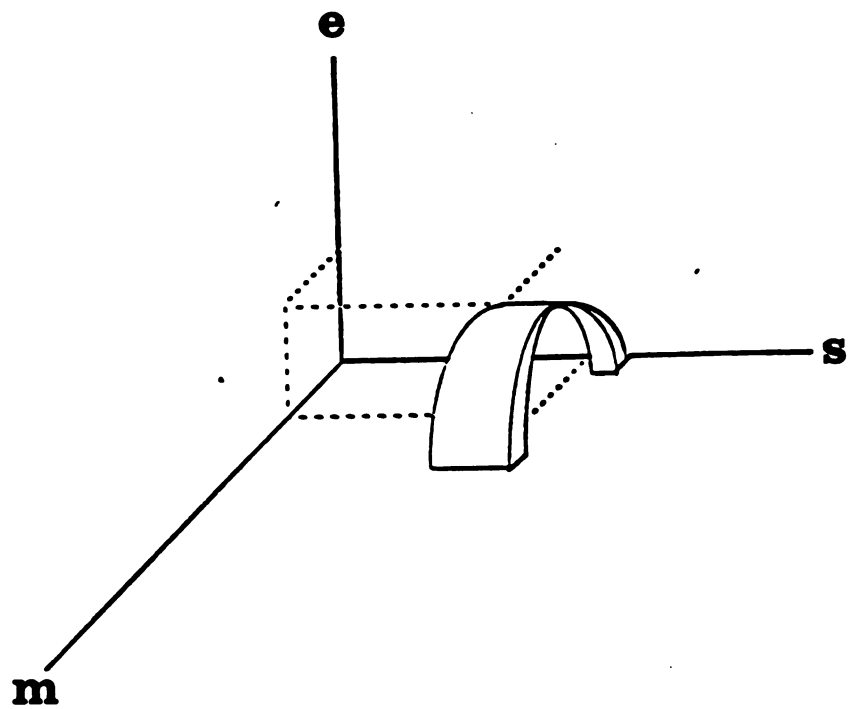


Figure 12(c). Expected Payoff and Treatment for Aptitude Level s_2

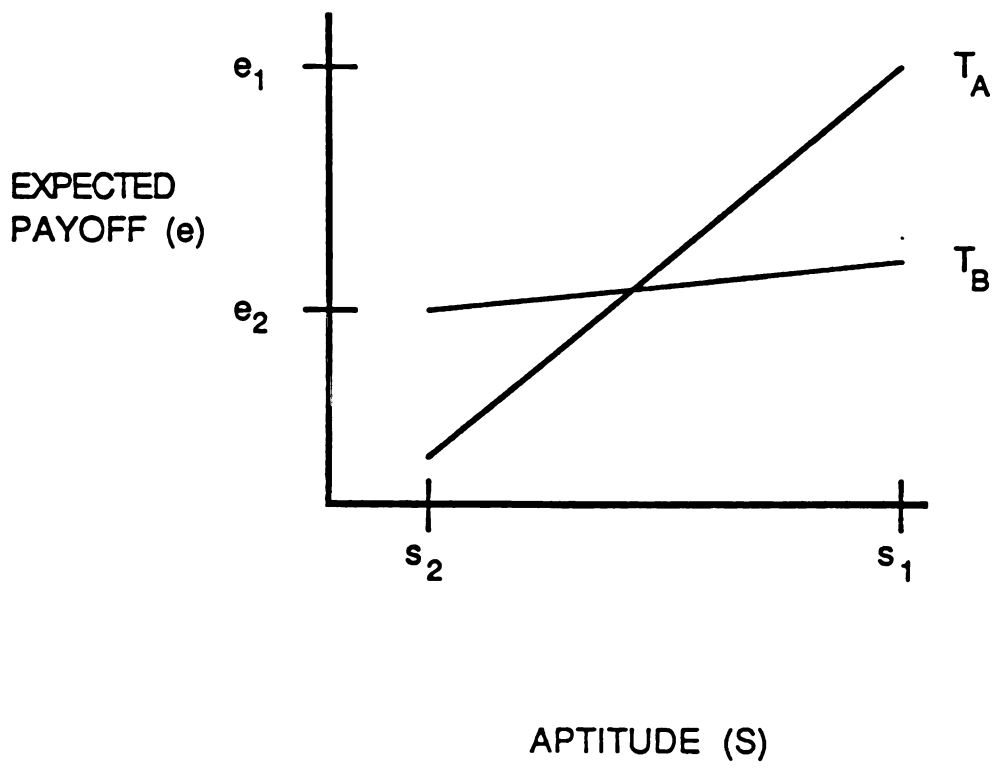


Figure 12 (d). Aptitude - Treatment Interaction for Aptitude S

best a posteriori treatment in which the selected treatment depends on the aptitude of the selected individual.

Cronbach and Gleser summarized the effects of testing when treatments are also considered in estimating payoff:

> Testing can bring about two types of change: an increase in the average quality of accepted men; and a further increase in benefits when treatment is adapted to fit this new level of quality....The relative advantages of selection without adaptation, adaptation without selection, and adaptation with selection depend on the parameters of the postulated payoff surface (and on the cost of adaptation). No unqualified generalizations as to proper practice can be made without considering particular surfaces established empirically. (1965, p. 49, italics added).

Appendix B

Formulas and Procedures Used in Meta-Analysis

1. Standardized regression weights as calculated from studies which used multiple regression: $B_i = [b_{T1}(sd_{x1}/sd_{y1}) - b_{T2}(sd_{x2}/sd_{y2})]$

Where b_{Ti} is the unstandardized regression weight for the ith treatment, sd_{xi} is the standard deviation for the aptitude for the ith treatment group, and sd_{yi} is the standard deviation for the criterion for the ith treatment group.

2. Weighted Average Standardized Regression Weight:

$$\bar{B} = [\ \Sigma(n_i)\ (B_i)] \ / \ n$$

Where n_i is the sample size for the ith study and B_i is the standardized regression weight calculated for the ith study.

3. Observed variance in the standardized regression weights:

$$\Sigma[n_i(B_i-\bar{B})^2] \ / \ n$$

4. Error variance across group of standardized regression weights:

$$([\ \Sigma\ (se_{Bi})^2] \ / \ n)$$

Where se_{Bi} is the standard error of the standardized regression weight for the ith study and n is the number of studies contributing standard errors.

5. Orwin's adaptation of Rosenthal's fail safe number:

$$[N_o\ (d_o - d_c)] \ / \ d_c$$

where d is equal to: $(2\bar{B})/(1 - \bar{B}^2)$--from Hunter et al, 1982, p. 98).

Appendix C

Measures

WONDERLIC

PERSONNEL TEST

FORM I

NAME_____ Date_____
(Please Print)

**READ THIS PAGE CAREFULLY. DO EXACTLY AS YOU ARE TOLD.
DO NOT TURN OVER THIS PAGE UNTIL YOU ARE
INSTRUCTED TO DO SO.**

**PROBLEMS MUST BE WORKED WITHOUT THE AID OF A CALCULATOR
OR OTHER PROBLEM-SOLVING DEVICE.**

This is a test of problem solving ability. It contains various types of questions. Below is a sample question correctly filled in:

REAP is the opposite of

1 obtain, 2 cheer, 3 continue, 4 exist, 5 sow _____ [5]

The correct answer is "sow." (It is helpful to underline the correct word.) The correct word is numbered 5. Then write the figure 5 in the brackets at the end of the line.

Answer the next sample question yourself.

Paper sells for 23 cents per pad. What will 4 pads cost? _____ [___]

The correct answer is 92¢. There is nothing to underline so just place "92¢" in the brackets.

Here is another example:

MINER MINOR — Do these words have

1 similar meaning, 2 contradictory, 3 mean neither same nor opposite? _____ [___]

The correct answer is "mean neither same nor opposite" which is number 3 so all you have to do is place a figure "3" in the brackets at the end of the line.

When the answer to a question is a letter or a number, put the letter or number in the brackets. All letters should be printed.

This test contains 50 questions. It is unlikely that you will finish all of them, but do your best. After the examiner tells you to begin, you will be given exactly 12 minutes to work as many as you can. Do not go so fast that you make mistakes since you must try to get as many right as possible. The questions become increasingly difficult, so do not skip about. Do not spend too much time on any one problem. The examiner will not answer any questions after the test begins.

Now, lay down your pencil and wait for the examiner to tell you to begin!

Do not turn the page until you are told to do so.

1. The Eleventh month of the year is
   1 October, 2 May, 3 November, 4 February _____ [___]
2. SEVERE is the opposite of
   1 harsh, 2 stern, 3 tender, 4 rigid, 5 unyielding _____ [___]
3. In the following set of words, which word is different from the others?
   1 certainty, 2 dubiousness, 3 assuredness, 4 confidence, 5 sureness _____ [___]
4. Answer by printing YES or NO. Does B.C. mean "before Christ"? _____ [___]
5. In the following set of words, which word is different from the others?
   1 sing, 2 call, 3 chatter, 4 hear, 5 speak _____ [___]
6. PURE is the opposite of
   1 immaculate, 2 indecent, 3 incorrupt, 4 innocent, 5 classical _____ [___]
7. Which word below is related to chew as smell is to nose?
   1 sweet, 2 stink, 3 odor, 4 teeth, 5 clean _____ [___]
8. How many of the five pairs of items listed below are exact duplicates? _____ [___]

Sharp, M. G.	Sharpe, M. G.
Fiedler, E. H.	Fiedler, E. H.
Conner, M. J.	Conner, M. J.
Woerner, O. W.	Woerner, O. W.
Soderquist, P. E.	Soderquist, P. E.

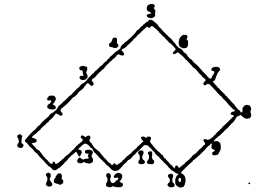
9. CLEAR is the opposite of
   1 plain, 2 obvious, 3 explicit, 4 distinct, 5 dim _____ [___]
10. A dealer bought some T.V.'s for $3500. He sold them for $5500, making $50 on each T.V. How many T.V.'s were involved? _____ [___]
11. ADOPT ADEPT — Do these words have
    1 similar meanings, 2 contradictory, 3 mean neither same nor opposite? _____ [___]
12. Lemon candies sell at 3 for 15 cents. How much will 1½ dozens cost? _____ [___]
13. How many of the six pairs of items listed below are exact duplicates? _____ [___]

5296	5296
66986	66486
834426	834426
7354256	7354256
61197172	61197172
83238234	83238234

14. FAMILIAR is the opposite of
    1 friendly, 2 old, 3 strange, 4 aloof, 5 different _____ [___]
15. Which number in the following group of numbers represents the smallest amount?
    6 .7 9 36 .31 5 _____ [___]
16. Suppose you arranged the following words so that they made a true statement. Then print the last letter of the last word as the answer to this problem.
    of salt the life Love is _____ [___]
17. One of the numbered figures in the following drawings is most different from the others. What is the number in that drawing? _____ [___]

18. Two men caught 36 fish; X caught 8 times as many as Y. How many fish did Y catch? _____ [___]
19. REFLECT REFLEX — Do these words have
    1 similar meanings, 2 contradictory, 3 mean neither same nor opposite? _____ [___]
20. Suppose you arrange the following words so that they make a complete sentence. If it is a true statement, mark (T) in the brackets, if false, put an (F) in the brackets.
    moss A stone gathers rolling _____ [___]
21. Assume the first 2 statements are true. Is the final one: (1)true, (2)false, (3)not certain?
    Most progressives are business men. Most progressives are Republicans. Some business men are Republicans. _____ [___]
22. Two of the following proverbs have similar meanings. Which ones are they? _____ [___]
    1. Straws show which way the wind blows.
    2. An empty sack can't stand straight.
    3. No doctor at all is better than three.
    4. All is not gold that glitters.
    5. Too many cooks spoil the broth.
23. Look at the row of numbers below. What number should come next?
    73 66 59 52 45 38 ? _____ [___]
24. The hours of daylight in SEPTEMBER are nearest equal to the hours of daylight in
    1 June, 2 March, 3 May, 4 November _____ [___]
25. Assume the first 2 statements are true. Is the final one: (1)true, (2)false, (3)not certain?
    Bill is the same age as Mary. Mary is younger than John. Bill is younger than John [___]
26. A train travels 75 feet in ¼ second. At this same speed, how many feet will it travel in 5 seconds? _____ [___]
27. Five pounds of feed sells for $2.00: how many pounds can you buy for 80 cents? _____ [___]
28. STRETCH SPREAD — Do these words have
    1 similar meanings, 2 contradictory, 3 mean neither same nor opposite? _____ [___]

29. This geometric figure can be divided by a straight line into two parts which will fit together in a certain way to make a perfect square. Draw such a line by joining two of the numbers. Then write the numbers as the answer. ..... [___]



30. Assume the first 2 statements are true. Is the final one: (1)true, (2)false, (3)not certain?
Fred greeted Mary. Mary greeted Ned. Fred did not greet Ned. ..... [___]

31. An automobile that costs $2490 has decreased 33⅓% in value by the end of the year.
What is its value at that time? ..... [___]

32. One of the numbered figures in the following drawings is most different from the others.
What is the number in that drawing? ..... [___]

33. A skirt requires 2⅓ yards of material. How many can be cut from 42 yards? ..... [___]

34. Are the meanings of the following sentences: 1 similar, 2 contradictory, 3 neither similar nor contradictory? No doctor at all is better than three. The more doctors, the more sickness. [___]

35. ENLARGE AGGRANDIZE — Do these words have
1 similar meanings, 2 contradictory, 3 mean neither same nor opposite? ..... [___]

36. Are the meanings of the following sentences: 1 similar, 2 contradictory, 3 neither similar nor contradictory? It is always well to moor your ship with two anchors. Don't put all of your eggs in one basket. ..... [___]

37. For $3.60 a grocer buys a case of fruit which contains 12 dozen pieces. He knows that two dozen pieces will spoil before he sells them. At what price per dozen must he sell the good ones to gain ⅓ of the whole cost? ..... [___]

38. PRETENSIONS PRETENTIOUS — Do these words have
1 similar meanings, 2 contradictory, 3 mean neither same nor opposite? ..... [___]

39. When wire is selling at $.0125 a foot, how many feet can you buy for fifty cents? ..... [___]

40. One number in the following series does not fit in with the pattern set by the others. What should that number be? ¼ ⅓ ⅛ ¼ ⅛ ⅛ ¼ ⅛ ⅛ ..... [___]

41. IMAGE IMAGINARY — Do these words have
1 similar meanings, 2 contradictory, 3 mean neither same nor opposite? ..... [___]

42. How many square yards are there in a floor which is 6 feet long by 21 feet wide? ..... [___]

43. Are the meanings of the following sentences: 1 similar, 2 contradictory, 3 neither similar nor contradictory? All good things are cheap, all bad things very dear. Goodness is simple: badness is manifold. ..... [___]

44. A soldier shooting at a target hits it 12½% of the time. How many times must he shoot to be certain he will register 100 hits? ..... [___]

45. One number in the following series does not fit in with the pattern set by the others. What should that number be? ¼ ⅛ ⅛ ⅑ 1/12 1/14 ..... [___]

46. Three men form a partnership and agree to divide the profits equally. X invests $4500, Y invests $3500, Z invests $2000. If the profits are $2400, how much less does X receive than if the profits were divided in proportion to the amount invested? ..... [___]

47. Two of the following proverbs have similar meanings. Which ones are they? ..... [___]

1. Perfect valor is to do without witnesses what one would do before all the world.
2. Valor and boastfulness never buckle on the same sword.
3. The better part of valor is discretion.
4. True valor lies in the middle between cowardice and rashness.
5. There is a time to wink as well as to see.

48. Are the meanings of the following sentences: 1 similar, 2 contradictory, 3 mean neither similar nor contradictory? After the event even a fool is wise. No man ever became wise by chance. ..... [___]

49. Three of the following 5 parts can be fitted together in such a way to make a triangle. Which 3 are they? ..... [___]

50. In printing an article of 24,000 words, a printer decides to use two sizes of type. Using the larger type, a printed page contains 900 words. Using the smaller type, a page contains 1200 words. The article is allotted 21 full pages in a magazine. How many pages must be in the smaller type? ..... [___]

JOB TITLE _____

APPLICANT# _____ EMPLOYEE# _____ AGE _____ SEX _____

EDUCATION 6 7 8 9 10 11 12 13 14 15 16 16+
ENCIRCLE NUMBER OF LAST GRADE COMPLETED

SCORES
RAW %-ILE
R W

DATE _____
MONTH DAY YEAR

NAME _____

COMPANY _____

PLANT _____ DEPT. _____

EMPLOYEE APTITUDE SURVEY

TEST 1 – Verbal Comprehension

FORM A, REVISED

Developed by

G. Grimsley, F. L. Ruch, N. D. Warren & J. S. Ford

Look at the sample problems below. Each of the words at the left is followed by four other words. One of these four means the same or about the same as the word at the left. You are to select this word and fill in the answer space in front of it. In example 1, "big" means about the same as "LARGE." Therefore, the answer space in front of "big" has been filled in. Now do the other examples yourself.

1	LARGE	bright	small	toy	▌ big
2	SPEAK	talk	run	stand	sleep
3	STORY	hat	ball	day	tale
4	ILL	well	cold	sick	strong

"Talk" means the same as "SPEAK" so you should have blacked in the answer space in front of "talk" in item 2. "Tale" means about the same as "STORY" so the answer space in front of "tale" should have been marked in item 3. The correct answer to item 4 is "sick."

Have you any questions?

When the signal is given, turn this sheet over and mark as many items as you can in five minutes. Put a heavy black mark between the little dotted lines in front of the word which means the same or about the same as the word at the left. If you want to change an answer, be sure to erase completely. Do not waste time making pretty marks. A heavy black line is all that is needed. Work as fast and accurately as possible.

PSYCHOLOGICAL SERVICES, INC.

LOS ANGELES, CALIFORNIA

PRINTED IN U.S.A. IBM FORM I.T.S. 1160 A 4071

1	KEEN	low	sharp	thick	blunt
2	ASSIGN	design	appoint	sell	address
3	AGED	old	useless	youthful	heavy
4	DONOR	benevolent	gift	forage	giver
5	AMPLE	plentiful	useful	doubtful	hopeful
6	TOLERATE	like	endure	dislike	fear
7	UNCOUTH	charming	spoiled	crude	proud
8	SHANTY	hut	cap	beggar	elf
9	NEEDY	cheap	weak	poor	dirty
10	ENCOMPASSED	hidden	surrounded	divided	clever
11	SCOFF	flatter	injure	fight	mock
12	MEANDER	measure	persist	wander	runner
13	FATEFUL	inevitable	fatal	zealous	uncertain
14	AQUEOUS	strong	abundant	watery	acetic
15	ORTHODOX	rectangular	conventional	grammatical	preventive
16	AVID	discolored	agitated	eager	fearless
17	EFFICIENT	industrious	good	valuable	effective
18	SAVOR	taste	feed	strength	blandness
19	HUMBUG	dwarf	termite	miser	impostor
20	RUSTIC	fend	rural	enigmatic	suave
21	SPASMODIC	continuous	epileptic	enthusiastic	convulsive
22	HAPHAZARD	random	dangerous	happily	fleeting
23	TAINT	infect	hue	vanish	paint
24	TRACTABLE	disagreeable	silent	docile	noisy
25	NEFARIOUS	infamous	greedy	friendly	tedious
26	ABEYANCE	hatred	vehicle	suspension	control
27	ABASH	crush	forsake	smite	embarrass
28	MINUEND	dance	magnitude	instant	innuendo
29	IMPOST	fraud	demand	jail	tax
30	PROLIX	fruitful	verbose	symptomatic	foreboding

EMPLOYEE APTITUDE SURVEY

TEST 2 – Numerical Ability
FORM A

Developed by
G. Grimsley, F. L. Ruch & N. D. Warren

JOB TITLE _____
APPLICANT _____ EMPLOYEE _____ AGE _____ SEX _____ M or F
EDUCATION 6 7 8 9 10 11 12 13 14 15 16 16+
ENCIRCLE NUMBER OF LAST GRADE COMPLETED

SCORES
RAW | %ILE
I
II
III
T

Look at the sample problems below. Each problem is followed by four possible answers and an "X." You are to work each problem and put a heavy black mark between the little dotted lines below the correct answer. If the correct answer is not given, make a heavy black mark between the dotted lines below the "X."

Now work the sample problems below. The first one has been answered correctly.

1. 5 + 5 =	8	11	9	12	X
2. 12 − 6 =	5	7	6	8	X
3. 5 × 5 =	23	24	25	26	X
4. 30 ÷ 3 =	6	7	9	8	X
5. 6 ÷ 7 =	15	13	14	12	X

Beginning with sample problem number 2, you should have marked 6, 25, X, and 13. Are there any questions?

On the back of this sheet are 75 problems. They are divided into 3 parts. When the signal is given you are to turn this sheet and work as many of these problems as you can beginning with Part I. At the end of 2 minutes the examiner will say, "Stop on Part I, go to Part II." After 4 minutes more the examiner will say, "Stop on Part II, go to Part III." You will then have 4 minutes for Part III. If you finish a part early, check your work while waiting for the signal to go on to the next part. Are there any questions?

Work as fast and accurately as possible. Remember that the correct answer is not always given. When the correct answer is not given, mark the space below "X." Make no marks except your answers on the reverse side of this sheet. If you want to change an answer, erase completely.

DATE
DIFF
NAME
COMPANY
PLACE

COPYRIGHT 1948 1952
PSYCHOLOGICAL SERVICES, INC.
LOS ANGELES, CALIFORNIA
PRINTED IN U.S.A. IBM FORM I.T.S. 1100 A 3366

EMPLOYEE APTITUDE SURVEY

TEST 2 – Numerical Ability

Form A

PART I

1. 7 + 8 =	14	13	15	16	X
2. 13 − 9 =	4	2	3	5	X
3. 8 × 3 =	21	25	23	26	X
4. 36 ÷ 4 =	8	9	7	8½	X
5. 17 + 15 =	31	33	32	31	X
6. 21 − 9 =	16	11	17	15	X
7. 7 × 8 =	53	59	55	58	X
8. 42 ÷ 7 =	5	6	5½	7	X
9. 27 + 29 =	56	46	55	57	X
10. 33 − 19 =	13	16	15	14	X
11. 8 × 9 =	71	71	72	62	X
12. 61 + 4 =	21	14	17	16	X
13. 28 + 46 =	74	64	76	75	X
14. 73 − 38 =	45	35	31	36	X
15. 13 × 13 =	149	179	169	159	X
16. 81 ÷ 6 =	12	15	16	13	X
17. 47 + 38 =	75	81	86	95	X
18. 81 − 27 =	56	66	64	54	X
19. 14 × 15 =	200	220	230	190	X
20. 128 ÷ 8 =	15	16	14	21	X
21. 58 + 25 =	83	73	74	81	X
22. 133 − 57 =	66	75	86	74	X
23. 16 × 17 =	292	272	282	262	X
24. 324 ÷ 18 =	16	19	26	18	X
25. 117 + 276 =	423	433	323	413	X

PART II

1. .2 + .6 =	.8	.08	.28	.82	X
2. .9 − .3 =	.39	.03	.06	6.0	X
3. .4 × .7 =	2.8	28	.28	.028	X
4. .6 ÷ 2 =	3.0	.3	.03	3.3	X
5. 6% of 10 =	6	.06	.6	6.0	X
6. .07 + .6 =	.67	6.7	.13	1.3	X
7. .5 − .03 =	.2	.47	.497	.047	X
8. .06 × .7 =	.42	4.2	0.42	.042	X
9. .26 ÷ .2 =	.13	13	130	1.3	X
10. 4% of 25 =	10	1.10	11	.10	X
11. 1.9 + .06 =	1.96	2.5	.25	.196	X
12. 1.78 − .7 =	.171	1.71	.108	1.08	X
13. 1.5 × 0.5 =	7.5	.65	.075	.85	X
14. .09 ÷ .03 =	.3	3	.03	.003	X
15. 5% of 10.6 =	5.3	.053	.53	.503	X
16. 2.08 + 1.52 =	3.6	.36	4.32	3.32	X
17. 1.07 − .47 =	.50	5.0	6.0	.60	X
18. 1.2 × .12 =	.154	14.4	1.44	1.54	X
19. 7.5 ÷ .05 =	.15	150	1.50	15	X
20. 3.5% of 80 =	2.8	.028	.28	28.0	X
21. .070 + .171 =	1.87	.241	.871	.178	X
22. 4.01 − 2.9 =	3.72	.111	1.11	2.11	X
23. .80 × .12 =	.960	9.6	.96	.096	X
24. 12.1 ÷ .11 =	.110	1.01	1.10	11.0	X
25. .06% of 150 =	9.0	0.9	.09	.90	X

PART III

1. ½ + ½ =	¼	½	1	2	X
2. 1 − ½ =	½	¾	0	½	X
3. 1 × ½ =	½	¼	½	1	X
4. 1 + ½ =	½	2	¾	1	X
5. ½ + 1 =	½	¾	1½	½	X
6. ½ − ¼ =	¼	½	¾	⅛	X
7. ½ × ½ =	¼	⅛	½	¾	X
8. ½ + ½ =	½	¼	⅛	1	X
9. ½ + ¼ =	¾	⅔	½	⅓	X
10. ¼ − ⅛ =	⅛	¼	⅜	⅝	X
11. ¼ × ½ =	⅛	⅜	⅝	⅞	X
12. ⅝ + ½ =	¼	3	1	⅛	X
13. ¼ + ⅓ =	⅓	⅔	11	11	X
14. ½ − ⅛ =	⅜	⅝	88	88	X
15. ⅔ × ¼ =	11	11	11	11	X
16. ⅛ + ½ =	11	⅔	3	11	X
17. ⅜ + ½ =	88	2⅔	11	3⅛	X
18. ½ − ⅛ =	¼	⅜	1	⅛	X
19. ½ × ½ =	½	11	⅓	⅛	X
20. 1¼ ÷ ⅛ =	⅛	⅛	5	6	X
21. 1⅓ + ⅓ =	1⅛	1⅓	1⅛	2	X
22. 4¼ − ⅔ =	1⅛	1⅓	1⅓	1⅓	X
23. 3½ × ⅓ =	1⅛	1⅓	2⅓	1⅓	X
24. 6½ + 1⅓ =	4⅜	3⅓	4⅛	4⅓	X
25. 5½ + 2⅓ =	7⅜	7⅜	6⅜	9	X

TAX COURSE STUDY DIARY

Use the grid below to help you keep track of how many hours you spend working on the Tax Course during the week. Please include time spent on all aspects of the course, for example, time in class, reading, working on homework or studying. In determining the amount of time subtract any breaks, interruptions, etc. That is, record only the time you actually spend on the Tax Course, to the nearest 1/2 hour.

Enter the number of Tax Course related hours you spent (even if it was 0) in each part of each day. At the end of the week (Saturday) please total the number of hours you spent on the Tax Course.

You do not need to carry this form around with you, making entries three times a day, but please do not wait more than two days before recording your activities, as it essential that this information be accurate. As with all other information you provide for this study, your entries will only be seen by the investigator.

Week beginning Sunday _____ and ending Saturday _____

DAY	MORNING	AFTERNOON	EVENING	TOTAL HOURS/DAY
SUNDAY				
MONDAY				
TUESDAY				
WEDNESDAY				
THURSDAY				
FRIDAY				
SATURDAY				
	TOTAL HOURS SPENT ON TAX COURSE THE WEEK OF X/XX TO X/XX			

Mid-Term Exam

INSTRUCTIONS: You will have 90 minutes to complete this two-part exam. Your answers are to be recorded on the separate answer sheets that will be found on pages R11.1 and R11.3 in the back of Volume 1 of the Student Workbook. Part 1 will be closed-book. As soon as you complete this part, turn answer sheet R11.1 in to your instructor and continue with Part 2, the open-book portion of the exam. All of the information necessary to complete this exam has been discussed in Chapters 1-10 of your Basic Tax Course text. Adding machines or calculators may be used to help you complete Part 2. When you complete the exam, hand it in and you may leave.

PART 1. The following statements are either true or false. Indicate your answer on the separate answer sheet by circling T for true or F for false.

1. A joint return may be filed even though one spouse has no income.

2. Mr. and Mrs. Fix are married and have one child. During 1986, they lived apart the entire year. They do not wish to file a joint return. Mr. Fix provided 100% of his child's support, and the child lived with him for the entire year. Mr. Fix paid all the costs of keeping up the home. Mr. Fix may use the head of household filing status for 1986.

3. Mr. Carter, a single taxpayer, age 43, provides all of the support for his parents. His father is 75. His mother is 73 and is blind. Mr. Carter may claim 6 exemptions on his return.

4. A couple is divorced. The divorce decree does not state which parent may take the dependency exemption for a child of the marriage. The child lived with the custodial parent all year and received all his support from his parents. The custodial parent may claim the exemption for the child unless he or she signs a statement allowing the noncustodial parent to claim the exemption.

5. If a taxpayer is unable to obtain Form W-2 from his employer, he cannot file his return until the form has been obtained from that employer.

6. Excess FICA withheld due to the employer's error can be used on the tax return as a payment of tax.

7. All interest income is taxable on the federal return, regardless of its source.

8. Alimony payments of $600 a month payable until death or remarriage are considered periodic.

9. A taxpayer must always be age 65 or older to qualify for the Credit for the Elderly or the Permanently and Totally Disabled.

10. Filing a joint return, if married, and having AGI under $11,000 are the only requirements for claiming the earned income credit.

11. For 1986, certain taxpayers were required to itemize deductions.

12. The cost of restaurant meals while away from home to obtain medical treatment is an allowable itemized deduction.

13. The cost of property donated to charity always is deductible as a charitable contribution.

14. All taxpayers who receive Form 1099-G showing the amount of a state tax refund must include the refund in taxable income.

15. A taxpayer who will recover his contribution in his pension in 36 months or less is required to exclude from taxable income the total amount received until his personal contribution is recovered if the pension starting date is prior to July 2, 1986.

16. The taxable portion of social security benefits varies from none to 50%, depending on the recipient's circumstances.

17. If a couple is married and living together, both must hold jobs in order to claim the child care credit.

18. A taxpayer who makes a rollover from one IRA to another has 60 days to complete the transaction.

19. A capital loss on personal-use property, like a taxpayer's car used only for personal purposes, is deductible.

20. Taxable alimony received is considered earned income for purposes of claiming an IRA deduction.

Give your instructor your answer sheet for Part 1 before beginning Part 2.

PART 2. The following questions are multiple choice. Indicate the correct answer on the separate answer sheet by putting a circle around the letter corresponding to the correct statement. There is only one correct answer for each question.

1. For 1986, what is the gross income filing requirement for a married couple, both age 43, using the married filing jointly status?
   A. $1,080
   B. $2,160
   C. $3,830
   D. $3,670

2. Which of the following items is not considered when determining the cost of maintenance of the home?
   A. Insurance on the home
   B. Utilities
   C. Food consumed in the home
   D. Mortgage principal payment

3. Which of the following dependent relatives does not have to live in the same household with the taxpayer to qualify a taxpayer for head of household status?
   A. Brother
   B. Aunt
   C. Mother
   D. Married child

4. Which of the following is not one of the five tests for dependency?
   A. Age
   B. Citizenship
   C. Gross income
   D. Support

5. Which of the following items is not included when determining the total support of a dependent?
   A. Education
   B. Child care
   C. Life insurance
   D. Recreation

6. Mr. and Mrs. House were both 61 years old in 1986. Mr. House died on January 1, 1986. How many exemptions may Mrs. House claim when filing her joint 1986 return?
   A. One
   B. Two
   C. Three
   D. Four

7. Mr. and Mrs. Burk, both under age 65, file a joint return. They are entitled to claim their two children as dependents. Their adjusted gross income is $45,000 and their excess itemized deductions total $4,300. Determine their tax.
   A. $6,761
   B. $6,319
   C. $7,957
   D. $9,865

8. A single taxpayer received the following income during 1986:

Wages	$35,000
Taxable interest	525
Taxable alimony	3,850
Nonqualifying dividends	100
Gain from sale of IBM stock	6,000

   Aside from the Income Information Worksheet, which forms is this taxpayer sure to need to complete her return?
   A. Form 1040A (both pages) and Schedule 1.
   B. Form 1040 (both pages), Schedule A, and Schedule B.
   C. Form 1040 (both pages), Schedule B, and Form 2441.
   D. Form 1040 (both pages), Schedule B, and Schedule D.

9. A taxpayer received the following income:

Interest credited to passbook savings account	$620
Interest on certificate of deposit	125
Dividends on share account in credit union	130
Interest on state municipal bond	340

   What is the final amount of interest income to be reported on line 3, Schedule B, Form 1040?
   A. $ 255
   B. $ 745
   C. $ 875
   D. $1,215

10. Which of the following income items is not taxable on the federal return?
    A. Income from an illegal activity
    B. Tips under $20 per month
    C. Fees received for jury duty
    D. Inheritance

11. A taxpayer's divorce decree requires him to pay $200 a month alimony and $100 per month child support. In 1986, he made total payments of $3,000. How much may he deduct as alimony on his return?
    A. $1,200
    B. $1,800
    C. $2,400
    D. $3,000

12. Which of the following has not been repealed by tax reform?
    A. Two-earner married couple deduction
    B. Credit for contributions to candidates for public office

C. Regular income averaging
D. Child and dependent care credit.

13. Which of the following expenses does *not* qualify as a deductible medical expense?
A. Fee paid to Christian Science practitioner
B. Premiums on a policy to cover loss of earnings if hospitalized
C. Maintenance costs of a guide dog for the blind
D. Cost of hearing aid batteries

14. A taxpayer paid the following interest in 1986:

Mortgage prepayment penalty	$500
Interest on brother's mortgage	400
MasterCard	100
Interest on money used to buy municipal bond	700

What is his interest deduction on Schedule A?
A. $ 500
B. $ 600
C. $1,000
D. $1,700

15. Which of the following items *cannot* be deducted as a charitable contribution?
A. Cash donation to church
B. Cost of volunteer scout master uniform
C. Cost of raffle tickets sold by church
D. FMV of property donated to Red Cross

16. Which of the following is *not* an allowable casualty loss?
A. Theft of registered dog
B. Accidental loss of diamond ring
C. Vandalism damage to personal residence
D. Roof damage from wind storm

17. Which of the following *cannot* be claimed as a miscellaneous deduction on Schedule A?
A. Tax preparation fee
B. Preparation of a will
C. Nurse's uniforms and shoes
D. $300 for lottery tickets (won $650)

18. Sam and Mary Fern are filing a joint return and have two dependent children, ages 3 and 4. Sam's wages are $15,800; Mary earned $8,400. AGI is $23,360. They paid total child care expenses of $2,500 to Childlove Nursery for both children. What is the amount of their child care credit?
A. $300
B. $575
C. $625
D. $750

19. Valerie Waters is an unmarried head of household with the following income for the year:

Wages	$13,000
Bank interest	300
Municipal bond interest	85
Lottery prize	200
Gift from her father	3,000

Valerie contributed $250 to her IRA. Compute her AGI.
A. $13,250
B. $16,335
C. $16,585
D. $13,000

20. A taxpayer contributed $10,800 to the cost of his pension. In June, 1986, he began receiving payments of $400 a month from the plan. How much of his pension income is taxable for 1986?
A. $ 0
B. $2,400
C. $2,800
D. $4,800

21. Which of the following payments qualifies as a lump-sum distribution for possible purposes of 10-year averaging?
A. Receipt of the entire balance from a qualified plan within one tax year because of an employee's retirement.
B. Total distribution of all funds from an IRA.
C. A loan from a qualified retirement plan.
D. Total distribution of all funds upon termination of a qualified plan; the employee continues to work for the same employer, and is not yet 59½.

22. Jane Smalley, a widow, received a lump-sum distribution of $15,000 from her husband's employer. Her husband had been employed at the time of his death. She received a 1099-R showing $15,000 as ordinary income. Mrs. Smalley plans to use special 10-year averaging to compute her tax on the distribution. What amount will be taxable?
A. $ 0
B. $10,000
C. $12,000
D. $14,000

23. Which of the following does *not* constitute earned income for purposes of making an IRA contribution?
A. Tips
B. Commissions
C. Bonuses
D. Unemployment compensation

24. A single taxpayer would like to make the maximum contribution to his IRA. His return shows the following:

Wages	$15.000
Loss from self-employment	(13.500)
Rental income	1.200
Employee business expenses	600

What amount may be contributed to his IRA?
A. $ 900
B. $1.500
C. $1.620
D. $2.000

25. Which of the following is *not* a prohibited transaction for an IRA?
A. Borrowing money from the IRA
B. Rolling over funds from one IRA to another within 60 days
C. Using IRA funds as loan collateral
D. Selling property to an IRA

26. Which of the following correctly describes the overhead projector in your classroom?
A. Tangible personal. investment-use property
B. Real. business-use property
C. Intangible personal. business-use property
D. Tangible personal. business-use property

27. Which of the following is *not* a capital asset?
A. Personal residence
B. Stock
C. Depreciable property used in a business
D. Municipal bonds

28. A single taxpayer with a 1986 AGI of $25.000 (without regard to capital losses) had a $4.000 short-term capital loss. How much of the loss may be deducted on his 1986 return?
A. $2.000
B. $2.400
C. $3.000
D. $4.000

29. A single taxpayer had 1986 taxable income on line 37. Form 1040 of $195.362. What is his tax on line 38. Form 1040?
A. $84.662
B. $87.643
C. $82.017
D. $77.611

30. A single taxpayer received the following income in 1986:

Wages	$4.000
Tax exempt interest	20.000
Social Security benefits	6.000

What portion of his Social Security benefits should be included on line 21b. Form 1040? (A copy of the Taxable SS/tier 1 RR Benefits Worksheet found in the exercises for Chapter 7 of your workbook may be helpful in computing your answer.)
A. $ 0
B. $1.000
C. $3.000
D. $6.000

TAX COURSE EXAM

NAME _____ TA (if applicable) _____

Please mark the answer sheet with the appropriate letter indicating the best answer to each question. Read each question carefully. Watch for "NOTs."

Note: Assume the question refers to the old tax law unless specified in the question by a reference to the Tax Reform Act.

1. When does a calendar year taxpayer's tax year end?

   a. June 30
   b. April 14
   c. December 31
   d. October 31

2. Bill Smith's birthday is October 6, 1945. For tax purposes his birthday is considered to be:

   a. January 1, 1945
   b. October 5, 1945
   c. October 6, 1945
   d. December 31, 1945

3. Which best describes who may file for their income taxes using Form 1040EZ?

   a. taxpayers with no itemized deductions
   b. taxpayers who have dividend income less than $400
   c. single taxpayers
   d. single taxpayers with 1 exemption and earned income less than $50,000

4. For 1986, what is the gross income filing requirement for a married couple, both age 43, using the married filing jointly status?

   a. $1,080
   b. $2,160
   c. $5,830
   d. $3,670

5. Which of the following dependent relatives does NOT have to live in the same household with the taxpayer to qualify a taxpayer for head of household status?

   a. brother
   b. aunt
   c. mother
   d. married child

6. Which of the following is NOT one of the five tests for dependency?

a. age
b. citizenship
c. gross income
d. support

7. Which of the following items is NOT included when determining the total support of a dependent?

a. education
b. child care
c. life insurance
d. recreation

8. Mr. and Mrs. House were both 61 years old in 1986. Mr. House died on January 1, 1986. How many exemptions may Mrs. House claim when filing her 1986 return?

a. zero
b. one
c. two
d. three

9. Roy Carpenter, a single taxpayer, supports and maintains a home for his mother who does not live with him. Which filing status should he use?

a. single
b. married filing separately
c. married filing jointly
d. head of household

10. Rhonda Martin (47) supports herself and her daughter, Michele. Rhonda provided the home where they lived since February when her husband, Alan (53), died. She has not remarried. What filing status should she use?

a. single
b. married filing jointly
c. qualifying widow(er)
d. head of household

11. Refer to question 10. How many dependency exemptions may Rhonda Martin claim?

a. zero
b. one
c. two
d. three

12. Michael Harris (32) is unmarried and lives with his unmarried sister who earned $2,000. Michael paid all of the household expenses and some of his sister's medical bills because his sister was injured and unable to work part of the year. What filing status should Michael use?

a. single
b. head of household
c. married filing jointly
d. qualifying widow(er)

13. Refer to question 12. How many dependency exemptions may Michael claim?

a. zero
b. one
c. two
d. three

14. Kenneth (64) and Robin (60) Reed, husband and wife, have two children, Cindy (20), a student and Ronald (22). Ronald was a full-time student from January through August; he completed requirements for a B.S. in business in August and spent the next four months touring the Orient (the trip was a gift from his father). Mr. and Mrs. Reed provided over one-half of the total support for both Cindy and Ronald. Mr. Reed also provided total support for his father, Robert (90), who is confined to a nursing home and had no income. What is the most favorable filing status for the Kenneth Reed?

a. single
b. married filing separately
c. married filing jointly
d. head of household

15. John Ford (30) is unmarried and maintains his home for himself and his son, Chris (12). Mr. Ford provides all of Chris's support. In 1984, Mr. Ford filed a joint return. His wife, Margaret, died November 5, 1984. What is the most favorable filing status for John?

a. single
b. married filing jointly
c. qualified widow(er)
d. head of household

16. Bernice Donlon (38) is unmarried and provided $6,400 toward the support of her widowed mother, Clara (68), during the year. This amount includes the cost of the nursing home in which her mother lives. Her mother paid the remaining $3,800 of her own support from her Social Security benefits, her only source of income. How many dependency exemptions may Bernice claim?

a. zero
b. one
c. two
d. three

17. Frank McDonald (45) is unmarried. His wife, Sara, died six years ago. He maintains a home for himself and his two children; Maureen (25) worked all year and earned $7,500 as a dental assistant and Jill (14) attended high school all year. Mr. McDonald provided the total support for the family. He deposited Jill's Social Security benefits in her savings account. Maureen spent $1,500 of her wages on clothing and recreation and saved the rest to purchase a car. What is the most favorable filing status for Mr. McDonald?

a. single
b. qualified widow(er)
c. head of household
d. married filing separately

18. Martin Black (43) has not heard from his wife since she left him a year and a half ago and he does not know where she is located. Martin provides the entire support for his five minor children. The children lived with Martin the entire year. He has not obtained a divorce or decree of separate maintenance. What is the most favorable filing status for Mr. Black?

a. single
b. married filing separately
c. married filing jointly
d. head of household

19. The W-2 does NOT provide information about

a. allocated tips
b. withheld social security tax
c. wages, tips and other income
d. unemployment compensation

20. If an employee is unable to obtain a W-2 from an employer,

a. the employee cannot file taxes
b. the employee reports wages and taxes paid on Schedule A
c. the employee reports wages and taxes paid on an attached statement after calling the IRS
d. the employee completes the 1040 without a W-2

21. Which taxpayers must file Form 1040 and use the tax rate schedules to compute their income tax?

a. taxpayers using married filing jointly status
b. taxpayers who receive unemployement compensation
c. taxpayers whose taxable income exceeds $49,999
d. taxpayers who have dividends under $400

22. The Zero Bracket Amount

a. is built into the tax table
b. must be deducted after finding the appropriate tax rate
c. is not built into the tax rate schedules
d. is based on the number of exemptions a taxpayer has

23. If an employee has tip income of less than $20 per month

a. he/she has to report the tip income to the IRS by the 10th day of the following month
b. he/she does not have to pay social security tax on the tips
c. Form 4137 must be used to report all tip income
d. he/she reports the income on Schedule A

24. Taxpayers who have Adjusted Gross Income of $50,000 or more

a. must file using the 1040A
b. must use the Tax Rate Schedule
c. must itemize their deductions
d. must return the zero bracket amount

25. A taxpayer received the following income in 1986:

Interest credited to passbook savings account	$620
Interest on certificate of deposit	$125
Dividends on share account in credit union	$130
Interest on state municipal bond	$340

What is the final amount of interest income to be reported for 1986 on Form 1040?

a. $ 255
b. $ 745
c. $ 875
d. $1,215

26. Which of the following income items is NOT taxable on the federal return?

a. Income from an illegal activity
b. Tips under $20 per month
c. Fees received for jury duty
d. Inheritance

27. A taxpayer's divorce decree requires him to pay $200 a month alimony and $100 per month child support. In 1986, he made total payments of $3,000. How much may he deduct as alimony on his return?

a. $1,200
b. $1,800
c. $2,400
d. $3,000

28. A taxpayer paid the following interest in 1986:

Mortgage prepayment penalty	$500
Interest on brother's mortgage	$400
MasterCard	$100
Interest on money used to buy municipal bond	$700

What is his interest deduction for 1986?

a. $ 500
b. $ 600
c. $1,000
d. $1,700

29. Albert Cole has a savings account at National Bank. The bank credited Albert with $53 interest. Albert had no other interest income. Where does Albert report this interest?

a. in Schedule A
b. in Schedule B
c. on the 1040
d. on Form 1099

30. Which of the following is NOT a periodic payment for alimony under divorce agreements made before 1985?

a. Payment of $175 per month for 5 years
b. Payment of 10% of the ex-spouse's monthly salary
c. Payment of $250 per month for 11 years
d. Payment of $325 per month until the death or remarriage of the recipient

31. Alimony is taxable income to the person receiving it when

a. the person paying the alimony may deduct it
b. the person paying the alimony may not deduct it
c. child support payments are in arrears
d. alimony is not periodic

32. If dividends or interest received during the year do NOT exceed a certain dollar amount, then they may be reported directly on the 1040 or 1040A. What is that dollar amount?

a. $1,080
b. $2,160
c. $ 600
d. $ 400

33. Which taxpayers must itemize deductions?

a. all married individuals who file separately
b. unmarried dependents with unearned income of $1080 or more and earned income less than $2480
c. married dependents with no unearned income
d. individuals filing under head of household status

34. Taxpayers who want to itemize and may itemize if

a. they have to use the tax rate schedules to calculate their taxes
b. the zero bracket amount exceeds the total of their itemized deductions
c. they are due to receive a tax refund
d. the total of their itemized deductions exceeds the zero bracket amount

35. Which of the following CANNOT be claimed as a miscellaneous deduction for 1986?

a. Tax preparation fee
b. Preparation of a will
c. Nurse's uniforms and shoes
d. $500 for lottery tickets (won $650)

36. Which of the following items is NOT considered when determining the cost of maintenance of the home?

a. Insurance on the home
b. Utilities
c. Food consumed in the home
d. Mortgage principal payment

37. Which of the following expenses does NOT qualify as a deductible medical expense?

a. fee paid to Christian Science practitioner
b. premiums on a policy to cover loss of earnings if hospitalized
c. maintenance costs of a guide dog for the blind
d. cost of hearing aid batteries

38. Which of the following items CANNOT be deducted as a charitable contribution?

a. cash donation to church
b. cost of volunteer scout master uniform
c. cost of raffle tickets sold by church
d. Fair Market Value (FMV) of property donated to Red Cross

39. Which of the following is NOT an allowable casualty loss?

a. theft of registered dog
b. accidental loss of diamond ring
c. vandalism damage to personal residence
d. roof damage from wind storm

40. Which of the following items is deductible as an itemized deduction?

a. vitamins for general health
b. mileage to take the family dog to the veterinarian
c. prescription medicines and drugs
d. federal income tax withheld

41. Which of the following items is deductible as an itemized deduction?

a. minimum education necessary to obtain a job in your profession
b. union dues
c. the cost of lunches you pack or buy
d. clothing for work which you also wear away from work

42. Which of the following items is NOT deductible as an itemized deduction?

a. state balance due from 1985 paid in 1986
b. interest on a loan for your son's car. You are not responsible for the loan.
c. home mortgage interest paid to a private individual
d. contributions to a qualified charity

43. Which of the following items is NOT deductible as an itemized deduction?

a. paying for damage to another person's property
b. safe deposit box (contains stock certificates)
c. book to assist you in preparing your tax return
d. the fair market value of clothing to a charity thrift store

44. Under the Tax Reform Act, the Zero Bracket Amount (ZBA) is being replaced by

a. the gross bracket amount
b. the standard deduction
c. the dependent exclusion
d. the earned income deduction

45. Beginning with 1987 tax returns, tax payers

a. will be required to report the social security numbers of all dependents age five and older
b. will be able to withhold the social security numbers of dependents
c. can get social security numbers for their dependent pets
d. will only be required to furnish the social security numbers of non-dependent children

46. Under the Tax Reform Act, unemployment compensation

a. must be reported on Form 1099
b. is fully taxable
c. will be taxable only for those individuals who make over $10,000 per year
d. is fully deductible

47. The Tax Reform Act

a. increased the deductible amount of capital gains to 70%
b. treats capital gains the same way as the old tax code
c. makes capital gains fully taxable
d. puts the maximum tax rate for capital gains at 40%

48. The dividend exclusion

a. has been increased under the Tax Reform Act
b. has been eliminated by the Tax Reform Act
c. is $300 for 1986
d. has been added to the interest exclusion under the Tax Reform Act

49. The Tax Reform Act changed the way scholarships and fellowships are handled. Taxes on scholarships and fellowships

a. are now fully eliminated
b. must be paid on the whole amount
c. must be paid on amounts not covering room and board
d. must be paid on amounts not covering qualified tuition and related expenses

50. Under the Tax Reform Act, exemption amounts

a. increase in phases to $2,000 by 1989
b. increase for 1987 to $2,000
c. decrease in phases to $2,000 by 1989
d. are phased out for all taxpayers

STUDENT EVALUATION OF INSTRUCTION

(This form is for district use only. This form should not be sent to Corporate Headquarters.)

☐ Basic ☐ Intermediate ☐ Level I ☐ Level II ☐ Level III ☐ T-Card ☐ Prework Training

Please assist us in evaluating the quality of the instruction and the classroom facilities used for your course by completing this questionnaire. Please circle the number that best represents your views.

Part One — Course Content

1. Did you learn what you expected to learn?
   No 1 2 3 4 5 Yes

2. What was your one MAIN reason for taking the course at this time?
   1 to help at my job
   2 to help get a new job
   3 for general interest
   4 for personal development
   5 to help solve difficulties I encounter

3. Was the material presented relevant and valuable to you?
   No 1 2 3 4 5 Yes

4. Was the material presented at an appropriate rate?
   No 1 2 3 4 5 Yes

5. Were the visual aids helpful to your learning
   No 1 2 3 4 5 Yes

6. To what extent were texts and suggested reading helpful to you?
   Not Helpful 1 2 3 4 5 Very Helpful

7. Was the course well organized, allowing a progression from one topic to another?
   No 1 2 3 4 5 Yes

Part Two — The Instructor

Please indicate how you felt about your instructor.

8. Never well prepared — 1 2 3 4 5 — Always well prepared

9. Poor knowledge of subject — 1 2 3 4 5 — Expert knowledge of subject

10. Provided no assistance — 1 2 3 4 5 — Provided adequate assistance

11. Poor class presentation — 1 2 3 4 5 — Class presentation excellent

12. Unfair in class and grading — 1 2 3 4 5 — Fair in class and grading

13. Visual aids not used effectively — 1 2 3 4 5 — Visual aids used effectively

14. Poor classroom control — 1 2 3 4 5 — Excellent classroom control

15. Poor communication with students — 1 2 3 4 5 — Effective communication with students

Part Three — The Classroom and Facilities

Please indicate your evaluation of the following:

16. Classroom arrangement — Needs Improvement 1 2 3 4 5 Excellent

17. Lighting — Needs Improvement 1 2 3 4 5 Excellent

18. Ventilation — Needs Improvement 1 2 3 4 5 Excellent

19. Furniture — Needs Improvement 1 2 3 4 5 Excellent

20. Temperature control — Needs Improvement 1 2 3 4 5 Excellent

Part Four — General Information

21. Do you feel you have mastered the subject matter of this course?

Do not know any more now 1 2 3 4 5 Totally

22. Considering all your previous answers, how do you rate this course overall?

Very Poor 1 2 3 4 5 Excellent

Average

General comments and suggestions for improvement: _____

Name _____ TA (if applicable) _____

STUDENT EVALUATION OF INSTRUCTIONAL MATERIALS

Please use the scale given below to answer the following questions.

1 - Strongly Agree
2 - Agree
3 - Neither Agree Nor Disagree
4 - Disagree
5 - Strongly Disagree

1.	I learned what I expected to learn.	1	2	3	4	5
2.	The material presented was relevant to me.	1	2	3	4	5
3.	The material presented was valuable to me.	1	2	3	4	5
4.	The text was written clearly.	1	2	3	4	5
5.	The text was well-organized.	1	2	3	4	5
6.	The text presented the tax material in a logical sequence.	1	2	3	4	5
7.	The examples were helpful illustrations of the material.	1	2	3	4	5
8.	The text is written in a comprehensible fashion.	1	2	3	4	5
9.	The text is written in an interesting fashion.	1	2	3	4	5
10.	The text presented the tax material in sufficient detail.	1	2	3	4	5
11.	The tables enhance the overall presentation of the material.	1	2	3	4	5
12.	I have mastered the subject matter of this course.	1	2	3	4	5

13. Considering all your previous answers, use the scale below to rate the instructional materials overall.

1 - very poor
2 - poor
3 - neutral
4 - good
5 - excellent 1 2 3 4 5

Name _____

The following questions are intended to give us an idea of the general demographic characteristics of the group of people who are taking this training. The information will be summarized only for the group. Please answer each question as accurately as possible. Please circle the one best answer to each question. Please do not skip any questions.

1. How much formal education have you had? (Please circle the highest formal education completed)

    1. Did not complete high school
    2. Completed high school
    3. Attended college and did not receive a bachelor's degree
    4. Currently enrolled as a bachelor's candidate
    5. Attended college and received a bachelor's degree
    6. Attended graduate school and did not receive an advanced degree (e.g., Masters, Doctorate, J.D., M.D.)
    7. Currently attending graduate school
    8. Attended graduate school and received an advanced degree (e.g., Masters, Doctorate, J.D., M.D.)

2. If you are currently attending college, what year are you?

    1. Freshman
    2. Sophmore
    3. Junior
    4. Senior
    5. Master's degree candidate
    6. Doctoral degree candidate
    7. Other (Please specify _____)

3. What is your age in years? (Please fill in the blank) _____

4. What is your sex? 1. Female 2. Male

5. If you attended or are attending college, what is your major area of study?

    1. Business
    2. Social Studies
    3. Humanities
    4. Engineering
    5. Education
    6. Other. Please specify _____

6. What is your exact area of study/major? _____

7. What is your marital status?
    1. Married
    2. Widowed
    3. Divorced
    4. Separated
    5. Never Married

Name _____

8. How many children do you have?

0. None
1. One
2. Two
3. Three
4. Four
5. Five or more

9. In which of these groups did your total family income, from all sources fall last year, before taxes?

1. Under $10,000
2. $10,001 - $15,000
3. $15,001 - $20,000
4. $20,001 - $25,000
5. $25,001 - $30,000
6. $30,001 - $35,000
7. $35,001 - $40,000
8. $40,001 - $45,000
9. $45,001 - $50,000
10. over $50,000

10. What is your previous experience with tax preparation?

1. I have never filed federal income tax returns.
2. A paid preparer completes my income tax returns for me.
3. A family member completes my income tax return for me.
4. I file jointly with my spouse who completes our tax return for us.
5. I file jointly with my spouse and we both work on the return.
6. I have completed and filed a tax return once.
7. I have completed and filed tax returns at least two times and no more than five times.
8. I have completed and filed tax returns at least six or more times.
9. I have completed tax returns for others for free.
10. I have been a paid tax preparer for others.

11. If you filed federal taxes, what form did you complete last year?

1. 1040-EZ (for singles with taxable income less than $50,000 and no itemized deductions)
2. 1040-A (for any one with taxable income less than $50,000 and no itemized deductions)
3. 1040 (for any one with any level of taxable income and itemized deductions)

Name _____

12. What is your general accounting background (i.e., what accounting or bookkeeping courses have you taken; what job experience have you had)?

13. What tax preparation courses have you had?

NAME_____

PART A.

Directions: After reading each of the following statements, use the following scale and enter on the line below each statement a number between 0 and 100 which best describes what you think the probability is of receiving AT LEAST a 70% in the course given the behavior described.

0	25	50	75	100
No chance at all	A slight chance	A 50/50 chance	A good chance	Maximum chance

1. If you don't study at all for the course, what are your chances of getting at least a 70% in the course?

_____

2. If you put a minimum amount of effort into the course, what are your chances of getting at least a 70% in the course?

_____

3. If you put a moderate amount of effort into studying for this course, what are your chances of getting at least a 70% in the course?

_____

4. If you put a maximum amount of effort into studying for this course, what are your chances of getting at least a 70% in the course?

_____

NAME _____

PART B.
Directions: Please read each of the following statements carefully. Then indicate how much you would desire or would not desire each of the following by circling the appropriate response.

1 - Extremely Undesirable
2 - Undesirable
3 - Does Not Matter
4 - Desirable
5 - Extremely Desirable

1. Increasing your knowledge of tax return preparation 1 2 3 4 5

2. Having a feeling of personal satisfaction 1 2 3 4 5

3. Having others think they are more intelligent than you 1 2 3 4 5

4. Being able to accurately complete your own tax return 1 2 3 4 5

5. Feeling depressed 1 2 3 4 5

6. Getting a job 1 2 3 4 5

7. Improving my present (or future) employer's evaluation of me 1 2 3 4 5

8. Being prepared for the business world 1 2 3 4 5

9. Receiving the research participation bonus 1 2 3 4 5

PART C.
Directions: In the preceding section you indicated how much you would desire or not desire each of the things listed. Now consider what your chances of attaining these things are if you get at least a 70% in this course. Please use the following scale and enter a number between 0 and 100 in the blank following each statement.

0	25	50	75	100
No chance at all	A slight chance	A 50/50 chance	A good chance	Maximum chance

1. Increasing your knowledge of tax return preparation _____

2. Having a feeling of personal satisfaction _____

3. Having others think they are more intelligent than you _____

4. Being able to accurately complete your own tax return _____

5. Feeling depressed _____

NAME _____

Part C (cont.)

0	25	50	75	100
No chance at all	A slight chance	A 50/50 chance	A good chance	Maximum chance

6. Getting a job _____

7. Improving my present (or future) employer's evaluation of me _____

8. Being prepared for the business world _____

9. Receiving the research participation bonus _____

PART D.

Directions: Please read each of the following statements carefully. Then indicate how much you agree or do not agree with the statement by circling the appropriate response.

1 - Strongly Disagree
2 - Disagree
3 - Does Not Matter
4 - Agree
5 - Strongly Agree

1. I am willing to exert considerable effort in studying for this tax course. 1 2 3 4 5

2. I am trying to learn as much as I can for this tax course. 1 2 3 4 5

3. I have a strong desire to learn the information emphasized in this tax course. 1 2 3 4 5

4. Doing well in this tax course is important to me. 1 2 3 4 5

5. I wish I didn't have to take this tax course. 1 2 3 4 5

6. I will get more out of this class than most people. 1 2 3 4 5

LIST OF REFERENCES

LIST OF REFERENCES

Baldwin, T. T. & Ford, J. K. (1988). Transfer of training: A review and directions for future research. Personnel Psychology, 41, 63-105.

Bangert-Drowns, R. L. (1986). Review of developments in meta-analytic method. Psychological Bulletin, 99, 388-399.

Baumgartel, H., Sullivan, G. J. & Dunn, L. E. (1978). How organizational climate and personality affect the payoff from advanced management training session. Kansas Business Review, 5, 1-10.

Berkowitz, L. & Donnerstein, E. (1982). External validity is more than skin deep: Some answers to criticisms of laboratory experiments. American Psychologist, 37, 245-257.

Berliner, D. C. & Cahen, L. S. (1973). Trait-treatment interaction and learning. Review of Research in Education, 1, 58-94.

Beswick, D. G. & Tallmadge, G. K. (1971). Reexamination of two learning style studies in the light of the cognitive process theory of curiosity. Journal of Educational Psychology, 62, 456-462.

Bobko, P., Karren, R. & Parkington, J. J. (1983). Estimation of standard deviations in utility analyses: An empirical test. Journal of Applied Psychology, 68, 170-176.

Boudreau, J. W. (1983a). Economic considerations in estimating the utility of human resource productivity improvement programs. Personnel Psychology, 36, 551-576.

Boudreau, J. W. (1983b). Effects of employee flows on utility analyses of human resource productivity improvement programs. Journal of Applied Psychology, 68, 396-406.

Boudreau, J. W. & Berger, C. J. (1985). Decision-theoretic utility analysis applied to employee separations and acquisitions. Journal of Applied Psychology [monograph], 70, 581-612.

Boudreau, J. W. & Rynes, S. R. (1985). Role of recruitment in staffing utility analysis. Journal of Applied Psychology, 70, 354-366.

Bracht, G. H. (1970). Experimental factors related to aptitude-treatment interactions. Review of Educational Research, 40, 627-645.

Bracht, G. H. & Glass, G. V. (1968). The external validity of experiments. American Educational Research Journal, 5, 437-474.

Brief, A. P. & Hollenbeck, J. R. (1985). An exploratory study of self-regulating activities and their effects on job performance. Journal of Occupational Behaviour, 6, 197-208.

Brogden, H. E. (1946). On the interpretation of the correlation coefficient as a measure of predictive efficiency. Journal of Educational Psychology, 37, 65-67.

Brogden, H. E. (1949). When testing pays off. Personnel Psychology, 2, 171-185.

Brooks, F. R., Ebner, D. G., Manning, D. T. & Balson, P. M. (1985). Influence of academic ability on partners' mastery of a stressful motor task. Journal of Instructional Psychology, 12, 19-23.

Buch, G. (1984). An investigation of the effectiveness of pilot judgment training. Human Factors, 26, 557-564.

Burns, R. B. (1980). Relation of aptitudes to learning at different points in time during instruction. Journal of Educational Psychology, 72, 785-795.

Carrier, C. & McNergney, R. (1979). Interaction research: Can it help individualize instruction. Educational Technology, 19, 40-45.

Cascio, W. F. (1982). Costing human resources: The financial impact of behavior in organizations. Boston: Kent.

Cascio, W. F. & Silbey, V. (1979). Utility of the assessment center as a selection device. Journal of Applied Psychology, 64, 107-118.

Cohen, J. (1968). Multiple regression as a general data-analytic system. Psychological Bulletin, 70, 426-443.

Cohen, J. & Cohen, P. (1983). Applied multiple regression/correlation analysis for the behavioral sciences, (2nd ed). Hillsdale, NJ: Lawrence Erlbaum.

Cook, T. D. & Campbell, D. T. (1979). Quasi-experimentation: Design and analysis issues for field settings. Boston: Houghton Mifflin.

Cronbach, L. J. (1957). The two disciplines of scientific psychology. American Psychologist, 12, 671-684.

Cronbach, L. J. (1975). Beyond the two disciplines of scientific psychology. American Psychologist, 30, 116-127.

Cronbach, L. J. & Gleser, G. C. (1965). Psychological tests and personnel decisions (2nd ed.). Champaign, IL: University of Illinois Press.

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

Dipboye, R. L. & Flanagan, M. F. (1979). Research settings in industrial and organizational psychology: Are findings in the field more generalizable than in the laboratory? American Psychologist, 34, 141-150.

Dodrill, C. B. (1981). An economical method for the evaluation of general intelligence in adults. Journal of Consulting and Clinical Psychology, 49, 668-673.

Dodrill, C. B. (1983). Long-term reliability of the Wonderlic Personnel Test. Journal of Consulting and Clinical Psychology, 51, 316-317.

Downs, S. (1970). Predicting training potential. Personnel Management, 3, 26-28.

Duncan, K. D. (1971). Long-term retention and transfer of an industrial search skill. British Journal of Psychology, 62, 439-448.

Eaton, N. K., Wing, H. & Mitchell, K. J. (1985). Alternative methods of estimating the dollar value of performance. Personnel Psychology, 38, 27-40.

Entwistle, N. J. & Entwistle, D. (1970). The relationships between personality, study methods and academic performance. British Journal of Educational Psychology, 40, 132-141.

Fisher, C. D. & Gitelson, R. (1983). A meta-analysis of the correlates of role conflict and ambiguity. Journal of Applied Psychology, 68, 320-333.

Fromkin, H. L. & Streufert, S. (1976). Laboratory experimentation. In M. D. Dunnette (Ed.), Handbook of Industrial-Organizational Psychology (pp. 415-465). Chicago: Rand McNally.

Gehlbach, R. D. (1979). Individual differences: Implications for instructional theory, research, and innovation. Educational Researcher, 8, 8-14.

Gill, R. W. T. (1982). A trainability concept for management potential and an empirical study of its relationship with intelligence for two managerial skills. Journal of Occupational Psychology, 55, 139-147.

Goldstein, I. L. (1980). Training in work organizations. Annual Review of Psychology, 31, 229-272.

Gordon, L. V. (1955). Time in training as a criterion of success in radio code. Journal of Applied Psychology, 39, 311-313.

Gordon, M. E. & Cohen, S. L. (1973). Training behavior as a predictor of trainability. Personnel Psychology, 26, 261-272.

Gordon, M. E. & Kleiman, L. S. (1976). The prediction of trainability using a work sample test and an aptitude test: A direct comparison. Personnel Psychology, 29, 243-253.

Grimsley, G. & Jarrett, H. F. (1973). The relation of past managerial achievement to test measures obtained in the employment situation: Methodology and results. Personnel Psychology, 26, 31-48.

Grimsley, G. & Jarrett, H. F. (1975). The relation of past managerial achievement to test measures obtained in the employment situation: Methodology and results-II. Personnel Psychology, 28, 215-231.

Guion, R. M. (1965). Synthetic validity in a small company: A demonstration. Personnel Psychology, 18, 49-63.

Hunt D. E. (1975). Person-environment interaction: A challenge found wanting before it was tried. Review of Educational Research, 45, 209-230.

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. & Jackson, G. B. (1982). Meta-analysis: Cumulating research findings across studies. Beverly Hills: Sage.

James, N. E. (1962). Personal preference for method as a factor in learning. Journal of Educational Psychology, 43, 43-47.

Jenson, A. R. (1977). An examination of culture bias in the Wonderlic Personnel Test. Intelligence, 1, 51-64.

Jensen, A. R. (1986). g: Artifact or reality? Journal of Vocational Behavior, 29, 301-331.

Johnson, D. W., Maruyama, G., Johnson, R., Nelson, D. & Skon, L. (1981). Effects of cooperative, competitive, and individualistic goal structures on achievement: A meta-analysis. Psychological Bulletin, 89, 47-62.

Keppel, G. (1982). Design and analysis: A researccher's handbook, 2nd ed. Englewood Cliffs, N.J.: Prentice-Hall.

Kirkpatrick, D. L. (1967). Evaluation of training. In E. L. Craig and L. R. Bittel (Eds.), Training and development handbook, pp. 87-112. New York: McGraw Hill.

Klein, H. J. (1987). Reactions for goal setting and feedback: A test of a control theory model of work motivation. Unpublished doctoral dissertation, Michigan State University, East Lansing, MI.

Kulik, J. A. (1981). Integrating findings from different levels of instruction. Paper presented at the 65th Annual meeting of the American Educational Research Association, Los Angeles, CA.

Kulik, J. A., Kulik, C. C. & Cohen, P. A. (1979). A meta-analysis of outcome studies of Keller's personalized system of instruction. American Psychologist, 34, 307-318.

Lewin, G. (1935). Dynamic theory of personality. New York: McGraw-Hill.

Love, K. G. & O'Hara, K. (1987). Predicting job performance of youth trainees under a Job Training Partnership Act program (JTPA): Criterion validation of a behavior-based measure of work maturity. Personnel Psychology, 40, 323-340.

Lunneborg, C. E. & Lunneborg, P. W. (1967). Uniqueness of selected employment aptitude tests to a general academic guidance battery. Educational and Psychological Measurement, 27, 953-960.

McCombs, B. L. & McDaniel, M. A. (1982). On the design of adaptive treatments for individualized instructional systems. Educatiionnal Psychologist, 16, 11-22.

McGehee, W. (1948). Cutting training waste. Personnel Psychology, 1, 331-340.

Merrill, M. D. (1975). Learner control: Beyond aptitude-treatment interaction. AV Communication Review, 23, 217-226.

Millman, J. (1974). Instruction planning and management. In J. Blaney, I. Housego & G. McIntosh (Eds), Program development in education. Vancouver: University of British Columbia.

Miles, M. B. (1965). Changes during and following laboratory training: A clinical-experimental study. Journal of Applied Behavioral Science, 1, 215-242.

Murray, H. G. (1983). Low-inference classroom teaching behaviors and student ratings of college teaching effectiveness. Journal of Educational Psychology, 75, 138-149.

Neel, R. G. & Dunn, R. E. (1960). Predicting success in supervisory training programs by the use of psychological tests. Journal of Applied Psychology, 44, 358-360.

Noe, R. A. & Schmitt, N. (1986). The influence of trainee attitudes on training effectiveness: Test of a model. Personnel Psychology, 39, 497-523.

Orwin, R. G. (1983). A fail-safe N for effect size in meta-analysis. Journal of Educational Statistics, 8, 157-159.

Parsons, C. K., Herold, D. M. & Turlington, B. (1981). Individual differences in performance feedback preferences. Academy of Management Proceedings, 166-171.

Pedhazur, E. J. (1982). Multiple regression in behavioral research: Explanation and prediction (2nd ed). New York: Holt, Rinehart & Winston.

Pintrich, P. R., Cross, D. R., Kozma, R. B & McKeachie, W. J. (1986). Instructional psychology. Annual Review of Psychology, 37, 611-651.

Ragsdale, R. G. (1980). Empirically based myths: Astrology, biorhythms, ATIs. Canadian Journal of Education, 5, 40-51.

Robertson, I. & Downs, S. (1979). Learning and the prediction of performance: Development of trainability testing in the United Kingdom. Journal of Applied Psychology, 64, 42-50.

Robertson, I. T. & Mindel, R. M. (1980). A study of trainability testing. Journal of Occupational Psychology, 53, 131-138.

Rosenthal, R. (1978). Combining results of independent studies. Psychological Bulletin, 85, 185-193.

Rosenthal, R. (1979). The "file drawer problem" and tolerance for null results. Psychological Bulletin, 86, 638-641.

Ross, P. C. & Houtz, J. C. (1979). Students' ratings of satisfaction with industrial training predicted from instructors' leadership style and trainees' personality characteristics. Psychological Reports, 45, 63-73.

Ruch, F. L. & Ruch, W. W. (1980). Employee Aptitude Survey: Technical Report. Los Angeles: Psychological Services, Inc.

Ryman, D. H. & Biersma, R. J. (1975). Attitudes predictive of diving training success. Personnel Psychology, 28, 181-188.

Salomon, G. (1972). Heuristic models for the generation of aptitude-treatment interaction hypotheses. Review of Educational Research, 42, 327-343.

Schlesinger, J. M. (1987, June 29). Job-security contracts are catching on: Unions refocus on employment guarantees. The Wall Street Journal, p. 6.

Schmidt, F. L., Hunter, J. E., McKenzie, R. C. & Muldrow, T. W. (1979). Impact of valid selection procedures on work-force productivity. Journal of Applied Psychology, 64, 609-626.

Schmidt, F. L., Hunter, J. E. & Pearlman, K. (1982). Assessing the economic impact of personnel programs on workforce productivity. Personnel Psychology, 35, 333-347.

Schmidt, F. L., Mack, M. J. & Hunter, J. E. (1984). Selection utility in the occupation of U. S. Park Ranger for three modes of test use. Journal of Applied Psychology, 69, 490-497.

Schmitt, N., Gooding, R. Z., Noe, R. A. & Kirsch, M. (1984). Metaanalyses of validity studies published between 1964 and 1982 and the investigation of study characteristics. Personnel Psychology, 37, 407-422.

Schneider, B. & Schmitt, N. (1986). Staffing organizations (2nd ed.). Glenview, IL: Scott, Foresman.

Semerad, R. D. (1987, February 9). 2000: Labor shortage looms. Industry Week, pp. 38-40.

Siegel, A. I. (1983). The miniature job training and evaluatin approach: Additional findings. Personnel Psychology, 36, 41-56.

Siegel, A. I. & Bergman, B. A. (1975). A job learning approach to performance prediction. Personnel Psychology, 28, 325-339.

Smith, M. C. & Downs, S. (1975). Trainability assessments for apprentice selection in shipbuilding. Journal of Occupational Psychology, 48, 39-43.

Snow, R. E. (1978). Theory and method for research on aptitude processes. Intelligence, 2, 225-278.

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

Solomon, D. (1980). Finding optimal instructional programs and procedures for individual students. Teaching Sociology, 7, 303-326.

SPSS Inc. (1986). SPSSX: User's Guide (2nd ed.). Chicago: Author.

Stone, E. F. & Hollenbeck, J. R. (1984). Some issues associated with the use of moderated regression. Organizational Behavior and Human Performance, 34, 195-213.

Tallmadge, G. K. & Shearer, J. W. (1969). Relationships among learning styles, instructional methods, and the nature of learning experiences. Journal of Educational Psychology, 60, 222-230.

Tallmadge, G. K. & Shearer, J. W. (1971). Interactive relationships among learner characteristics, types of learning, instructional methods, and subject matter variables. Journal of Educational Psychology, 62, 31-38.

Taylor, C. W. (1952). Pre-testing saves training costs. Personnel Psychology, 5, 213-239.

Taylor, E. K. & Tajen, C. (11948). Selection for training: Tabulating equipment operators. Personnel Psychology, 1, 341-348.

Taylor, H. C. & Russell, J. T. (1939). The relationship of validity coefficients to the practical effectiveness of tests in selection. Journal of Applied Psychology, 23, 565-578.

Tenopyr, M. L. (1969). The comparative validity of selected leadership scales relative to success in production management. Personnel Psychology, 22, 77-85.

Tobias, S. (1976). Achievement treatment interactions. Review of Educational Research, 46, 61-74.

Tobias, S. (1982). When do instructional methods make a difference? Educational Researcher, 11, 4-9.

Tubiana, J. H. & Ben-Shakhar, G. (1982). An objective group questionnaire as a substitute for a personal interview in the prediction of success in military training in Israel. Personnel Psychology, 35, 349-352.

UAW's Bieber lists aims for future negotiations. (1987, April 13). The Wall Street Journal, p. 5.

Wernimont, P. F. & Campbell, J. P. (1968). Signs, samples and criteria. Journal of Applied Psychology, 52, 372-376.

Wexley, K. N. (1984). Personnel training. Annual Review of Psychology, 35, 519-551.

Williams, B. B., Sauser, W. I., Kemery, E. R. & Dyer, F. N. (1982, March). Intelligence and physical fitness as predictors of success in early infantry training. Paper presented at the 28th Annual Meeting of the Southeastern Psychological Association, New Orleans, LA.

Winer, B. J. (1971). Statistic principles in experimental design (2nd ed). New York: McGraw-Hill.

Wise, K. C. & Okey, J. R. (1983). A meta-analysis of the effects of various science teaching strategies on achievement. Journal of Research in Science Teaching, 20, 419-435.

Wonderlic, E. F. & Associates, Inc. (1983). Wonderlic Personnel Test Manual. Northfield, IL: Author.

Zedeck, S. (1971). Problems with the use of moderator variables. Psychological Bulletin, 76, 295-310.

Zink, D. L. (1982, October). Standards for time taken in self-paced training. Paper presented at the 26th Annual Meeting of the Human Factors Society in Seattle, WA.

Studies with statistics included in Meta-Analysis.

Adams, V. M. & McLeod, D. B. (1979). The interaction of field dependence/independence and the level of guidance of mathematics instruction. Journal for Research in Mathematics Education, 10, 347-355.

Cicchelli, T. (1982). Effects of direct and indirect instruction patterns and prior achievement on post course achievement. Journal of Instructional Psychology, 9, 176-187.

Elawar, M. C. & Corno, L. (1985). A factorial experiment in teachers' written feedback on student homework: Changing teacher behavior a little rather than a lot. Journal of Educational Psychology, 77, 162-173.

Gettinger, M. (1983). Effects of learner ability and instructional modifications on time needed for learning and retention. Journal of Educational Research, 76, 362-369.

Greene, J. C. (1980). Individual and teacher/class effects in aptitude treatment studies. American Educational Research Journal, 17, 291-302.

Janicki, T. C. & Peterson, P. L. (1981). Aptitude-treatment interaction effects of variations in direct instruction. American Educational Research Journal, 18, 63-82.

Kallison, J. M., Jr. (1986). Effects of lesson organization on achievement. American Educational Research Journal, 23, 337-347.

Koran, M. L. & Koran, J. J., Jr. (1980). Interaction of learner characteristics with pictorial adjuncts in learning from science text. Journal of Research in Science Teaching, 17, 477-483.

Kozma, R. B. (1982). Instructional design in a chemistry laboratory course: The impact of structure and aptitudes on performance and attitudes. Journal of Research in Science Teaching, 19, 261-270.

Lehman, J. R., Koran, J. J., Jr. & Koran, M. L. (1984). Interaction of learner characteristics with learning from three models of the periodic table. Journal of Research in Science Teaching, 21, 885-893.

McKinney, C. W., Larkins, A. G., Ford, M. J. & Davis, J. C., III. (1983). The effectiveness of three methods of teaching social studies to fourth-grade students: An aptitude-treatment interaction study. American Educational Research Journal, 20, 663-670.

McLeod, D. B. & Adams, V. M. (1980). Aptitude-treatment interaction in mathematics instruction using expository and discovery methods. Journal for Research in Mathematics Education, 11, 225-234.

McLeod, D. B. & Briggs, J. T. (1980). Interactions of field independence and general reasoning with inductive instruction in mathematics. Journal for Research in Mathematics Education, 11, 84-103.

Pascarella, E. T. (1978). Interactive effects of prior mathematics preparation and level of instructional support in college calculus. American Educational Research Journal, 15, 275-285.

Peterson, P. L. (1977). Interactive effects of student anxiety, achievement orientation and teacher behavior on student achievement and attitude. Journal of Educational Psychology, 69, 779-792.

Peterson, P. L. (1979). Aptitude X treatment interaction effects of teacher structuring and student participation in college instruction. Journal of Educational Psychology, 71, 521-533.

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, 677-687.

Peterson, P. L., Janicki, T. C. & Swing, S. R. (1980). Aptitude-treatment interaction effects of three social studies teaching approaches. American Educational Research Journal, 17, 339-360.

Peterson, P. L., Janicki, T. C. & Swing, S. R. (1981). Ability X treatment interaction effects on children's learning in large-group and small-group approaches. American Educational Research Journal, 18, 453-473.

Ross, S. M. & Rakow, E. A. (1981). Learner control versus program control as adaptive strategies for selection of instructional support on math rules. Journal of Educational Psychology, 73, 745-753.

Smith, L. R. (1985). The effect of lesson structure and cognitive level of questions on student achievement. Journal of Experimental Education, 54, 44-49.

Stinard, T. A. & Dolphin, W. D. (1981). Which students benefit from self-paced mastery instruction and why. Journal of Educational Psychology, 73, 754-763.

Threadgill, J. (1979). The interaction of learner aptitude with types of questions accompanying a written lesson on logical implications. Journal for Research in Mathematics Education, 10, 337-346.

Winn (1980). The effect of block-word diagrams on the structuring of science concepts as a function of general ability. Journal of Research in Science Teaching, 17, 201-211.

Winn (1981). Effect of attribute highlighting and diagrammatic organization on identification and classification. Journal of Research in Science Teaching, 18, 23-32.

Studies not included due to insufficient statistics.

Bing, S. (1982). Role of adjunct questions and reading ability levels on rote and conceptual learning from prose. Instructional Science, 11, 129-138.

Corno, L. (1979). A hierarchical analysis of selected naturally occurring aptitude-treatment interactions in the third grade. American Education Research Journal, 16, 391-409.

Hamilton, R. J. (1986). Role of adjunct questions and subject ability levels on the learning of concepts from prose. American Educational Research Journal, 23, 87-94.

Koran, J. J., Jr., Koran, M. L. & Baker, S. D. (1980). Differential response to cueing and feedback in the acquisition of an inductively presented biological concept. Journal of Research in Science Teaching, 17, 167-172.

Ross, S. M., Rakow, E. A. & Bush, A. J. (1980). Instructional adaptation for self-managed learning systems. Journal of Educational Psychology, 72, 312-320.