A METHODOLOGY FOR THE EVALUATION OF TRAINING EFFECTIVENESS DURING EARLY PHASE DEFENSE ACQUISITION

A Dissertation
Presented to
The Academic Faculty

by

Cynthia Chalese Brown

In Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy in the
School of Aerospace Engineering

Georgia Institute of Technology
August 2014

A METHODOLOGY FOR THE EVALUATION OF TRAINING EFFECTIVENESS DURING EARLY PHASE DEFENSE ACQUISITION

Approved by:

Dimitri N. Mavris, Professor
Advisor and Committee Chair
School of Aerospace Engineering
Georgia Institute of Technology

Daniel P. Schrage, Professor
School of Aerospace Engineering
Georgia Institute of Technology

Kelly A. Griendling, Ph.D.
School of Aerospace Engineering
Georgia Institute of Technology

Steven J. Tourville, Ph.D.
Human Performance Engineering Fellow
and Chief Training Architect
Lockheed Martin Corporation

James H. Crutchfield Jr., Ph.D.
Human Factors Design Engineer
Lockheed Martin Corporation

Date Approved: June 27, 2014

To my family, my friends, and to my co-workers.

ACKNOWLEDGEMENTS

I would like to thank my husband, Marvin, for his unwavering support of all that I set out to do. Without his constant fortitude and understanding spirit, this effort would not have been as fulfilling. To my parents, thank you for the steadfast motivation you have instilled in me. I would be remiss, if I did not thank Dr. Johnson for his vision and sponsorship. To Dr. Bloschock, thank you for being such an encouragement to me; your positive attitude and support is always uplifting. I would like to thank my friends and colleagues at ASDL. Special thanks to Michael Ellis, Dr. Salmon, Burak Bagdatli, Annie Jones, Michael Miller, and Eric Brown for the time they have spent providing insight and assistance in various areas. I would also like to express gratitude to my committee, Dr. Dimitri Mavris, Dr. Kelly Griendling, Dr. Daniel Schrage, Dr. Steven Tourville, and Dr. James Crutchfield. To my committee chairman and advisor, Dr. Mavris, thank you for sharing the vast engineering knowledge you have, and your willingness and dedication in developing scholars. I am grateful to you for allowing me the opportunity to be a part of ASDL. To Dr. Griendling, you have been an absolute pleasure to work with. The constructive feedback, criticism, and support throughout the last two years have increased my abilities and the results of my research. To Dr. Tourville, thank you for being fully committed and generous with your time. You have exceeded my expectations. Being able to discuss your perspectives has proven invaluable. Finally, I am grateful for all the support and encouragement I have received from my family, friends, and co-workers. The development of this thesis has been both rewarding and challenging, but your prayers, kind words, and constant belief in me have pushed me through. *Thank you.*

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF SYMBOLS AND ABBREVIATIONS

AA	Application and Analysis
ABET	Accreditation Board for Engineering and Technology
ACC	Air Combat Command
ACF	Aircraft Flying
ACL	All Critical Learning
ACS	Aircraft Standing
AFSOC	Air Force Special Operations Command
ASDL	Aerospace Systems Design Laboratory
ASTD	American Society for Training and Development
ATD	Aircraft Training Device
BCR	Benefit-Cost Ratio
BMDS	Ballistic Missile Defense System
CB	Cost Benefit
CBA	Capability Based Assessment
CBAT	Common Battlefield Airmen Training
CBT	Computer-Based Training
CDT	Component Display Theory
CE	Cost Effectiveness
CEAT	Cost-Effective Analysis of Training
CER	Cost-Effectiveness Ratio
CJCS	Chairman of the Joint Chiefs of Staff

CoE	Consequence of Error
COEA	Cost and Operational Evaluation Analysis
CONOPS	Concepts of Operations
CRM	Crew Resource Management
CU	Cost Utility
CUR	Cost-Utility Ratio
DAS	Defense Acquisition System
DAU	Defense Acquisition University
DIFE	Difficulty, Importance, Frequency, Consequence of Error
DMO	Distributed Mission Operations
DoD	Department of Defense
DOTMLPF-P	Doctrine, Organization, Training, Materiel, Leadership and Education, Personnel, Facilities, and Policy
DCR	DOTMLPF-P Change Recommendation
ECR	Effectiveness-Cost Ratio
ECS	Expeditionary Combat Skills
Ex	Example
e-learning	Electronic Learning
etc	Etcetera
FAA	Federal Aviation Administration
FY	Fiscal Year
GAO	Government Accountability Office
GT	Georgia Institute of Technology

Hr.	Hour
IADS	Integrated Air Defense System
ICD	Initial Capabilities Document
ICW	Interactive Courseware
INCOSE	International Council on Systems Engineering
ISD	Instructional Systems Development
JCA	Joint Capabilities Areas
JCCD	Joint Concepts to Capabilities Division
JCIDS	Joint Capabilities Integration and Development System
JCS	Joint Chiefs of Staff
JROC	Joint Requirements Oversight Council
KC	Knowledge and Comprehension
KPP	Key Performance Parameters
KSA	Knowledge, Skills, Attitudes
LB	Lower Bound
LO	Learning Objective
M&S	Modeling and Simulation
MCS	Monte Carlo Simulation
MDA	Missile Defense Agency
MEC	Mission Essential Competency
MoE	Measure of Effectiveness
MPEET	A Methodology to Predict and Evaluate the Effectiveness of Training

MT	Method Type
NATO	North Atlantic Treaty Organization
NRC	Nuclear Regulatory Commission
NTL	National Training Laboratories
NTSB	National Transportation Safety Board
OEC	Overall Evaluation Criterion
OJT	On-the-Job Training
ONR	Office of Naval Research
PPBE	Planning, Programming, Budgeting and Execution
PTT	Part Task Trainer
QTP	Qualification Training Program
r	Correlation
R&D	Research and Development
ROE	Return on Expectation
ROI	Return on Investment
SAT	Systems Approach to Training
SE	Synthesis and Evaluation
SIS	Suggested Instructional Strategies
SME	Subject Matter Expert
SOLO	Structure of Observed Learning Outcomes
SoS	System of Systems
TA	Training Asset
TACWAR	A Tactical Warfare Computer-Based Simulation Model

TEM	Threat Error Management
TM	Training Method
TRADOC	United States Army Training and Doctrine Command
UB	Upper Bound
U.S.	United States
USA	United States of America
USAF	United States Air Force
WP	Warsaw Pact
WWII	World War II
3-D	Three Dimensional

SUMMARY

Today's economic environment requires for a greater emphasis to be placed on the development of cost-effective solutions to meet military capability based requirements. The Joint Capabilities Integration and Development System (JCIDS) process is designed to identify materiel and non-materiel solutions to fill defense department capability requirements and gaps. Non-materiel solutions include: Doctrine, Organization, Training, Materiel, Leadership and Education, Personnel, Facilities, and Policy (DOTMLPF-P) changes. JCIDS specifies that all non-materiel solutions be analyzed and recommendations be made accordingly following a capability-based assessment (CBA). Guidance for performing CBA analysis provides minimal information on how to predict training effectiveness and as a result training investments are not properly assessed and considered as a viable alternative. Investigations into the ability to predict versus evaluate training performance and to quantify uncertainty in training system design are two identified gaps in the capability of existing training evaluation methods. To address these issues, a Methodology to Predict and Evaluate the Effectiveness of Training (MPEET) has been developed. To address the gap in predictive capability MPEET uses primary elements of learning theory and instructional design to predict the cost-effectiveness of a training program, and recommends training alternatives based on decision-maker preferences for each of the cost and effectiveness criteria. The use of educational and instructional theory involves developing and ensuring human performance requirements will be met after training. Utility theory is used to derive an overall criterion consisting of both cost and effectiveness attributes. MPEET uses this criterion as a key variable in determining how to properly allocate resources to gain maximum training effectiveness.

To address the gap in quantifying uncertainty in training performance, probability theory is used within a modeling and simulation environment to create and evaluate previously deterministic variables. Effectiveness and cost variables are assigned probability distributions that reflect the applicable range of uncertainty. MPEET is a systems engineering based decision-making tool. It enhances the instructional design process, which is rooted in the fields of education and psychology, by adding an objective verification step to determine how well instructional strategies are used in the design of a training program to meet the required learning objectives.

A C-130J pilot case study is used to demonstrate the application of MPEET and to show the plausibility of the approach. For the case study, metrics are derived to quantify the requirement for knowledge, skills, and attitudes in the C-130J pilot training system design. Instructional strategies were defined specifically for the C-130J training program. Feasible training alternatives were generated and evaluated for cost and effectiveness. Using information collected from decision-maker preferences for cost and effectiveness variables, a new training program is created and comparisons are made to the original. The case study allows tradeoffs to be performed quantitatively between the variable importance weightings and mean value of the probabilistic variables.

Overall, it is demonstrated that MPEET provides the capability to assess the cost and effectiveness of training system design and is an enabler to the inclusion of training as an independent non-materiel alternative solution during the CBA process. Although capability gaps in the defense acquisition process motivated the development of MPEET its applicability extends to any training program following the instructional design process where the assumed constraints are not prohibitive.

CHAPTER 1

INTRODUCTION

The continued operating budget deficit of the United States (U.S.) challenges all Government funded programs to reduce cost while simultaneously meeting program objectives. For many programs, budgets have already been cut and leaders are figuring out ways to maintain organizational goals. This is the case with American military policy. The new strategic guidance for the Department of Defense (DoD) changed the requirement of having the ability to fight and win two simultaneous wars in two different regions [1]. Now the requirement is to fight and win one war, and "be capable of denying the objectives of – or imposing unacceptable costs on – an opportunistic aggressor in a second region" [2]. As President Obama presented this strategy at a Pentagon briefing in January 2012 he stated [3]:

> "our military will be leaner, but the world must know the United States is going to maintain our military superiority."

Maintaining military superiority with fewer forces is a challenging goal, but the Joint Chiefs of Staff (JCS) have taken action by developing plans for the military future coined "Joint Force 2020." Because all Services will have to decrease the number of active duty personnel, the military will rely on bringing together personnel from different Services to conduct joint operations. However, this is not a completely new course of action for the JCS [4]:

> "The strength of any Joint Force has always been the combining of unique Service capabilities into a coherent operational whole. Future Joint Forces will routinely employ more such combinations than ever before, with partners as well

as within the Joint Force, to achieve efficiencies and synergies not previously feasible. The assertion is that through globally integrated operations, Joint Forces will remain able to protect U.S. national interests despite constrained resources."

The JCS recognizes that it will take more than just a combining of personnel from varying Services in order for Joint Forces to be effective. "The ability to shift forces fluidly from one combatant command to another necessitates a certain amount of standardization between those theaters. Forces must train and exercise standardized tactics, techniques, and procedures in both joint and Service-specific training" [4]. Communication is also extremely important for Joint Forces, along with improving strategic and operational mobility, and enhancing tactical maneuvering. Training plays a major part in effectively integrating specialized skills and force structure to successfully execute tactical procedures and concepts of operations (CONOPS). Determining the most cost-effective set of alternatives to meet the current defense strategy requires careful analysis and planning.

More recently, the Pentagon presented a proposal to drastically reduce military personnel as part of the 2015 fiscal year (FY) budget[1] [5]. The proposal would shrink the U.S. Army by six percent to pre-World War II (WWII) levels, eliminate the Air Force fleet of A-10 close air support planes, and decrease military benefits. Defense Secretary Chuck Hagel explained that these "cuts are necessary to deal with the tight budgets and a changing battlefield" [6]. The proposed reforms for FY 2015 are an attempt to balance tradeoffs between pay and benefits versus training and modernization. Because of the financial constraints and reduction in funds, "[w]e chose further reductions in troop

---

[1] FY 2015 begins October 1, 2014.

strength and force structure in every military service - active and reserve - in order to sustain our readiness and technological superiority and to protect critical capabilities," said Hagel [5]. The operating budget is not sufficient to support the current number of military personnel and maintain superiority against the technological advancements of the adversary.

The DoD operates within the Defense Acquisition Policy, which established the Joint Capabilities Integration and Development System (JCIDS) process. The JCIDS process exists to support the Joint Requirements Oversight Council's (JROC) and the Chairman of the Joint Chiefs of Staff's (CJCS) responsibilities in identifying, assessing, validating, and prioritizing joint military capability requirements. Outputs of the JCIDS process are used to facilitate Doctrine, Organization, Training, Materiel, Leadership and Education, Personnel, Facilities, and Policy (DOTMLPF-P) changes, to drive the Defense Acquisition System (DAS), and to inform the Planning, Programming, Budgeting, and Execution (PPBE) processes [7]. Understanding how to meet DoD capabilities, in terms of training, is especially important considering the current defense strategy and development of Joint Force 2020. However, there is a lack of information on how to effectively analyze "Training" during the JCIDS process so that it can be compared with other DOTMLPF-P solution alternatives. Without the ability to compare "Training" with all other DOTMLPF-P solutions, capability based assessments (CBAs) are missing a potentially significant factor that could result in increased mission effectiveness and cost savings. This research attempts to enhance CBAs by developing and demonstrating the plausibility of a methodology to predict and evaluate the effectiveness of training (MPEET), which enables "Training" to be assessed as a DOTMLPF-P alternative.

The remainder of this chapter provides more details about the defense acquisition process and concludes with an overall research objective. Analyzing training effectiveness has been studied for decades and many models have been recommended. Several of these models will be surveyed in Chapter 2. Along with a description of the existing models of training effectiveness, the advantages and disadvantages of each method will be discussed. While the existing training effectiveness models are useful and have contributed to the current state-of-the-art in evaluating training, none meet all the criteria necessary to be used in a predictive analysis. Chapter 2 also examines the various types of information required to predict the effectiveness of a training system design. It includes a thorough review of learning theory, instructional strategies, and cost analysis methods. Chapter 3 discusses the model and simulation approach in the development of MPEET. Chapter 4 presents MPEET. Chapter 5 describes the case study used to demonstrate MPEET, and Chapter 6 provides a summary of this research effort, conclusions, and recommendations.

1.1 Defense Acquisition Process Overview

JCIDS is one of three key processes in the DoD Decision Support Systems (DSS), which must work together to ensure consistent decision-making and delivery of timely and cost effective capability solutions to war fighters [7]. In 2003, the defense industry acquisition guidance was completely reformed. Instead of the previous bottom-up requirements generation process, a top-down approach called JCIDS is now used. The intent was to replace statements such as “we need a more advanced fighter,” with “we need the capability to defeat enemy air defenses” [8]. JCIDS is responsible for developing capability requirements by identifying and prioritizing capability gaps and proposing solutions to fill those gaps. The JCIDS process is designed to include equal consideration of materiel as well as non-materiel solutions, but solutions are only recommended for JCIDS review after a CBA is complete. CBAs must therefore be conducted with the same regard for materiel and non-materiel solutions. The other two processes are the DAS process and PPBE process. The DAS process transforms validated capability requirements into materiel capability solutions. They are responsible for developing and/or buying the materiel solution [9]. This includes overseeing acquisitions from the materiel solution analysis through production and deployment phase of the program life cycle. The PPBE process enables funding for various JCIDS and DAS activities. They allocate resources and budgets based on the national security, defense and military strategies, and defined capability needs. Figure 1.1 shows the three DoD DSS processes interacting with the overseeing organization and official guiding documents. The DSS process has historically focused on the acquisition of materiel solutions. Due to the economic downturn, there is a shift to find cost-effective solutions that can be implemented more quickly and with less risk than traditional defense procurements.

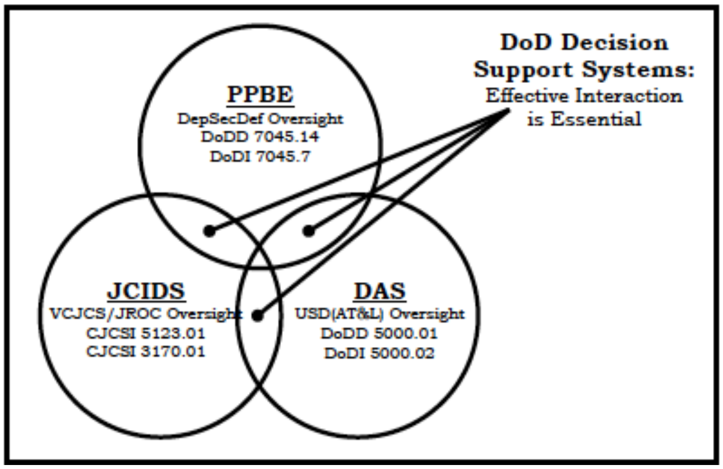


Figure 1.1: DoD Decision Support Systems [7]

1.2 Capability Requirements and Gaps

A simple definition for capability in JCIDS is the ability to achieve an effect in a military operation [8]. A capability gap refers to the inability to execute a specified course of action. The gap may be the result of no existing capability, lack of proficiency or sufficiency in an existing capability solution, or the need to replace an existing capability solution to prevent a future gap [10]. JCIDS is responsible for developing capability requirements by identifying and prioritizing capability gaps and proposing solutions to fill those gaps. Before any action can be taken in the JCIDS process related to reviewing and validating requirements documents, a capability requirement(s) must be identified related to functions, roles, missions, and operations of the Sponsor. They then must determine if there are any capability gaps which present an unacceptable level of risk and warrant further action in JCIDS [10]. Sponsor refers to the organization submitting a JCIDS document. Typical sponsors of JCIDS analysis are the Training and Doctrine Command in the Army, the Center for Naval Analysis and/or the Office of the Chief of Naval Operations staff in the Navy, the Marine Corps Combat Developments Command in the Marine Corps, and the operational commands (e.g., Air Combat Command or Air Mobility Command), supported by the Office of Aerospace Studies in the Air Force [11, 12]. Any of these organizations can bring forward a capability requirement to JCIDS.

The most common approach to defining capability requirements is through a CBA. "A CBA identifies the mission to be studied, the capabilities required to perform that mission, the operational characteristics and attributes of each capability, existing capability gaps and operational risks, an assessment of the viability of non-materiel solutions, and, if needed, a recommendation on the type of materiel solution to be pursued. A CBA also justifies that a solution is needed for the identified gaps, as opposed to accepting the operational risk and making no changes" [13]. CBAs are required to examine materiel and a set of defined non-materiel solutions. A materiel solution is a new item, such as a weapon system or aircraft, necessary to equip, operate, maintain, and support joint military activities. The predefined set of non-materiel solutions includes Doctrine, Organization, Training, Materiel, Leadership and Education, Personnel, Facilities, and Policy (DOTMLPF-P). Materiel in DOTMLPF-P are existing items, such as a weapon system or fighter jet, but rather than advocating for something new, the solution is to increase the current quantity or use the item in a different application [10]. The output of a CBA depends on the recommended solution. Figure 1.2 illustrates the flow of the JCIDS process and how CBAs fit in. If a non-materiel solution is recommended, a DOTMLPF-P Change Recommendation (DCR) is created. If a materiel solution is preferred, then an Initial Capabilities Document (ICD) is presented to the JROC for review and approval. Both of these documents initiate the process for further analysis and program development. Figure 1.3 summarizes the interaction between the capability requirements process and the defense acquisition process. After a decision is made to move forward with a new acquisition request in lieu of a DOTMLPF-P change request during the JCIDS process, the capability requirement is fulfilled through the program acquisition process. The capability requirement is periodically reviewed throughout the product life cycle to ensure proper alignment with changes in knowledge, circumstances, budgets, and requirements [14].

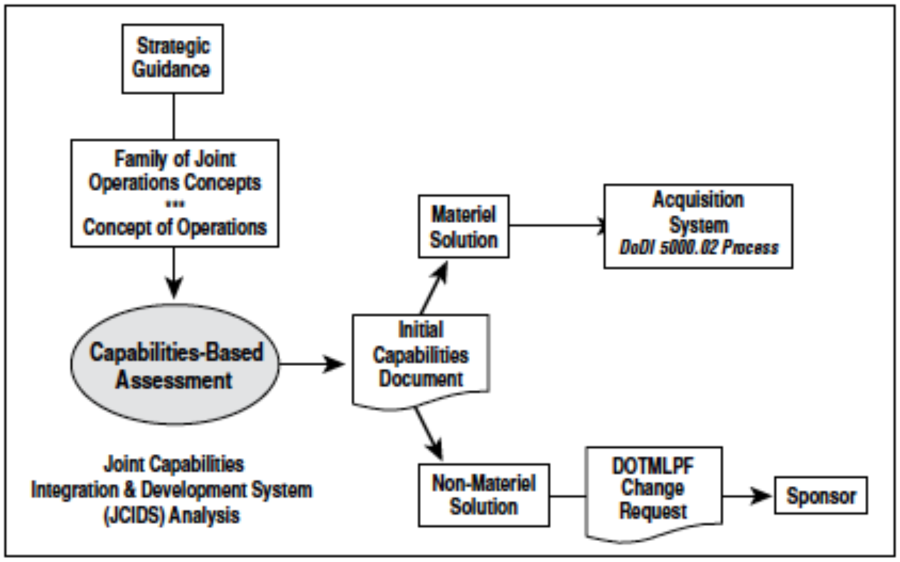


Figure 1.2: Capabilities-Based Assessment Process [12]

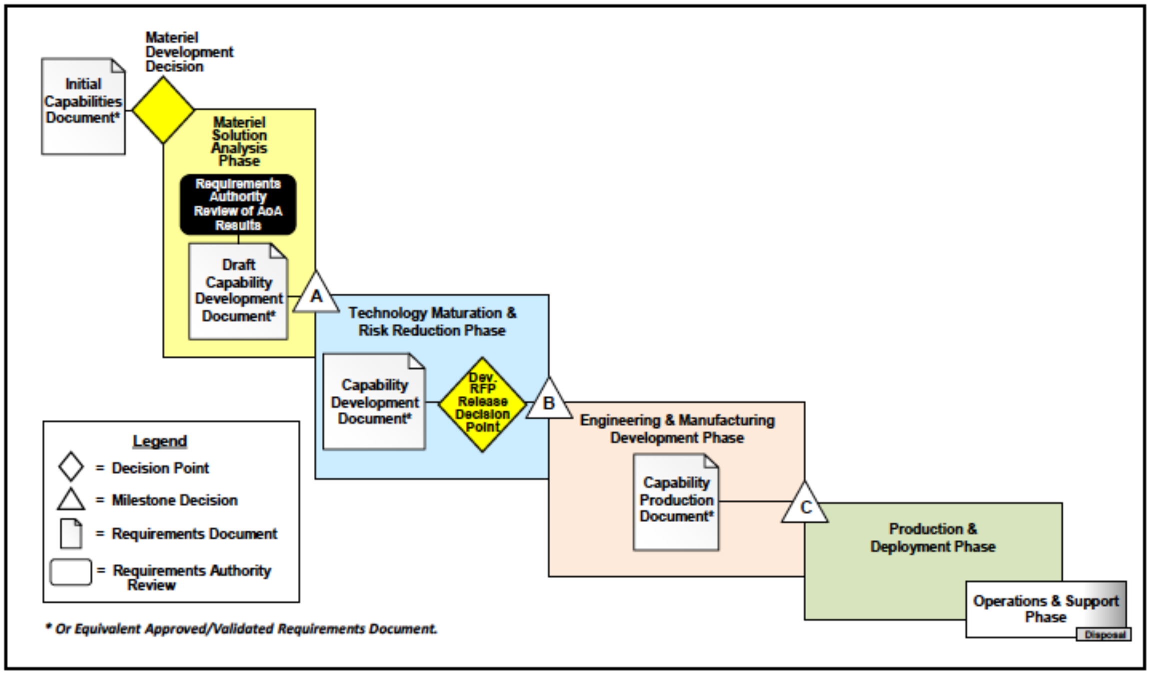


Figure 1.3: Defense Acquisition Decision Process [14]

1.3 Challenges in Analyzing Training as a DOTMLPF-P Alternative

Capability Based Assessments (CBAs) are performed well before any solution is designed or developed, as illustrated in Figures 1.2 and 1.3. Conducting any analysis this early creates significant challenges in CBA trade studies because the evaluations are theoretical and based on existing designs or specification data. What is expected from a solution versus its final capabilities can vary based on many factors including, but not limited to, development times, costs, budgets, requirement changes, and stakeholder inputs. As part of a CBA one must determine to what extent capabilities are provided now and the current plan for how they will be provided in the future [8]. This involves being able to objectively measure the current capability and predict future performance.

Information on how to perform a CBA is provided in the following documents: *Appendix A to Enclosure B of the Manual for the Operation of the JCIDS, Capabilities-Based Assessment (CBA) User's Guide* by JCS J-8, and *Capabilities-Based Assessment (CBA) Guide, Version 3.1* by The U.S. Army Training and Doctrine Command (TRADOC) [15-17]. *Appendix A to Enclosure B of the Manual for the Operation of the JCIDS* is the official CBA guide provided by the JCIDS authors. It includes a summary of the CBA process steps and references for detailed guidance and best practices [15]. The suggested references that address DOTMLPF-P solution alternatives are the above-mentioned CBA User Guides by JCS J-8 and TRADOC. All three of these documents emphasize the requirement to determine if a non-materiel approach can partially or entirely mitigate any identified capability gaps by recommending changes to existing capabilities in one or more of the DOTMLPF-P areas. Definitions of each alternative are given, but details and recommendations of what to include or where to find information to properly assess each area is only provided for certain solution alternatives. In the *Appendix A to Enclosure B of the Manual for the Operation of the JCIDS* details are discussed for what they refer to as “the most common non-materiel approaches” [15]. The most common non-materiel approaches are identified as alternative doctrine and CONOPS, policy alternatives, organizational alternatives, and personnel alternatives. The *CBA User’s Guide* authored by JCS J-8 offers practical advice on how to conduct a CBA. It advises one on how to assemble a CBA that meets the goals of JCIDS, captures lessons learned from previous CBAs, and discusses the techniques and practices that have led to successful CBA completions in the past [16]. Details on DOTMLPF-P alternatives are limited to the same non-materiel approaches from the JCIDS manual. The only document that lists examples of what an analyst should consider when evaluating the training alternative during a CBA analysis is the *CBA Guide, Version 3.1* authored by TRADOC. They suggest that the analyst consider several questions such as: “Is existing training being delivered effectively? Are training results being monitored and analyzed for effectiveness? Is training properly staffed and/or funded? Are there training devices, simulators, or simulations that, if developed and fielded, would close or mitigate the gap?” [17]. These questions provide an analyst, who may or may not have a background in training, a place to begin doing research and seek subject matter expert (SME) input. However, TRADOC does not provide any references or information on how to answer these questions. To date documents that provide assistance for conducting CBAs have yet to address how to include training in the DOTMLPF-P analysis process.

Investigation of how to effectively analyze training revealed that the DoD has been heavily criticized for poor cost-effective analysis of training (CEAT). A 1995 study concluded that CEAT methods are inadequately defined, DoD policy guidance for CEAT is ambiguous, CEAT procedural guidance is inadequate, and CEAT programs differ among the Services [18]. Surprisingly, the same study found that the cost analysis part of CEAT is fairly well defined but methods for performing the related training effectiveness analysis are not. Methods varied between the Services for how to evaluate training, and within some Services there was inconsistency in training evaluation methods. At the time of that study there were several military standards and handbooks available for instructions on analyzing and evaluating training programs [19, 20]. The Navy primarily developed these with participation from all other military branches. However, other Services, such as the Air Force, have written their own handbooks for specific training activities [21]. In 1999, MIL-HDBK-29612 was released. It is a five part handbook providing guidance on all facets of training, from identifying training requirements, solution analysis and approaches to training and training evaluation [22]. Although, contractors and government acquisition heavily use this handbook series, handbooks are not requirements and the Services still have their own instructional documents. More recently, a 2009 U.S. Government Accountability Office (GAO) report found the Navy and Air Force were not fully applying best practices in the development and management of combat skills training [23]. The Navy's Expeditionary Combat Skills (ECS) course was examined based on the intent to standardize the training curriculum by eliminating inefficiencies and wide divergences in existing combat skills training. The Air Force planned to provide similar training as part of their Common Battlefield Airmen Training (CBAT) program. After reviewing the Air Force program, it was discovered that the Air Force "did not tie the need for the expansion of CBAT training to an identified gap in combat skills training, knowledge, and abilities." This contributed to the eventual cancellation of the training program. The Navy had clear goals to provide training to all forces that lacked entry-level individual combat skills and to establish a training pipeline for all expeditionary troops. Even though they had a clear implementation strategy and were in the process of training, they were criticized for not creating or operating on a timeline to complete the combat skills training for all forces. Another GAO report in 2011 discovered actions that are needed to improve training integration and training cost estimations of the Ballistic Missile Defense System (BMDS) [24]. Operation of the BMDS involves the Missile Defense Agency (MDA) and multiple Services. GAO found gaps between training requirements and budget resources where MDA and Services had not completed training cost estimates before fielding BMDS elements. DoD is attempting to follow "train the way you operate" Joint Staff instructions, but will fall short if training goals and costs are not aligned.

These varying GAO reports show the inconsistency with training evaluation and budget allocation within the DoD. Without an understanding of how effective training is in terms of performance measures and cost, there is no way someone could reasonably investigate the ability of training to achieve an effect in a military operation, which is the JCIDS definition of a capability. Without a means for analyzing training in terms of capability, training will not be properly assessed as part of the DOTMLPF-P non-materiel solution alternatives. It is extremely difficult to include training as an alternative in a CBA if the mission goals, training deficiencies, and effectiveness of training are not understood.

1.4 Primary Research Objective and Research Questions

Two major observations are evident from literature regarding CBAs and their approach to analyzing DOTMLPF-P alternatives. The first observation is that there is minimal information in the referenced JCIDS documentation and guidelines provided to an analyst who wants to include “Training” as an alternative solution. Currently, most of the CBA guidance treats the training alternative as an afterthought or a subsidiary requirement for materiel solutions. Secondly, there are numerous methods recommended and used for evaluating training effectiveness within the DoD, but all the methods provided in Government standards and guidebooks are post-training evaluation techniques. This likely contributes greatly to the lack of including training as part of CBAs. To determine the capability gaps that training can fill, in comparison with alternate solutions during early phase defense acquisition decision-making, there must be a method that can predict training effectiveness, as opposed to post-training evaluation procedures; and the results must be expressed in comparable terms of alternate solutions. This research addresses the second observation. Once the effectiveness of training can be predicted, and not just evaluated, then guidance can be developed for properly including training as part of the non-materiel solution alternatives.

The goal of this research is to develop a methodology for evaluating the cost and effectiveness of training, and to demonstrate how this new ability can enable the inclusion of training as an independent DOTMLPF-P alternative.

In order to meet this research objective several questions must be answered as summarized below.

RQ1. What is an appropriate method of measuring training effectiveness during early phase defense acquisition to aid decision makers in DOTMLPF-P alternative selection?

RQ2.1 How does one quantify the benefits of soldiers training in terms of effectiveness?

RQ2.2 For a given set of monetary resources, how should one allocate resources to gain maximum training effectiveness?

RQ3: How does one quantify increased knowledge, skills and attitudes in training system design?

The first research question, RQ1, "what is an appropriate method of measuring training effectiveness during early phase defense acquisition to aid decision makers in DOTMLPF-P alternative selection," stems directly from the primary research objective. An investigation into the existing methods for evaluating training effectiveness, detailed in Chapter 2, reveals that training evaluations are commonly performed upon completion of a training program. This tradition is not sufficient for CBA analysis, or for any organization interested in estimating the potential return on investment in advance of executing a training program. It will be shown that a process for predicting training effectiveness is required. With any predictive capability there exist uncertainty between predicted and actual performance, which is accounted for in this research effort. Historical training evaluation methods use post-training surveys and follow up on-the-job (OJT) assessments to determine training effectiveness. For CBA analyses, indicators of training effectiveness are necessary pre-training. RQ2.1 and RQ3 were derived from the need to understand the contributing factors in training system design that are available for pre-training effectiveness predictions. To answer RQ2.1, "how does one quantify the benefits of soldiers training in terms of effectiveness," and RQ3, "how does on quantify increased, knowledge, skills, and attitudes in training system design" required research in learning theory and instructional design. Discussed in Chapter 2 are taxonomies for various learning domains, methods and types of media to use for meeting instructional objectives, and analysis methods for determining the criticality of training lessons. An emphasis is placed on measuring training effectiveness beyond evaluating if and how well a soldier can perform a task. Training that includes emotional control and cognitive problem solving ability is just as important as physical and technical skills training. It is understood that not every possible scenario can be experienced as part of training, but soldiers are expected to perform in real life based on their training. There has to be a measurement technique within instructional design that assures the design team that the training system developed adequately prepares trainees for both nominal and off-nominal conditions and enhances the probability of student learning. RQ2.2, "for a given set of monetary resources, how should one allocate resources to gain maximum training effectiveness," is necessary because of the current economic situation of the DoD. The DoD is operating with constrained resources, which is one of the primary reasons for the shift to capability-based requirements [9]. The philosophy of maximizing capability and minimizing cost is essential. Maximizing training effectiveness must be done with an understanding of the associated costs. Cost analysis and decision-making techniques are both investigated to create a balance between the effectiveness and cost of training. Weightings are applied to criteria used to evaluate the pre-training variables based on decision-maker preferences for training costs and effectiveness. This permits objectivity in determining the effectiveness of training. Answering these research questions provides information used to develop a methodology for evaluating the cost and effectiveness of training, and to demonstrate how this new ability can enable the inclusion of training as an independent DOTMLPF-P alternative.

CHAPTER 2

BACKGROUND

The effectiveness of training is measured by any organization that values its investments. Investments could be made in terms of time, people, monetary or any combination. People are trained with the expectation that there will be a return on investment for the person or organization sponsoring the training. How this return on investment is measured can depend on multiple factors, but one of the most important is always the cost effectiveness or cost benefit analysis. In the past half-century, many models have been published to address the need to quantify training effectiveness. These methods and models were reviewed in order to answer RQ1, “what is an appropriate method of measuring training effectiveness during early phase defense acquisition to aid decision-makers in DOTMLPF-P alternative selection?” The methods surveyed in this section have been used for military applications and provide a basis for the training effectiveness method proposed herein.

Defining Training

Ask anyone to define training and you are sure to get an answer, but what is the best definition of training? It is necessary to define training and training effectiveness before proceeding to review training effectiveness models. Oxford Press University provides a general definition of training as to teach (a person or animal) a particular skill or type of behavior through sustained practice and instruction [25]. Dictionary.com provides a similar definition but without reference to a subject, “to give the discipline and instruction, drill, practice, etc., designed to impart proficiency or efficiency” [26]. Both definitions contain two distinct points concerning training. First, a specific result is expected, such as a certain skill, behavior or proficiency level. Second, this change in ability is a direct result of instruction and practice over a period of time. Moving from general definitions to military applications, two other definitions are useful for understanding training. The Joint Capability Areas (JCA) under the Joint Chiefs of Staff (JCS) define training as [27]:

> The ability to enhance the capacity to perform specific functions and tasks using institutional, operational, or self-development (to include distance learning) domains in order to improve the individual or collective performance of personnel, units, forces, and staffs.

The definition of training according to Defense Acquisition University (DAU) Glossary is [28]:

> The level of learning required to adequately perform the responsibilities designated to the function and accomplish the mission assigned to the system.

The DAU and JCA training definitions contain the same points as the general definitions, an expectation of a specific outcome based on practice, but they also include a beneficiary from training. JCA states that training can benefit individual or group performance. Recognizing that training benefits individuals as well as groups is important in a military context because mission performance is dependent upon teamwork. Individuals by themselves do not win the war, but the contributions of each individual are necessary to meet the mission objectives. The benefit of training according to the DAU is mission accomplishment. Defining training for overall mission accomplishment is most applicable to the primary research objective of developing a methodology for modeling the effects of warfighter training, because it allows flexibility to include all types of skill acquisitions and links learning to successfully completing the mission. Both the DAU and JCA definitions will be used within the context of this research. This provides not only the definition of training, but also how it occurs, who is effected, and why it is necessary from a military prospective. The JCA training definition will be considered a sub-level definition to the DAU training definition. Where the DAU describes training from a very high, system level, top-down viewpoint, JCA provides insight from a bottom-up approach by actually specifying types of training and whom it directly affects.

Training effectiveness is the study of the individual, training, and organizational characteristics that influence the training process before, during, and after training [29]. The focus of this research is studying the characteristics that influence training outcomes before training occurs; however, the literature on training evaluations is dominated by post-training evaluations. Training evaluations are generally defined as a measurement technique that examines the extent to which training programs meet their intended goals. In a training effectiveness study evaluation of something is still required. A literature search has been conducted and presented in this thesis to discover the best variables to use during a pre-training evaluation to represent effectiveness of the training system.

2.1 Existing Models of Training Effectiveness

Training of people has existed since the beginning of human kind. The simple, or not so simple, act of rearing children from babies to young adulthood is filled with training exercises. As children transition to working adults, they become employed with some basic knowledge and skill level. As employees, people are trained to enhance their job performance. On-the-job training is provided with the expectation that upon training completion the employee will be capable of performing their job at a certain proficiency level. The same is true about training military personnel. In order for the military to properly invest in training, it is necessary to quantify the effectiveness of training. Several models have been published that attempt to quantify training effectiveness. A summary of models that have been applied to military applications are listed in Table 2.1, followed by a detailed discussion of each model.

Table 2.1: Training Effectiveness Models Summary

Author	Year	Model Summary	Development Context
Kirkpatrick	1959	1) Reaction 2) Learning 3) Behavior 4) Results	General technique
Deitchman	1988	1) Quantify training needed to maximize performance 2) Determine realistic performance	Assessment of the military value of unit training
Bell & Waag	1998	1) Utility Evaluation 2) Performance Improvements 3) Transfer to Alternate Simulation Environment 4) Transfer to Flight Environment 5) Extrapolation to Combat Environment	Evaluating the effectiveness of flight simulators for training combat skills
Bahlis & Tourville	2005	1) Discovery 2) Strategize 3) Prioritize 4) Optimize	Training impact assessment during upfront planning phase
Schrieber, Schroeder & Bennett	2011	1) Performance Improvements 2) SME ratings 3) Reaction	Evaluating the effectiveness of DMO simulator training (F-16)
Clark	2012	1) Results 2) Performance 3) Learning 4) Motivation	General Technique

2.1.1 Kirkpatrick Four Level Model

In 1959, Kirkpatrick published a four level model for evaluating training effectiveness. He has updated his publications, as recently as 2006, with detailed case studies and current examples, but the model has remained the same. The four levels of Kirkpatrick's model include 1) Reaction, 2) Learning, 3) Behavior and 4) Results [30] as shown in Figure 2.1. Level one, reaction, asks how the trainee reacted to the training session. Did they like it, and do they see immediate application to their job [31]? Determining the employee's reaction is typically done using a post-training evaluation filled out by the trainee. It may include questions with ratings and/or open-ended questions and comments. Level two, learning, estimates how much the trainee learned in comparison to the specific learning objectives for the training session. This is evaluated by exams or practice sessions at the culmination of a training activity. Levels three and four focus on how much the training activity has an impact outside of the training environment. Level three, behavior, measures actual changes in behavior on the job for tasks that specifically relate to the training objectives. For example, does the trainee use the techniques and skills taught in the training program or some other methods? Level three is assessing how much training transferred to the work environment. Level four, results, measures the impact that training has to the organizations' bottom line. Examples of measurable results include time to complete a task or reduced number of errors, if efficiency was the objective. Efficiency can be translated directly to financial value. Results do not have to be a financial measurement. Improved morale and reduced personnel turnover are examples of measurable results that are not as easy to quantify with cost. Whichever factors are used to measure results, they should be selected to correspond with the original training objectives. Otherwise the training effectiveness assessment will be inaccurate.

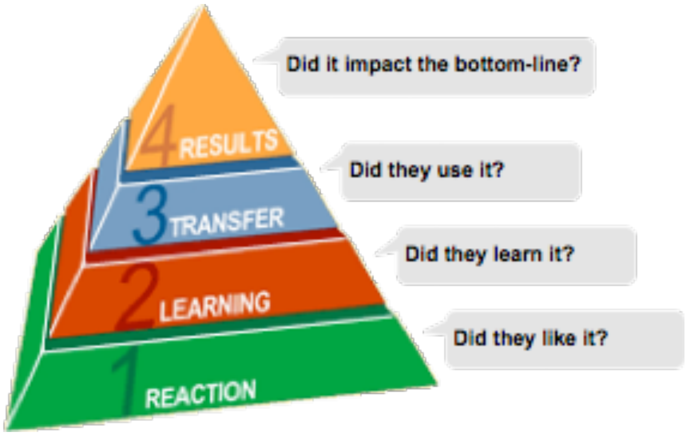


Figure 2.1: Kirkpatrick Training Evaluation Model [32, 33]

The “Kirkpatrick Model” is the most commonly used method for evaluating training effectiveness [34-36]. It is used throughout government, corporate, and academic institutions [37-41]. It has proven to be a successful model for evaluating the effectiveness of training after a training activity has completed, and aids in design and preparation of training materials and methods. According to the ASTD in 2010, over 90% of companies surveyed measured trainee reactions, over 80% measured trainee learning, over 50% measured on-the-job behavior, and nearly 40% reported measuring results [42]. Kirkpatrick encourages using return on expectations (ROE) in place of standard return on investments (ROI) as the general training effectiveness metric [39]. Using ROE versus ROI as the value indicator is not useful for building a military system-of-systems (SoS) training effectiveness model because the decision makers need to know the estimated costs to satisfy the mission objectives from investing in training compared to other

DOTMLPF-P alternatives. Kirkpatrick's idea of negotiating and compromising on expectations fails when the results being measured are defense of our country, allies, and protection of our troops. The concept that results have to be measurable in terms of meeting the training objectives is useful in developing MPEET for modeling the effects of soldiers' training and demonstrating how this modeling technique can be used in trade studies with other DOTMLPF-P alternatives when evaluating military SoS performance. The difference is that the results must be come from a predictive model rather than a post-training evaluation.

2.1.2 Clark Model

Although no other model is so widely used, Kirkpatrick's training effectiveness model has received praise and criticism over the decades. Donald Clark believes that the Kirkpatrick model includes the necessary elements, but it is presented incorrectly [43]. Clark makes two primary changes to Kirkpatrick's model as shown in Figure 2.2. First, he reverses the order of all four levels. Where Kirkpatrick's final step is measuring results against business objectives, Clark makes this step one. Actual results are the primary interests to business leaders and decision makers. It is satisfying to know that employees are motivated and interested in their training activities but enhancing performance is the reason for investing in training and it weighs significantly more than the employee's reaction. Many programs spend so much time in levels one and two of Kirkpatrick's method that according to an American Society for Training and Development (ASTD) Value of Evaluation Study in 2009, very few organizations performed the level four evaluation [44]. Level four is the most important, yet least executed. Along with Clark, many researchers, including Kirkpatrick's son James (part of

Kirkpatrick Partners, LLC), argue that reversing the order of Kirkpatrick's model by presenting results first allows the model to be used as a training-preplanning tool, in addition to an evaluation tool [43-47].

The other modification Clark made to Kirkpatrick's model is to rename step one from "Reaction" to "Motivation", and step three from "Behavior" to "Performance". Training should be conducted as a result of some identified performance or capability gap, therefore Clark and Wick et al. recommend that the learner be made aware of the fact there is a gap and the evaluation then focuses on the learners motivation to close the identified gap [43, 48]. Reaction, according to Clark, can result in trainers developing fancy graphics and humorous games, which may or may not have an effect on the trainee's response to the training session. Performance and behavior are similar, however, as Rudman states, "performance is focused behavior or purposeful work" [49]. Performance is results driven, which is the intent of training and evaluations [50]. These name changes to steps one and three are commonly used in the business / human resource industry and often appear as the original Kirkpatrick model [51].

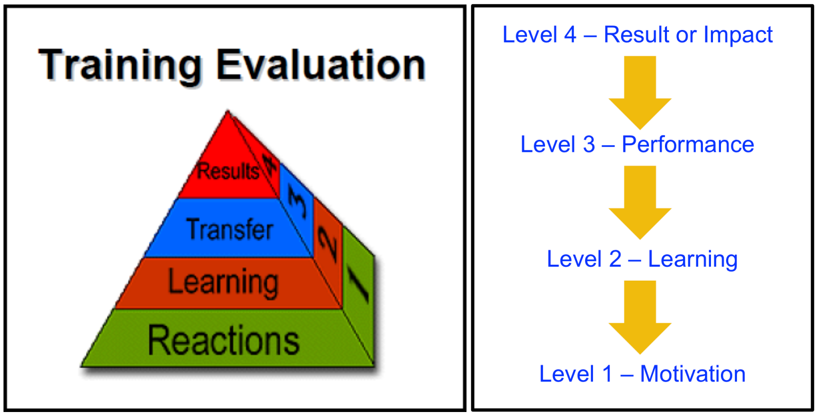
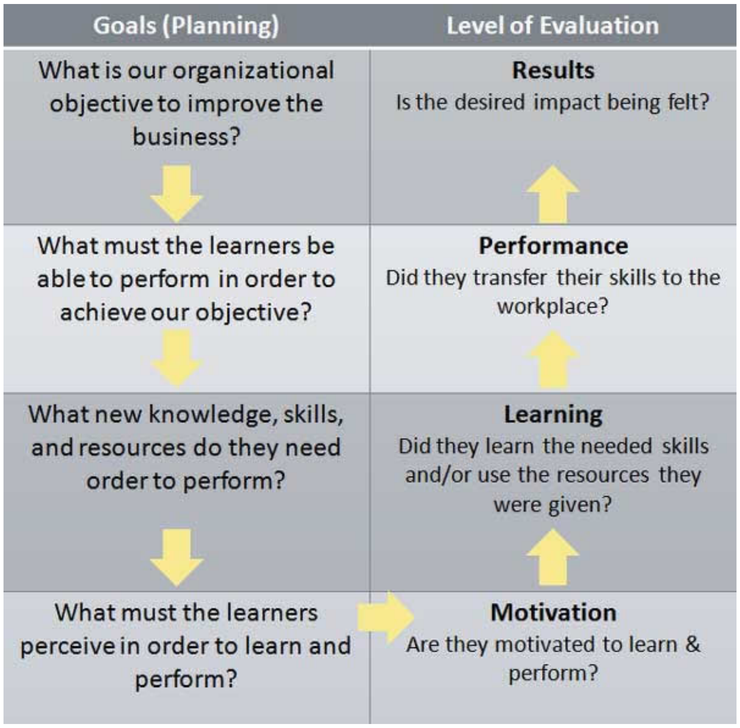


Figure 2.2: Kirkpatrick Training Evaluation Models: Original (Left) [52] and Modified (Right) [45]



Goals (Planning)	Level of Evaluation
What is our organizational objective to improve the business?	**Results** Is the desired impact being felt?
What must the learners be able to perform in order to achieve our objective?	**Performance** Did they transfer their skills to the workplace?
What new knowledge, skills, and resources do they need order to perform?	**Learning** Did they learn the needed skills and/or use the resources they were given?
What must the learners perceive in order to learn and perform?	**Motivation** Are they motivated to learn & perform?

Figure 2.3: Clark Training Effectiveness Model - Top Down Approach [46]

Don Clark's modified Kirkpatrick Four Level Training Effectiveness Model is useful to the proposed research because it provides a top-down approach to training, as shown in Figure 2.3. The proposed research focuses on providing decision makers with level four information (predicted results) they can use to decide if training is the best DOTMLPF-P solution when weighing costs versus benefits. Kirkpatrick and Clark contributed training evaluation models that can be generally be applied to any problem. They provide a starting point for military training evaluations, training prediction models and cost analyses.

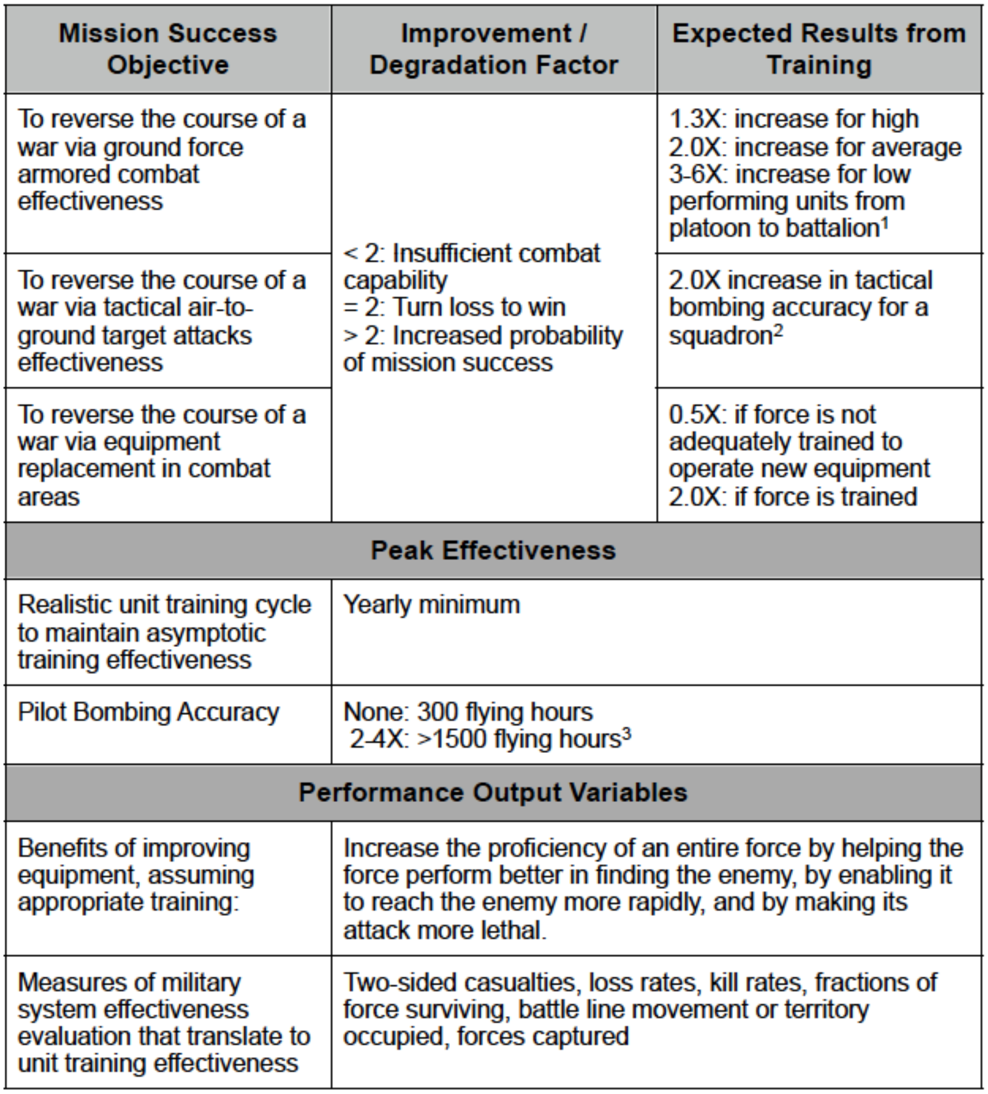
2.1.3 Deitchman Model

Seymour J. Deitchman conducted a study in 1988 under the Institute for Defense Analyses to "assess the military value of unit training in the same quantitative cost and effectiveness terms used to assess investments in other areas such as acquisition of new weapon systems or of more forces of various kinds" [53]. To complete this analysis Deitchman experimented with a large-scale Tactical Warfare computer-based simulation model (TACWAR) which represented a war between the North Atlantic Treaty Organization (NATO) and Warsaw Pact (WP) in the central region of Europe. Model parameters were changed to reflect weapon system performance based on user proficiency, with the assumption that increased training results in higher proficiency. For example, it could be assumed that target identification rate or bombing accuracy can be doubled through training. Deitchman increased the proficiency of controlled parameters until the outcome of a conventional conflict was reversed. The outputs of the model were initially reviewed for reasonableness. Deitchman then solicited military data and expert judgments "that would indicate the nature and extent of training necessary to achieve effects such as those that emerged from the 'test'." Once Deitchman had evidence that increased training could provide the required improvements in proficiency, he estimated the type of training and costs associated with changing the outcome of war in TACWAR. At the conclusion of this study, Deitchman was able to give coarse approximations or "rules of thumb" regarding attribute factors and cost comparisons that future analyst can use to assess the value of unit training versus force and hardware investments. For more details on Deitchman's modeling approach, the reader is referred to *Preliminary Exploration of the Use of a Warfare Simulation Model to Examine the Military Value of Training* by Seymour J. Deitchman [53].

Deitchman showed that the military value of unit training can be quantified in measures similar to that of investing in force structure and new or upgraded equipment, and he demonstrates how it can be done. First order cost approximations were provided for ground combat and tactical air-to-ground training investments to reverse a losing conflict in the TACWAR war model. These are the types of results that Kirkpatrick level four evaluations seek, but Deitchman used current and past data to baseline the TACWAR model so that his results can be used for future planning. This is a necessary modeling practice that is also proposed as part of this research and experimental plan. Conclusions drawn by Deitchman are summarized in Table 2.2, for attribute ranges and variability during development of the training effectiveness predictability model and testing.

Two additional points should be noted from Deitchman's observations upon completion of his work. The results showed that training yields quantifiable improvements in performance in the areas of warfare examined (ground combat and tactical air-to-ground target attacks), but the benefits gained from training were not enough or cost the same as the cost for new equipment. Most of Deitchman's data came from field training exercises and simulation trainings of tanks (M1 and M-60) and aircraft (A-7, A-10, F-16, F/A-18). Deitchman stated that "if more training could be done as simulation exercises the cost benefit of training would likely be much better" [54]. Deitchman conducted his study over twenty years ago, when simulation training was not as widely practiced. Advancements in technology now permit distributed network simulation, where not only do pilots train in simulators, but also all roles including mission support can participate in the training exercises. Simulated training costs less, and allows for units to gain more experience in mission based scenarios than traditional field exercises. This will be discussed in more detail in the Schreiber et al. training effectiveness model. The other observation from Deitchman's results, "obvious when it is stated but not so obvious a priori, is that training is needed with new as well as with existing equipment, and this changes the way the equipment [versus] training trade-off question must be formulated" [54]. The issue is not whether funds should be put into improvement in training *or* equipment, because both contribute to force improvement, and both are needed. The proper way to view the training versus equipment trade-off during early phase planning is to separate it into parts. First, find out how much training is needed to maximize performance with either current or new equipment, then decide at what point training has carried the force as far as it can go, so that equipment and force size change will be necessary to carry it further. Following this method, resources can be allocated to training to make the best of the existing forces, then funds can be allocated to improve the forces' equipment and/or to change their size. This is opposite to how funds have been historically allocated where equipment improvement occurs on a regular renewal cycle, changes in force size are driven by external events, and funds are allocated to training from any remaining residual. Deitchman's thoughts regarding how the process for allocating training funds occurs is considered in the proposed training effectiveness model methodology. It is especially noted that Deitchman degrades performance metrics by half when new equipment is added without adequate training.

Table 2.2: Deitchman Conclusions Summarized [53-56]



Mission Success Objective	**Improvement / Degradation Factor**	**Expected Results from Training**
To reverse the course of a war via ground force armored combat effectiveness	< 2: Insufficient combat capability = 2: Turn loss to win > 2: Increased probability of mission success	1.3X: increase for high 2.0X: increase for average 3-6X: increase for low performing units from platoon to battalion[1]
To reverse the course of a war via tactical air-to-ground target attacks effectiveness		2.0X increase in tactical bombing accuracy for a squadron[2]
To reverse the course of a war via equipment replacement in combat areas		0.5X: if force is not adequately trained to operate new equipment 2.0X: if force is trained
Peak Effectiveness		
Realistic unit training cycle to maintain asymptotic training effectiveness	Yearly minimum	
Pilot Bombing Accuracy	None: 300 flying hours 2-4X: >1500 flying hours[3]	
Performance Output Variables		
Benefits of improving equipment, assuming appropriate training:	Increase the proficiency of an entire force by helping the force perform better in finding the enemy, by enabling it to reach the enemy more rapidly, and by making its attack more lethal.	
Measures of military system effectiveness evaluation that translate to unit training effectiveness	Two-sided casualties, loss rates, kill rates, fractions of force surviving, battle line movement or territory occupied, forces captured	

1. These factors were provided by Gen. Gorman concerning realistic training above the capability that routine peacetime training and other military activity generates [55].
2. Bombing accuracy for a squadron increases at the cost of increasing tactical bombing practice flying hours from 10 to 40 hours per month. Improvements in survivability, through Red Flag Exercises, are also needed to ensure the maximum contribution of tactical air-to-ground training to winning a war.
3. Data from Air Force Study [56]: 300 flying hours represent a pilot just beginning their career specialty; pilots reached a highly experienced performance level around 1500 hours. Experienced pilots who's immediate practice increased from 10 to 40 hours a month showed four times the improvement in bombing accuracy compared to beginners.

2.1.4 Bell and Waag Model

From the late 1970s throughout the 1990s the popularity of simulator-based training grew. Resources were allocated and efforts were continually put forth to develop technologies for training combat skills in flight simulators. In 1998, Bell and Waag, did a research review of approximately 25 total air-to-surface and air-to-air combat simulator training evaluations [57]. They discovered that the majority of claims, which stated that simulation training was valuable, were based upon trainee opinion data or subjective rankings, with very little objective evidence. They also observed that most evaluations of military training systems closely paralleled Kirkpatrick's four criterions. Bell and Waag proposed a five-stage sequential process training evaluation model, shown in Table 2.3, which they believed "would provide an estimate of the military value of combat training using simulation." The last step, stage five, consists of developing an analytical model by extrapolating data collected in an alternative simulator and actual flight environments to weigh tradeoffs between weapons systems enhancements, increased flying hours, and advanced simulation-based training. Bell and Waag considered Deitchman's approach of initially using arbitrary estimates to represent the potential impacts of training, but decided that using the data collected from stages one through four will result in a more exact training value estimate. The acknowledged fallacy in their method is the difficulty in gathering the data and magnitude of resources, labor, and cost that is needed from the first four stages to build the analytical model. It is agreed that the most accurate training value estimate is always wanted, but there must also exist a balance between accuracy, uncertainty, and the cost of being precise, especially when results will come from analytic tools. Ultimately, the most accurate training value estimate comes from actually performing the training and conducting post-training evaluations. When using analytical tools to estimate training value, the results are limited in accuracy by how relevant the model inputs are and the sensitivity of each variable. For any input that has a significant impact to the resultant output, it may be worth considering an investment in gathering that information as accurately as possible. Accuracy of input variables in an analytical model can also be limited by the amount of time available to develop a value estimate and the availability of resources.

Table 2.3: Bell and Waag Five Stage Training Evaluation Model

Stage Number & Title	Description
1 - Utility Evaluation	Evaluate the accuracy or fidelity of the simulation environment, and gather opinions about the potential value of the simulation within a training environment
2 - Performance Improvement	Determine the extent to which simulator based training improved performance within the simulation environment.
3 - Transfer to Alternate Simulation Environment	Test out improvements in another simulated environment that is built like a real mission.
4 - Transfer to Flight Environment	Due to peace time restrictions only a subset of data may be collected when transferring to an actual live flight environment. Compare the performance of simulator trainees versus those untrained.
5 - Extrapolation to Combat Environment	Use actual data generated from previous stages as inputs to an analytic model of a mission scenario to determine military value of simulator training.

Bell and Waag took Kirkpatrick's training evaluation model and modified it to have a more engineering-based approach. In stage one, utility evaluation, they solicit weightings for parameters that are affected by training. These weightings are used in stage five for the analytical model. The concept of weighting parameters is common in utility theory and is part of the methodology proposed herein to model the effects of soldiers' training. The idea that empirical data is desirable, but not necessary to predict the value of simulator training can be expanded to general training. A model can always be updated and revised when better data becomes available. If reasonable data exists, then using a model to aid in determining the effectiveness and the value of military training, similar to Deitchman's method of varying parameter ranges and including parameter weightings from Bell and Waag, is a valid approach.

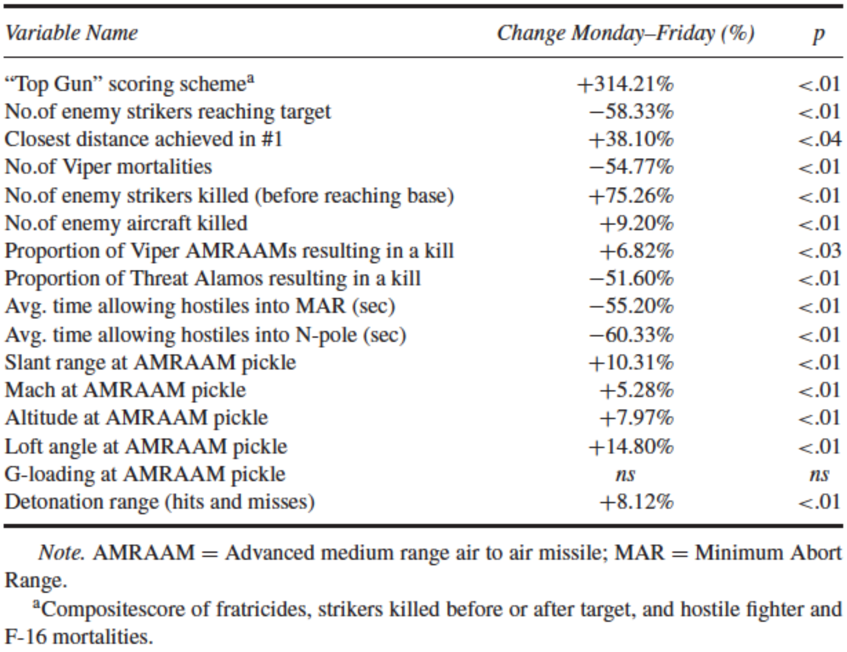
2.1.5 Schreiber, Schroeder & Bennet Model

It has been established that by using analytical modeling techniques one can estimate the cost of training effectiveness with a level of uncertainty [56, 57]. The question remains what data is considered "reasonable enough" that the model will be credible, and after verification and validation, deemed acceptable. This is why Deitchman used as much real field exercise data as he could find, in addition to simulator training data. This is also why Bell and Waag proposed gathering pre and post-training results from multiple environments (simulator, alternate simulator, actual flight practice exercise, real mission). Recognizing the need for credible data, the U.S. Air Force (USAF) in 2002 funded a F-16 squadron study to validate the cost effectiveness of Distributed Mission Operations (DMO) training within a virtual environment [58]. DMO training is generally defined as events that can bring multiple war fighters together to train for complex individual or team tasks during large scale, realistic combat missions. Prior to the late 1990s, complex tactical mission training was provided infrequently during sizable “live” range exercises. Surveyed combat pilots reported that simulated higher order training, like DMO network training, allowed them to gain battle-like experiences not frequently encountered outside of real war [59]. This F-16 pilot study was led by Scientists Schreiber, Schroeder, and Bennett who, aware that limitations exists on how DMO within-simulator training results extend to the real world mission, chose to include multiple types of assessment to make their findings more robust. They performed this research by adopting a training effectiveness methodology that resembles more of Bell and Waag's model than Kirkpatrick's; nonetheless, it includes objective, subjective, and user opinion data. A total of 76 teams (384 pilots and Airborne Warning and Control System operators) participated in a one week (Monday - Friday) DMO training while objective data was collected by measuring improvements on outcomes and skill proficiency [58]. Summary results for the objective metrics tracked are shown in Table 2.4. Subjective data was collected by trainer expert observation data. These results favorably compared with the objective data, where the trainer subject matter expert (SME) rated trainee's competency higher at the end of the week long training compared to their initial evaluation. User opinion data was captured based on questionnaires and out-brief sessions, which captured the trainees opinions about the usefulness of the training system as well as pros and cons.

Since the 2011 publications, Schreiber et al. recommend that future work include live-fly sessions where data can be measured to determine how much training transfers from the DMO simulated environment to the real world. That effort had not been funded at the time of communications [B. T. Schreiber (personal communication, March 15, 2013)]. The authors are supporting research initiatives where data from the F-16 DMO experiment is being used to build models that will predict the effectiveness of DMO training rather than only conducting post-training evaluations [58]. Schreiber et al. also recommend that the metric data be used as maximum expected improvement due to issues such as training decay, inclement weather during missions, and potentially negative training issues (e.g. lack of performing emergency procedures, g-force effects not simulated and lack of consequences for running out of fuel) [58]. Considering the potential negative training issues and concerns, the metrics should not be modeled as uniformly distributed parameters, but should be assigned a distribution that has a smaller probability of reaching the maximum values, such as the normal or triangular distribution.

Note that the objective metrics from the DMO study, Table 2.4, compare very well to the performance output variables that Deitchman concluded would translate as measures reflecting training effectiveness shown in Table 2.2. Deitchman's study primarily used field data because simulation data was scarce at the time. For any mission scenario, mission essential competencies (MECs) can be found and used for data collection, but many skills apply to diverse scenarios and tend to be more commonly needed. These include Deitchman's recommended variables and Schreiber's et al. top seven outcome metrics [60].

Table 2.4: DMO Training Objective Measures Result Summary [58]



Variable Name	*Change Monday–Friday (%)*	*p*
"Top Gun" scoring scheme[a]	+314.21%	<.01
No.of enemy strikers reaching target	−58.33%	<.01
Closest distance achieved in #1	+38.10%	<.04
No.of Viper mortalities	−54.77%	<.01
No.of enemy strikers killed (before reaching base)	+75.26%	<.01
No.of enemy aircraft killed	+9.20%	<.01
Proportion of Viper AMRAAMs resulting in a kill	+6.82%	<.03
Proportion of Threat Alamos resulting in a kill	−51.60%	<.01
Avg. time allowing hostiles into MAR (sec)	−55.20%	<.01
Avg. time allowing hostiles into N-pole (sec)	−60.33%	<.01
Slant range at AMRAAM pickle	+10.31%	<.01
Mach at AMRAAM pickle	+5.28%	<.01
Altitude at AMRAAM pickle	+7.97%	<.01
Loft angle at AMRAAM pickle	+14.80%	<.01
G-loading at AMRAAM pickle	*ns*	*ns*
Detonation range (hits and misses)	+8.12%	<.01

Note. AMRAAM = Advanced medium range air to air missile; MAR = Minimum Abort Range.

[a] Compositescore of fratricides, strikers killed before or after target, and hostile fighter and F-16 mortalities.

2.1.6 Bahlis and Tourville Model

In 2005, Bahlis and Tourville proposed that a shift from modeling training effectiveness through historical data collection and reduction to a predictive analysis is more beneficial in determining where training resources should be allocated [61]. The ultimate goal of traditional training models is to derive training value from it's resultant impact on mission objectives [40, 62]. Unfortunately, in practice, the final measurement to determine training's effect on the business is rarely ascertained and this lack of data leads to organizations devaluing training. To potentially increase the value of training, Bahlis and Tourville presented six strategies that can be applied individually or in any combination during the initial training planning phase [61]. The first strategy, aligning training with mission goals, defines a method to ensure resources are invested in training programs that will have maximum impact on the unit's overall target performance. In step one, the organization's or unit's mission and/or goals are defined and prioritized. Next, all the tasks and subtasks associated with achieving the goals are identified and assigned to a team(s) or specific job, depending on the task level. Step four reviews each task and determines if the task is one that would require training by assessing "the attributes of each task, i.e., level of difficulty, level of importance/criticality, and frequency of performance." For the tasks that require training, the knowledge, skills, and attitudes needed to perform those tasks are prioritized. A gap analysis is then performed by comparing the current capability of the team(s) or person assigned to the task with the required knowledge and skill set. At step seven, the availability of funding and resources is examined and any training implementation issues should be identified. Now that training tasks have been aligned to mission goals, tasks prioritized in order of importance, capability gaps identified, and budget constraints understood, the final step is to prepare a plan of action from the compiled data to help determine which training programs will generate the greatest impact. The data provides justification for particular training investments based on the organizational or unit's priorities and goals. This method establishes tangible benefits between training activities and mission objectives before training investments are expended using a predictive analysis model.

2.1.7 Training Effectiveness Models Summary

The six training effectiveness models reviewed above are methods that contribute to the current state-of-the-art in assessing the impact of training activities relative to meeting performance goals and objectives. The described models reveal criteria necessary in answering RQ1 "what is an appropriate method of measuring training effectiveness during early phase defense acquisition to aid decision-makers in DOTMLPF-P alternative selection?" To address the problem of training being properly considered as a solution during the military SoS CBA process, a method must exists that: 1) connects training results to mission specific goals, 2) is based primarily on objective data (can be supported by subjective data), 3) accounts for variation of skill levels, 4) includes uncertainty analysis, and most importantly, 5) can be used to predict, rather than simply evaluate, performance results after training is complete. Table 2.5 identifies whether each model addresses these five criteria in full (✓), partially (◎), or not at all (**x**).

Table 2.5: Training Effectiveness Model Summary Criteria

Criteria Model	Predictability Model	Objective Data	Links Training to Goals	Competency Level	Uncertainty
Kirkpatrick	×	○	○	×	×
Deitchman	✓	✓	✓	○	○
Bell & Waag	×	✓	○	×	×
Bahlis & Tourville	✓	✓	✓	✓	×
Schreiber et al.	×	○	✓	×	×
Clark	×	○	○	×	×

2.2 Theories of Learning

Part one of research question two, RQ2.1, asks how to quantify the benefits of soldiers' training in terms of effectiveness. Research question three, RQ3, delves deeper into this question and asks how to quantify increased knowledge, skills, and attitudes in training system design. Aguinis and Kraiger define training effectiveness as "the study of individual-, group-, or organizational-level factors that influence learning in training and transfer after training [35]." Traditionally, training effectiveness is measured using post-training data collection and surveys as described in the review of existing training evaluation methods in section 2.1. But to predict the effectiveness of training, a measurement system has to be used that does not rely on any post-training material. The factors influencing learning in training that are available for pre-training assessments are contained within the instructional philosophy used to create the system design. The benefits of instruction are indicated in terms of what the learner is to accomplish, called instructional or learning objectives (LOs) [63]. LOs perform two important functions for instructional designers, instructors, and evaluators. They provide a development tool for selecting and organizing training activities and resources to facilitate effective learning. Secondly, LOs provide a structure for formulating ways to evaluate student learning.

In the past, military training involved teaching a person or thing how to perform a task, but that is not satisfactory in today's military environment. Now soldiers must know how to think and make real time problem solving decisions [64]. With the advancements of technology so prevalent in today's society, machines and robots have been designed and are used when tasks only require the ability to follow set rules, commands, and instructions. To prepare soldiers for the situations they will encounter today, it is important to understand and enhance their human thinking ability. There is a difference between educating a soldier and training a soldier. "At its most basic level, training can be thought of as the planned and systematic activities designed to promote the acquisition of knowledge (i.e., need to know), skills (i.e., need to do), and attitudes (i.e., need to feel) [65]." In order for trainees to obtain the required competencies to successfully perform the overall mission objectives, training must result in sustainable changes in behavior and cognition. "The performance of any system is directly dependent on the training of the warfighters who operate and maintain the system [10]."

Learning is a desired outcome of training, but sometimes training fails to produce any learning [65]. Learning involves the acquisition of "new knowledge and behaviors as a result of practice, study, or experience. It involves relatively permanent changes in behavior and affect." Teaching a warfighter how to think and make the most appropriate decisions, as part of their training, requires the proper alignment between learning and training objectives. In the past, training research and theory have been criticized for lagging behind developments in learning theory and other areas of psychology [66, 67].

Over the last couple decades, training research has made great advancements incorporating practical applications of general learning theory and models [68]. Three traditional learning theories will be examined herein: Behaviorism, Cognitivism, and Constructivism. This information is presented to provide a brief background into the theories that have been debated for over a century about how people learn. This review is not an attempt to describe all of learning theory. But if training is to be evaluated for effectiveness and efficiency, there has to be an understanding of the individual learners' experience. An appreciation for the science of psychology and the work of psychologists who have studied and researched human learning is warranted. Stemming from these learning theories are the taxonomies that should be used in evaluating training effectiveness. For an exhaustive discussion on the psychology of learning, the reader is referred to one of the numerous psychology textbooks or handbooks; here are a few recommendations in addition to the references used in the following paragraphs [69-71].

While investigating learning theories for training effectiveness evaluations, a proper examination of the learning domains used while developing training programs is necessary. Learning occurs in one of three educational activity domains, as identified by educational psychologist Dr. Benjamin S. Bloom and his colleagues in 1956: cognitive, affective, or psychomotor [72]. The goal of this committee was to develop a method that captured how to foster higher forms of thinking in education, such as analyzing and evaluating, rather than rote learning (remembering and regurgitating facts). They believed this classification was essential in order to discuss and compare student achievement, while taking on a larger effort to develop standardized testing. Bloom spent his career contemplating and researching the very nature of thinking, resulting in authoring or co-authoring 18 books [73]. Since its original publishing, *Taxonomy of Educational Objectives: The Classification of Education Goals*, has been translated into 22 languages and is one of the most cited and applied educational references for curriculum development and teaching strategies nationally and internationally [74, 75]. Here is a brief summary of the three learning domains. The following paragraphs will provide further details, and examine multiple taxonomies that have been and are currently in use by varying educators and industries, to describe and evaluate the learning process. Cognitive learning relates to one obtaining intellectual skills or knowledge from facts, procedures, rules, and principles [76]. It includes heightened knowledge and better mental representations. Affective skills are emotion based, such as a person learning to contend with stress or changing their attitude and feelings. The Psychomotor domain involves the use of motor skills and physical movement, where skills are developed from practice and execution. This may include developing a new skill or improving an existing one.

2.2.1 Educational Terms

In addition to the definitions of learning and training previously discussed, it will be useful to define some of the terms already used and that are forthcoming.

2.2.1.1 Learning Theory

Learning has occurred when a person can exhibit a change in behavior or performance potential resulting from a specific experience and interaction with another being or thing [67]. Learning theory specifies the link between what is learned and the conditions under which learning occurs [77]. There are three basic components to any learning theory: the results, the means, and the inputs. The results are the changes in performance, which the theory tries to explain. The means are the processes used to make the changes; and the inputs are anything that triggers the process to occur, including resources and experiences. Learning theories are useful for understanding why an instructional design works by explicitly addressing which features of the learning environment promote intentional learning and how they may be developed [78]. In order to successfully predict the effectiveness of a training program the design of the training system must be examined to see which learning techniques are used in the instructional design, and if they are appropriately applied. It is also important to compare the methods selected to industry standards to determine if other proven practices exist that better satisfy the training objectives.

2.2.1.2 Instructional Design

Instructional design refers to the methodical process of translating learning principles and teaching practices into plans for instructional materials, activities, information resources, and evaluation [79]. It consists of four fundamental components: objectives, learners, methods, and evaluation. Every experience developed is focused towards one or more goals for learning. The work of an instructional designer parallels to that of an engineer. Both plan their work based on doctrines that have been successful in the past, and design their solutions for functionality as well as end-user appeal, effectiveness, and efficiency. At a most basic level, engineers follow the laws of physics, and an instructional designer follows the psychological principles of instruction and learning that have been researched over centuries. They both follow the problem-solving process to aid in decision-making, and most times the final result for both in the development process is a specification or plan. Engineers normally hand off their specifications or drawings to a production facility, and an instructional designer gives their plans to a media production specialist or training facility for implementation.

2.2.1.2.1 Instruction Versus Education

Education is a broad term that can describe any and all experiences where people learn. This includes unplanned, incidental, and informal activities. For example, after a person receives their driver's license they continue learning how to maneuver through traffic via trial and error. There are only so many hours designated to driving instruction, and all situations that the driver will encounter are not learned in driving school, especially those involving heavy highway traffic. Instruction, on the other hand, is "the deliberate arrangement of learning conditions to promote the attainment of some intended goal" [80]. The idea behind instruction is that the student gains a capability that can vary in qualitative and quantitative form. All instruction is part of education, but not all education is instruction [79]. The difference is in the systematic planning and development that goes into instruction, to ensure effective, efficient, and attractive experiences as the student learns a particular objective. The majority of instruction in business, military, and government settings can be called 'training' because the experiences are focused on preparing personnel with specific on-the-job skills (identified learning goals).

2.2.1.2.2 Design

In general, design involves an orderly and thorough planning and iteration process prior to the development of something or the implementation of some plan for the purpose of solving a problem [25]. Within instructional design, design refers to the level of precision, care and expertise used to develop the instructional material [79].

Instructional designers understand that poor planning can result in serious consequences, such as wasted time and resources, unmotivated and withdrawn students, and, in the most extreme cases, loss of life.

2.2.1.3 Taxonomy

Taxonomies are orderly classification systems designed to operate in a specific field of study based upon the natural relationships within that field [25]. In education and training, classification systems are useful for setting objectives based on the level of student understanding or skills needed to achieve particular learning outcomes. It is desirable when learning and teaching new skills that a well-structured pattern is followed [81]. This accounts for students' cognitive abilities, and aids instructors in assessing the qualitative leaps in students' learning and development. Taxonomies of the learning domains presented in this chapter can be used to assess the progression in learning objectives for training evaluation purposes. The taxonomies of learning skills (psychomotor, cognitive, and affective) provide generally useful metric evaluation systems [82]. Critics of taxonomies argue that they turn complex subject matters into simple and rigid hierarchical processes [83]. But models or taxonomies also act as catalysts to inspire thought, and have proved useful in categorizing differences in evaluations [76, 84]. Taxonomies should not be accepted as theories. Rather they serve as heuristics designed for clarification and classification.

2.3 Behaviorism

The Behaviorist theory focuses on how the environment helps to shape the learning process of an individual, and places a heavy emphasis on observable conduct. It associates learning with changes in either the form or frequency of observable action [85]. In behaviorism, learning is accomplished when a proper response is demonstrated after a specific environmental stimulus is presented to the learner. For example, when presented with a geometric problem showing a triangle with side lengths of 3 and 4 with a right angle between them and question “what is the length of the hypotenuse?”, the student replies with the answer of “5.” The stimulus is the right triangle with dimensions and the proper answer is the associated response. The key elements in behaviorism are the stimulus, the response, and the association between the two. How the connection between the stimulus and response is made, reinforced, and sustained is a primary focus. The consequences of a response are observed and the custom of rewarding learners for correct responses, and punishing or ignoring incorrect responses was practiced to increase the learner’s correct response probability. Behaviorism does not try to define the structure of a student’s knowledge, or consider the mental processes necessary for the learner to use [86]. The student is characterized as being reactive to conditions in the environment in contrast to taking an active role in discovering the environment. This creates a weakness because if a person finds him or herself in a situation where the impetus for the correct response does not occur, the person will not respond. For instance, a pilot who has been conditioned to respond to a certain cue for landing as part of the autopilot sequencing could have a jarring and possibly catastrophic landing if an anomaly occurs, because he or she does not understand the aircraft system. On the other hand, when a person is focused on a clear goal and can respond automatically to stimuli or that goal, the results of behaviorism are valuable. In World War II, pilots were given ‘spotter cards’ that showed silhouettes of enemy aircraft and naval ships as stimuli to enable quick life saving reactions [87].

Many psychologists have rejected behaviorism as originally published, because it lacks the explanations of why and how people actually learn. However, the fact that it consistently observed the action-reaction relationship of student learning should not be discounted because there is evidence of behaviorism throughout our learning environment. There are at least two enlightenments that resulted from behaviorism that relate directly to instructional design and training. One, the goal of behavioral objectives was to identify the actual behavior that the learner would be able to display at the end of instruction, the conditions under which it would be displayed, and the criteria that would determine acceptable performance [88]. This principle is used in all of our education systems from elementary to post-secondary levels, and even in the workplace. A very simple example is how everyone responds to the stimuli of time during a day. In high schools, when the school bell rings, students and teachers alike know it is either time to start the first class, move to another class, or be dismissed from school for the day [89]. As we move to higher education and the workplace, in most cases, the ringing bell is no longer present, yet students and workers all follow specific behaviors based on the time of day. Although the school bell, or time of day, does not denote any type of learning response, it does establish in every student and worker a conditioning – how to act and what to expect. The other tradition developed from behaviorism, that impacts training effectiveness, is the idea behind how to best present information to a student to enable them to demonstrate a specific behavior. In behaviorism, target behaviors are divided into small, easy-to-achieve steps presented in a logical sequence that builds toward the final complete behavior [88]. By carefully sequencing the components of the final desired behavior, students can master each step before moving on to higher-level concepts and/or tasks. This is an instructional approach consistently used today, because it works, and was observed during the studies of student conduct.

One of the most noted psychologists who advocated behaviorist learning theory stated, the "ultimate goal of education is to bring about behavior that will ensure survival of the human species, societies, and individuals" [90]. It can be argued that teaching someone to survive is not really learning. Of course, it all depends upon which definition of learning one refers to. The type of learning that behaviorism stimulates is debatable, but in terms of training, behaviorism can be linked directly to the psychomotor conditioning of a person and this is one important aspect in evaluating a training program for military soldiers. The psychomotor, or motor skill, learning domain is concerned with the general area of muscle development and coordination, and several taxonomies exist in literature[91, 92]. Equipment and/or tools may be needed to perform psychomotor skills, and speed maybe a factor. Testing usually requires more than just paper-based assessments. A demonstration where the student physically acts out the desired behavior is normal in order to determine the mastery skill level. In the early 1970s, several psychomotor taxonomies were published and three that remain popular today are shown in Figure 2.4 and discussed below [93-95]. A more recent taxonomy by Dr. Timothy Ferris is also presented that not only captures the psychomotor learning process, but accounts for advanced intellectual ability and where cognition and motor skills work together [96]. In general, psychomotor taxonomies describe a progression from simple observed behavior to mastery of physical skills. However, no taxonomy is universally accepted for motor skill development [63].

Mastery learning is an instructional design technique that can be applied directly for teaching motor skills during training. Mastery learning is based on the premise that learners must acquire skills in incremental, sequential progression, with pre-requisite skills being learned (mastered) prior to attempting more difficult and complex tasks [97]. It is believed that with proper instruction and enough time all students will achieve a satisfactory learning outcome. Thus, in mastery learning, students progress through a syllabus only after acquiring pre-requisite skills [98]. In teaching any new behavior, the student should demonstrate a firmly established ability prior to moving on to tasks that are more closely approximated to the goal. If too large a gap between previously learned and currently expected skills is presented to the student, their behavior may fail and training may have to resume at the point where the learner has repeatedly demonstrated success. Most people have a good conceptual understanding of motor skill development via interactions with children, or playing instruments and sports. Researchers created taxonomies to help identify stages of psychomotor progress, and gain a better understanding how one can teach mastery skills.

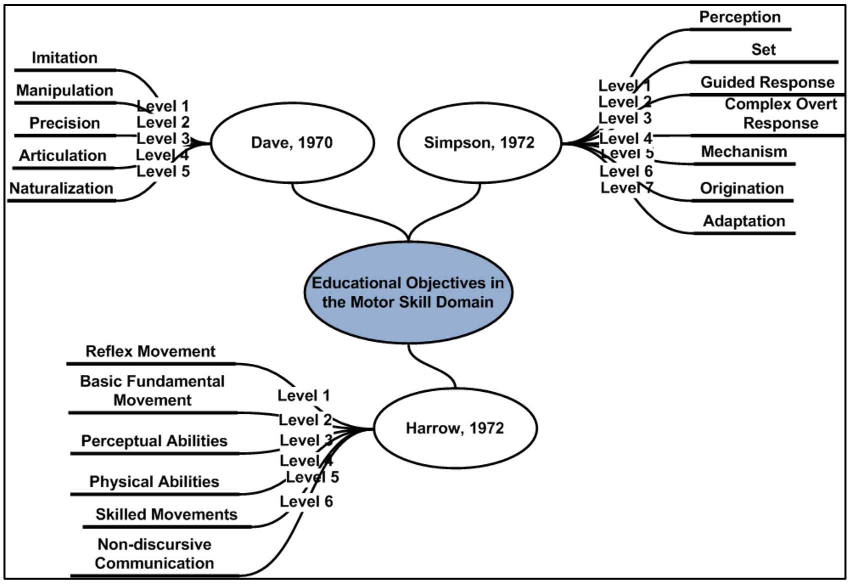


Figure 2.4: Psychomotor Taxonomies [78]

2.3.1 Dave's Psychomotor Taxonomy

As student of Bloom, Dave developed one of the most widely cited taxonomies in the psychomotor domain [99]. It is very simple and organized based on the refinement that occurs in order to complete a skill with increasing difficulty. The five categories include: imitation, manipulation, precision, articulation, and naturalization as shown in Table 2.6 [93] [2]. The first level, manipulation, involves observing and copying someone else's performance. Once the student reaches the final skill level, naturalization, the actions inherent in the skill have become automatic and no longer require deliberate focus. The naturalization level is most desirable because the knowledge and skills are ingrained [100]. The student does not have to focus on doing the task or remembering

---

[2] Dave's Psychomotor Taxonomy was presented at the International Conference of Educational Testing in Berlin in 1967. It was also published as a chapter in *Developing and Writing Behavioral Objectives* by Armstrong in 1970. Currently this book is out of print and is one of the most difficult publications to find at a reasonable price (Amazon Seller Cost exceeds $2,000). The taxonomy presented is based on credible sources that cite 93. Dave, R.H., *Psychomotor Levels*, in *Developing and Writing Behavioral Objectives*, R.J. Armstrong, Editor. 1970, Educational Innovators Press: Tucson, AZ..

procedures and can instead focus on environmental factors that may be necessary in decision-making. In this manner, Dave has indirectly related the psychomotor and cognitive domains. Although there are three learning domains, most on-the-job tasks require abilities that cross at least two, if not all three, realms [67]. The three domains are not mutually exclusive [101].

Table 2.6: Dave's Plan for Taxonomy of Psychomotor Outcomes [93, 102]

Primary Classification	Example and Key Words (Verbs)
1. Imitation: Observing and copying the behavior of someone else	Ex: Shifting the car gear from park to drive. Verbs: Copy, follow, mimic, repeat, replicate, reproduce, trace, observe, try
2. Manipulation: Guided via instruction to perform a skill. Instruction could come from taking a lesson, watching a video or reading.	Ex. Driving your car to work. Verbs: Act, execute, perform, re-create, build, implement
3. Precision: Accuracy, proportion and exactness exist in the skill performance without the presence of the original source.	Ex. Teaching a teen how to drive. Verbs: Calibrate, demonstrate, show, control, master, perfect
4. Articulation: Two or more skills combined, sequenced, and performed consistently.	Ex. Texting and driving (not recommended). Verbs: Adapt, construct, create, modify solve, formulate, improve, teach
5. Naturalization: Two or more skills combined, sequenced, and performed consistently and with ease. The performance is automatic with little physical or mental exertion.	Ex. Maneuvering your car into a tight parallel parking spot. Verbs: Design, development, specify, manage, invent

2.3.2 Simpson's Psychomotor Taxonomy

Elizabeth Simpson first attempted to classify skill performance in 1966 with a five level taxonomy. In 1972, she updated her publications and added two additional skill levels. The seven levels now include perception, guided response, mechanism, complex response, adaption and origination [94]. They are defined in Table 2.7 along with examples and key words (verbs). Similar to Dave's, this taxonomy attempts to categorize the hierarchy of classes required to learn a motor task or skill, but Simpson stretches beyond physical ability by including the idea of invention as mastery of a skill. The final level, origination, involves creating new movements, actions or expressions. This relates even more to the cognitive domain and higher order thinking capability. The student has developed the ability to not only physically perform a task, but also to create new methods and techniques that can enhance their own or someone else's performance.

Table 2.7: Simpson's Plan for Taxonomy of Psychomotor Outcomes

SOURCE: Adapted from [94]

Primary Classification	**Example and Key Words (Verbs)**
1. Perception (awareness): Becoming aware of stimulation and the need for action using sensory cues to guide motor activity	Ex: Noticing a car coming down the on-ramp to merge on to the highway in your lane. Verbs: Associate, compare, feel, hear, identify, inspect, listen, notice, recognize, scan, select, smell, taste
2. Set: Preparing for action mentally, physically or emotionally	Ex. Checking mirrors and turning on your signal to prepare to change lanes so the car can merge. Verbs: Adjust, arrange, comprehend, identify, locate, organize, recognize, respond, select.
3. Guided Response: Responding with assistance from a teacher or coach	Ex. Driving instructor gives the okay to switch lanes cautiously by turning the wheel and maintaining speed. Verbs: Adapt, correct, imitate, match, practice, repeat, produce, simulate
4. Mechanism: Responding habitually	Ex. Driving your car to work daily. Verbs: Assemble, fasten, manipulate, mix, mold, set-up, shape
5. Complex Response: Resolving uncertainty and performing difficult tasks automatically	Ex. Maneuver a car into a tight parallel parking spot. Verbs: Adjust, combine, coordinate, integrate, manipulate, regulate
6. Adaption: Altering responses to fit new situations	Ex. Respond effectively to unexpected driving situations, line inadvertent actions of other drivers. Verbs: Adapt, adjust, alter, convert, correct, revise, vary, integrate, order, standardize
7. Origination: Creating new acts or expressions	Ex. Develops a new and comprehensive driving instructional program. Verbs: Construct, create, design, develop, makes, formulate, invent

2.3.3 Harrow's Psychomotor Taxonomy

Anita Harrow developed a psychomotor taxonomy that was heavily influenced by her work with children with special needs. It focuses on assessing physical ability to perform behavioral tasks or activities. For this reason it is used more in physical education, sports, or recreational activities in comparison to the type of physical activities performed in the work environment. The taxonomy is comprised of six classification levels as shown in Table 2.8 [95]. Comparing the taxonomy of Dave, Simpson, and Harrow, it is obvious there was a difference in focus between the Dave and Simpson versus Harrow. Dave and Simpson attempted to capture phases of learning motor skills, whereas Harrow created groupings of different types of motor behavior. Harrow's taxonomy is not useful in the context of training evaluation, but is included here because of level six. Body language such as gestures and facial expressions can be good indicators to instructors about how the student feels. This crosses over with the affective domain, which is important and is discussed below. Most psychomotor taxonomies cross into the cognitive domain, but rarely are any presented with relationships to affectivity [101].

Table 2.8: Harrow’s Plan for Taxonomy of Psychomotor Outcomes [95]

Primary Classification	Example and Key Words (Verbs)
1. Reflex Movements: Automatic reactions. Reactions that are not learned.	Ex: Touching the stove and instinctively releasing your hand because it is hot. Verbs: Extend, flex, stretch, react, respond
2. Basic Fundamental Movement: Simple movements that can build up to more complex movements.	Ex. Walking Verbs: Grasp, throw, catch, punt, run, push, twist
3. Perceptual: Environmental cues that allow one to adjust movements. Visual, auditory, kinesthetic, or tactile discrimination.	Ex. Tracking a moving object with your eyes. Verbs: Draw, write, catch, coordinated movements
4. Physical Activities: Things requiring endurance, strength, vigor, and agility.	Ex. Running a marathon Verbs: All activities that require strenuous effort for long periods of time, muscular extension, a quick wide range of motion at the hip joints, or quick precise movement.
5. Skilled Movements: Activities where a level of efficiency is achieved.	Ex. Playing basketball Verbs: Adapt, construct, create, modify
6. Non-Discursive Communication: Effective body language, such as gestures and facial expressions and sophisticated choreography.	Ex. Ballet dancing Verbs: Interpretation

2.3.4 MIL-HDBK-29612-2A Psychomotor Taxonomy

Some of the most popular industry utilized learning domain taxonomies published have heavily influenced military standards and guidebooks. In the case of U.S. DoD handbook for instructional and training system design, the psychomotor taxonomy used parallels, and for half of the levels is exactly the same as, Simpson’s taxonomy. The DoD handbook refers to psychomotor learning as the “skills” learning domain [103]. Table 2.9, lists the taxonomy with definitions and standardized verbs used to distinguish the skill levels. Level one of the military handbook and Simpson’s taxonomy both acknowledge the student’s ability to recognize something that prepares them for physical movement. In level two of the DoD handbook, the student reacts immediately with their own physical movements. This does not occur until level three in Simpson’s taxonomy and is more of a guided process. In the DoD handbook taxonomy, the focus of levels two and three is on the trainee's ability to make a movement or respond, but it does not ensure that the student's response is correct. Level four, readiness, is the first time the physical movement must correspond to the specific desired action. Once the trainee's actions follow the learning objective, the top three skill levels are the same as Simpson's. The student moves to a state where they demonstrate complex skills. They then begin adapting what they have learned so they can accomplish tasks in different situations. Lastly, they begin creating new complex skills to successfully complete duties in an unknown environment and conditions. One key observation in the DoD handbook is that levels five, six, and seven refer to mental and physical skills. The bottom three skill levels focus on physical abilities and movements, but at level four there is a shift to recognize that cognitive and psychomotor skills are required to perform, modify, and invent new complex tasks.

Table 2.9: MIL-HDBK-29612-2A Taxonomy of Psychomotor Outcomes [103]

Primary Classification	**Key Words (Verbs)**
1. Perception (encoding): Sensory stimuli that translate into physical performance.	Detect, feel, hear, scan, see, smell, taste, visualize
2. Gross Motor Skills: Manual dexterity in the performance of physical skills.	Assault, carry, creep, depart, fall, hold, jump, lift, pull, run, stay, swim, throw, turn, twist, wear
3. Continuous Movement: Tracking or making compensatory movements based on feedback.	Advance, control, follow, guide, hover, land, maneuver, regulate, steer, take off, track, traverse
4. Readiness: Having readiness to take a particular action	Able, assist, challenge, cross, delay, guard, prepare, prime, ready, set, stand to
5. Mechanism: Performing a complex physical or mental skill	Adjust, assemble, balance, clear, cover, diagnose, disengage, display, elevate, enter, establish, fuel, ground, hoist, initialize, integrate, launch, load, maintain, navigate, perform, replace, retrieve, stow, support, transfer, troubleshoot, update, write
6. Adaption: Modifying a complex physical or mental skill to accommodate a new situation.	Acclimatize, accommodate, adapt, ambush, attack, bypass, conduct, deploy, direct, draw, evade, infiltrate, lead, occupy, patrol, prevent, protect, reconcile, relieve, suppress, tailor, temper, train
7. Origination: Creating a new complex physical or mental skill to accommodate a new situation.	Cause, construct, contrive, correct, initiate, invent, make, originate

2.3.5 Ferris' Psychomotor Taxonomy

After years of lecturing and supervising engineering courses that required laboratory tasks, Dr. Ferris noticed a significant discrepancy between student performances on written assignments versus their ability to complete the practical work [96]. This observation led him to investigate the three learning domains and eventually propose one that best fit his purposes as an engineering educator. Ferris stresses the importance of contextual dependency when defining a learning domain, "for example, the foundational psychomotor development of young children, sporting capabilities and trade and professional skill development" have differences. According to Ferris, the variety of interests associated with skill development makes it difficult to formulate a context-free description of psychomotor skills. In his development of a psychomotor taxonomy he "interpret[s] the psychomotor domain as concerning the whole of the interface between the person and the things and environment with which they interact, including physical action skills, the ability to use the five senses to perceive, and the ability to decide and to do appropriate action." The taxonomy is mapped to three descriptions of knowledge: 'know that', 'know how', and 'knowing'. 'Know that' describes declarative knowledge or knowledge expressed in representation of concepts. It is mainly associated with cognitive domain learning objectives, but is many times the first step in psychomotor skill development. A trainee should have some understanding of what they are doing before they take an action. 'Know how' describes student's ability to actually perform the task. 'Knowing' stresses the learners' ability to suitably choose and perform an action effectively. 'Knowing' is different than 'know how' because it is usually requires some judgment-call or decision, made with foundational knowledge, regarding when someone should act, or situational constraints, or other factors that impact the course of action. Therefore, ‘knowing’ is hierarchically superior to ‘know how’. The context of the psychomotor taxonomy proposed by Ferris is to assist educators in achieving learning outcomes as described in accreditation processes, such as used by Engineers Australia and ABET [104-106].

Ferris’ psychomotor taxonomy consists of seven primary classifications in the context of an electrical engineering circuit board design laboratory class, shown in Table 2.10. Recognition of tools and materials is the most basic skill concern and is the first step. It is important for trainee effectiveness and safety. This stage involves learning the names and descriptors of the tools and materials. This category applies to tools of all sizes from small and simple to larger or complex machinery. The larger and more complex the tools and materials, the more increased risk for safety and damages if used inappropriately. Step two moves the training from ‘knowing that’ to the initial ‘know how’. Students are taught to handle the tools and materials according to the methods for holding, lifting, moving, and setting them down. Each of these four handling processes can have implications for safety, security, and effectiveness, not only for the tools and materials, but also for people and the operating environment. Once the student has learned to properly handle the tools, they learn the basic tool operations. This involves using the tool to perform rudimentary tasks for single operations, which do not have to be productive by themselves. For example, if teaching someone to drive a car one may start by teaching how to press the brake. With the car in park this accomplishes little, but it is a critical basic operation for driving. Once the student learns the basic operations, they move on to performing a range of sequence specific tasks. This is step four, competent operation of tools. In the car example, this is where a student may engage in the following sequence of tasks to drive the car: press and hold brake, change the gear shift to drive, apply signal if necessary, check mirrors for surrounding traffic, when clear remove foot from brake and press gas pedal appropriately. In this stage the trainee is actual producing useful outcomes. The next step is to expertly operate the tools. Now that the student understands the purpose of the tools and materials, and can operate them, they need to perfect those skills to perform tasks on a regular basis without so much focus on what they are doing but on the context of what is going on all around them. Continuing the car-driving example, the student can move from driving in the open parking lot to actually driving on the road with other vehicles. At this point, the student shouldn't have to focus so much on what and how they are driving, in comparison to where they are going and how other drivers are maneuvering around them. This is where 'knowing' arises. Judgments and decisions are now being made based on what the student knows about the tools and materials and how to operate them. Once the trainee has become efficient and effect with the tools, they can begin planning work operations, category six. This level requires an understanding of the work to be done, possible courses of actions based on the equipment available, and the ability to choose the appropriate method. At this point, the desire is for a student to take what they have learned and build a set of processes or tasks to deliver a final product or service, based on specification document(s). The final classification level is the ability to evaluate outputs and plan means for improvement. It heavily mirrors the final level of cognitive ability because it involves examining a product and reviewing it for quality and deficiencies, and providing a course of action for correcting and preventing future faults. Here it is very clear how Ferris takes into account all five senses in his taxonomy. He captures the “how to” of skill ability within the psychomotor domain but also requires some ability of higher order thinking.

The advantages of using this taxonomy are 1) it ensures trainees can not only perform a task, but also have a full understanding of why they are following a certain process, and 2) trains them to evaluate and identify any alternate methods that may be better suited for the work being performed. This is very important for not training people to imitate robots, but enabling them to make decisions and adapt to the operational environment. This psychomotor taxonomy works well for trade and professional education, where the student practices through doing practical actions. It is also good for those in design or supervisory roles. In order to effectively write out a sequence of actions for other people to follow, an appreciation and knowledge of the tools and processes being used is required. One must also be able to identify and solve issues that may arise from unexpected maintenance issues, different equipment used, environmental changes, etc. A possible criticism of Ferris’ psychomotor taxonomy is that it does overlaps with some of the cognitive and affective learning domain aspects. It parallels very well to Bloom’s cognitive taxonomy. This can be advantageous for someone searching for an encompassing taxonomy to assess all three learning domains, but may cause confusion for someone who only associates the psychomotor domain with physical motor skills (which is the more common belief).

Table 2.10: Ferris' Taxonomy for Psychomotor Outcomes [96]

Primary Classification	**Sub-classification**	**Mapping to Knowledge Type**
1. Recognition of tools and materials	1.1 Recognition of tools 1.2 Recognition of materials	Know that Know that
2. Handling tools and materials	2.1 Holding tools and materials 2.2 Lifting tools and materials 2.3 Moving or transporting tools and materials 2.4 Setting-down tools and materials	Know how Know how Know how Know how
3. Basic operation of tools	3.1 Holding tools steady for use 3.2 Operating the tool 3.3 Method to do each of the unitary actions with the tool	Know how Know how Know how
4. Competent operation of tools	4.1 Moving form one unitary task to another 4.2 Reliably performing tasks to an acceptable standard	Know how Know how
5. Expert operation of tools	5.1 Efficiently and effectively using the tools 5.2 Ability to focus on the broader context of the work	Know how, Knowing Knowing
6. Planning of work operations	6.1 Ability to conceive tool capability abstractly 6.2 Ability to envision the effect of a sequence of operations 6.3 Ability to develop novel work processes to achieve specified outcomes	Know that, Knowing Know that, Knowing Knowing
7. Evaluations of outputs and planning means for improvement	7.1 Ability to recognize the cause of product characteristics 7.2 Ability to pre-emptively judge the effect of modification of work process 7.3. Ability to recommend improved work methods 7.4 Ability to critically review the effectiveness of methods to perform novel tasks	Know that, Knowing Know that, Knowing Know that, Knowing Know that, Knowing

2.3.6 Psychomotor Taxonomy Summary

The psychomotor taxonomies presented are a small representation of the many classification systems that have been developed to capture the motor skill learning process. As can be seen in Table 2.11, they tend to follow a pattern of very low abilities, such as natural reflex movements or imitating observed behavior, and move into categories of learning from practice. They all include some type of expert or habitual performance level, which diverts the trainee's direct focus from what they are doing and enables them to assess their surroundings. Primary differences arise in that some taxonomies, such as Simpson and Ferris, which consider the ultimate motor skill ability as creation and evaluation of actions and processes. Choosing which taxonomy is best for evaluating general training effectiveness depends on the type of training and amount of details provided. For the C-130J case study that is a part of this research it is recommended to use the psychomotor taxonomy proposed by Ferris. It logically progresses through the motor skill development process and can be easily applied for any crew position. The last two steps are also very important for this type of military training. Once the crew learns how to operate the aircraft, they need the ability to plan, evaluate, and adjust to mission requirements. The fact that this crosses into the cognitive learning domain is seen as a positive. The level descriptions are clear enough that any confusion between learning domains can be avoided. Use of Ferris' psychomotor taxonomy is not recommended without a cognitive and affective taxonomy, because lower order thinking skills and affectivity are not addressed.

Table 2.11: Classification Comparison of Psychomotor Learning Objectives

Skill Levels	Dave (1970)	MIL-HDBK-29612-2A (1972)	Simpson (1972)	Harrow (1972)	Ferris (2010)
Low (Imitation)		Perception (Encoding)	Perception (Awareness)	Reflex Movement	Recognition
	Imitation (Copy)	Gross Motor Skills	Set	Basic Fundamental Movements	Handling
	Manipulation	Continuous Movement	Guided Response	Perceptual Abilities	Basic Operation
Practice (Development)	Develop Precision	Readiness	Mechanism	Physical Abilities	Competent Operation
	Articulation	Mechanism	Complex Overt Response	Skilled Movements	Expert Operation
High (Beyond Habits)	Naturalization	Adaptation	Adaptation	Non-Discursive	Planning of work instructions
		Origination	Origination		Evaluation of outputs and planning means for improvement

2.4 Cognitivism

The genesis of Cognitivism as a learning theory can be traced back to the early twentieth century, but only gained prominence after psychologists began challenging the assumptions and limitations of behaviorism [107]. The shift from behaviorism to Cognitivism grew from the behaviorist tradition's failure to explain why and how individuals make sense of and process information; in short, how the mental process actually works. As opposed to the emphasis on behavior, the cognitive school focuses on meaning and semantics [108]. According to cognitivist, "The human mind is not simply a passive exchange-terminal system where the stimuli arrive and the appropriate response leaves [behavior theory]. Rather, the thinking person interprets sensations and gives meaning to the events that impinge upon his consciousness" [109]. In cognitivist theory, the major emphasis is placed on how knowledge is acquired, processed, stored, retrieved, and activated by the learner during various phases of the learning process [110] [111]. Cognitive psychologists place more emphasis on *what* learners know, and on an understanding of *how* they have come to attain that knowledge, than on what they actually *do*. Therefore, the cognitive approach focuses on making knowledge meaningful, and helping learners organize and relate new information to prior knowledge in memory.

Many instructional methods exist today that are based on principles of Cognitive Learning Theory. Some of the more distinctive methods are cognitive apprenticeship, reciprocal teaching, anchored instruction, inquiry learning, discovery learning, and problem-based learning [107]. From a training system design perspective, where physical skill development is as important as mental and emotional growth, cognitive apprenticeship instructional method is best suited because it helps students grasp concepts as well as procedures. Cognitive apprenticeship is a blend of an old instructional process (apprenticing) mixed with strategic consultations [112]. The student (apprentice) initially observes the instructor. The instructor then gradually increases the problem solution responsibilities onto the student until the student is solving problems on their own. During this transition, the instructor coaches the student with questions and encourages them to think aloud about their solution process. This instructional approach includes the following phases as described by Yillmaz.

- Modeling: The teacher performs a task or explains a process for students to observe, which helps them understand what it takes to accomplish the learning task. Modeling provides students with the opportunity to generate conditionalized knowledge (i.e., when, where, and how to use knowledge to solve problems of different kinds).
- Coaching: While students do the same task, the teacher observes students and provides hints, cues, feedback, and help, if needed.
- Articulation: Students are asked to think out loud about how they performed the task and offer reasons for the strategies that they used. Having students articulate their implicit knowledge and strategies makes them explicit. The teacher can detect whether students have any misconceptions or use improper and inadequate strategies.
- Reflection: Students retrospectively think of their performance upon completing the task and compare their actions with the teacher's or other students' actions.
- Exploration: The teacher urges students to identify a problem, formulate a hypothesis, and seek needed information to solve it. Students look at the different

aspects of the problem from different perspectives on their own. This strategy is intended to promote students’ ability to think independently.

A key component of the cognitive apprenticeship instructional approach is the role of the student throughout the learning process. The student is actively involved in all stages of learning. In training, students are learning for a purpose, to achieve a set goal. To achieve that goal, they must assess the task requirements, their current ability, and what it will take to move forward toward the goal. When a person is fully aware of why they are learning and actively participating in the process, it is called metacognition, or “thinking about thinking” [113]. This is a much more learner-centered and learner-directed model for learning, and fits into the higher order cognitive skill ability. Unfortunately, many if not most, students do not practice metacognitivism because they have not been made aware of this practice [88]. Like any other skill, metacognition requires practice and a person will grow in this ability [112]. So not only can the instructor assess the students’ cognitive ability, but also the trainee can independently compare themselves and know if they are on track with course material. Building in instructional methods that support metacognition in the design of training curriculum allows the role of the instructor to be more passive. This is a optimal for computer based and media driven training environments, which can be cheaper and provide more opportunities for trainees because they are not restricted to a set classroom scheduling. Instructional techniques that support metacognition include writing journals, describing problem solutions in prose as well as mathematical format, and discussing problem-solving strategies in a group context.

2.4.1 Bloom’s Taxonomy

After developing the framework that learning consists of three domains, Bloom and fellow researchers established a specific taxonomy for the cognitive learning domain. Commonly referred to as “Bloom’s Taxonomy”, it is one of the first systematic classifications of the processes of thinking and learning [114]. Bloom’s Taxonomy is a six-level categorization system that uses observed student behavior to infer the level of a student’s achievement, Figure 2.5. The three lower levels are: knowledge, comprehension, and application. The higher three higher levels are: analysis, synthesis, and evaluation. The categories of Bloom’s Taxonomy are defined in Table 2.12, along with key words that are used to assess the student’s competence and suggested instructional strategies for use with each level.

Bloom presented the taxonomy as a progression from simple to more complex cognitive ability. Over time, users began dividing the levels into lower and higher levels of thinking. The original taxonomy is a stringent cumulative hierarchy [76]. In order for a student to reach the more complex or higher order thinking they must have mastered the simpler lower order categories; the higher levels were said to encompass the lower levels. However, this is one area where Bloom has been criticized. Since its original publication, there has been further research and empirical evidence showing that the three middle levels, Comprehension, Application, and Analysis, are cumulative and hierarchical, but the evidence was faint for ordering the top two categories [76]. Today many researchers would argue, “although the construct is hierarchical, subsequent classes of behavior include some, but not necessarily all, of the behaviors found in the lower levels. Thus, this is a hierarchical framework of conceptual sophistication and not a prescriptive model

[115].” Viewing the taxonomy as more of a framework is what many fields outside of traditional education have done. This allows for use of the cognitive thinking process principles while leaving room for more creative and industry specific application.

The idea behind Bloom’s Taxonomy is that the classification system can be used for teachers to evaluate their student’s ability, but also so that the student can recognize and evaluate their own learning process and progression. Within the elementary education (K-12) learning environment Bloom’s Taxonomy is primarily used in curriculum development and teaching strategies, but is rarely made aware to the students directly. Putnam suggested that this is because metacognitive skills are context-dependent, and that younger students lack the ability to focus on abstractions well enough to transfer them from one context to another [116]. This leaves elementary education focusing on assurance that students can comprehend and analyze within a certain subject.[3] In this case they are not reaching the highest level of thinking; where they have the ability to take what was learned in one context and move to another to create, interpret, and/or defend based on past knowledge and their own insights. This leaves opportunity for increasing cognitive ability at the post-secondary level. Forms of post-secondary education include college and university, vocational and trade schools, as well as military training. Remember that Bloom developed this taxonomy with other university professors, for evaluation of their students. Since inception, its use has extended beyond colleges and universities and is found in government standards and

---

[3] It should not be implied that all elementary students are incapable of exercising higher order thinking skills. One of many examples is a study conducted with sixth graders to understand the connection between cognitive tool use and cognitive processes 117. Liu, M., et al., *Understanding the Connection Between Cognitive Tool Use and Cognitive Processes as Used by Sixth Graders in a Problem-Based Hypermedia Learning Environment.* Journal of Educational Computing Research, 2004. **31**(3): p. 309-334.. In this experiment the students demonstrated the evaluation process within the problem-based environment proposed.

handbooks for instructional design. To determine the effectiveness of any training program; it is important to 1) understand the level of cognitive ability that will be required to successfully perform the task(s) and 2) to decide if the training program is designed to ensure the student achieves the corresponding cognitive level. Military training programs that involve complex missions should include lessons that allow soldiers to exercise the highest form of cognitive ability ensuring they are able to handle the situations that will undoubtedly arise outside of the simulations and training flights used for practice.

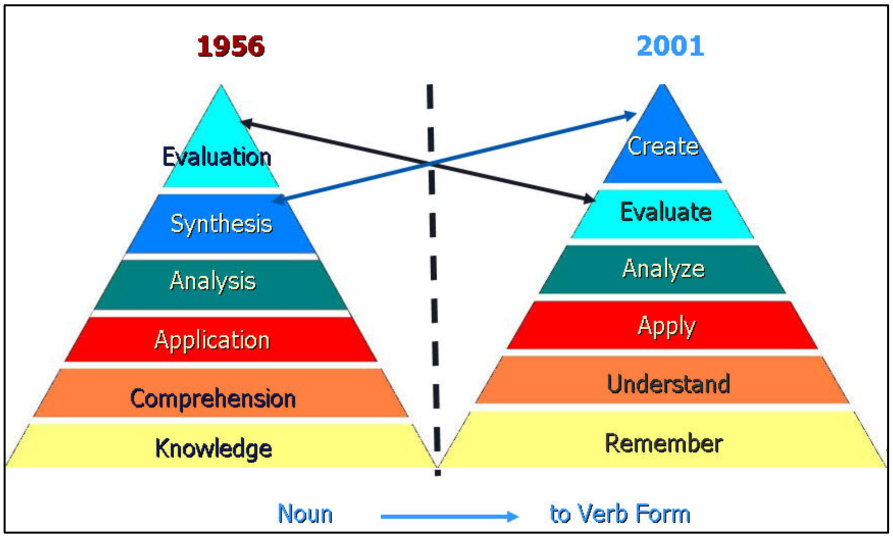


Figure 2.5: Bloom’s Taxonomy Original and Revised [118]

Table 2.12: Bloom’s Taxonomy of Cognitive Outcomes [72, 119]

Primary Classification	**Suggested Instructional Strategies (SIS) Key Words (verbs)**
1. Knowledge: Recall data or information.	**SIS**: lecture, visuals, video, audio, example illustrations, analogies. **Key Words**: arranges, defines, describes, identifies, knows, labels, lists, matches, names, outlines, recalls, recognizes, reproduces, selects, states.
2. Comprehension: Understand the meaning, translation, interpolation, and interpretation of instructions and problems. State a problem in one's own words.	**SIS**: questions, discussion, review, test, assessment, reports, learner presentations, writing. **Key Words**: comprehends, converts, defends, distinguishes, estimates, explains, extends, generalizes, gives an example, infers, interprets, paraphrases, predicts, rewrites, summarizes, translates.
3. Application: Use a concept in a new situation or unprompted use of an abstraction. Applies what was learned in the classroom into novel situations in the work place.	**SIS**: exercises, practice demonstrations, projects, sketches, simulations, role-play, micro-teach. **Key Words**: applies, changes, computes, constructs, demonstrates, discovers, manipulates, modifies, operates, predicts, prepares, produces, relates, shows, solves, uses.
4. Analysis: Separates material or concepts into component parts so that its organizational structure may be understood. Distinguishes between facts and inferences.	**SIS**: problems, exercises, case studies, critical incidents, discussion questions, test. **Key Words**: analyzes, breaks down, compares, contrasts, diagrams, deconstructs, differentiates, discriminates, distinguishes, identifies, illustrates, infers, outlines, relates, selects, separates.
5. Synthesis: Builds a structure or pattern from diverse elements. Put parts together to form a whole, with emphasis on creating a new meaning or structure.	**SIS**: projects, problems, case studies, creative exercises, develop plans, constructs, simulations. **Key Words**: categorizes, combines, compiles, composes, creates, devises, designs, explains, generates, modifies, organizes, plans, rearranges, reconstructs, relates, reorganizes, revises, rewrites, summarizes, tells, writes.
6. Evaluation: Make judgments about the value of ideas or materials.	**SIS**: case studies, projects, exercises, critiques, simulations, appraisals. **Key Words**: appraises, compares, concludes, contrasts, criticizes, critiques, defends, describes, discriminates, evaluates, explains, interprets, justifies, relates, summarizes, supports.

2.4.2 Revised Bloom’s Taxonomy

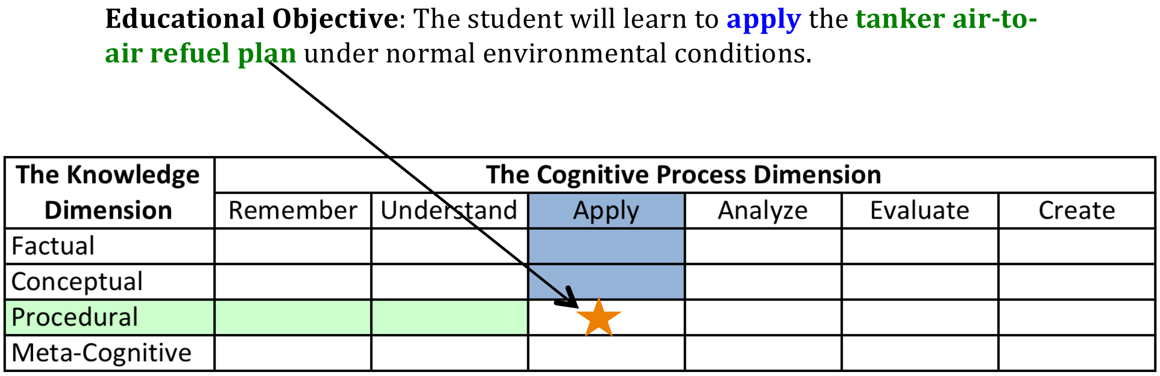
A former student of Benjamin Bloom, Lori Anderson, gathered a new group of cognitive psychologists, curriculum theorists, instructional researchers, and testing and assessment specialist together during the 1990’s to update Bloom’s Taxonomy, hoping to add relevancy for 21st century students and teachers [76]. The revision they published in

2001 has three primary changes. There was a terminology change where all six categories were changed from nouns to verbs and a slight realignment in the order placement. They believed that teaching objectives are meant to describe the learner's thinking processes rather than the learner's behaviors. Starting from the lowest level of cognitive thinking they changed knowledge to remember, comprehension to understand, application to apply, analysis to analyze, synthesis to create, and evaluation to evaluate, as shown in Figure 2.5. Changing these terms all fits within the definition or description of the original intent of Bloom's Taxonomy. Table 2.13, lists the terms and description of both the original and revised classifications for ease of comparison. Changing the names from nouns to verbs is a minor modification. The second change, which can be noticed in Figure 2.5, is that the top two higher order thinking categories are flipped. In the original Bloom's Taxonomy, category five, is synthesis and culminates with evaluation; the revised version has evaluation at level five and the final process in cognitive learning is the ability to create. This change more accurately depicts the process of active learning [120]. Once a person has practiced applying and analyzing something, they then make their own judgment about how useful the information is to what they are trying to accomplish. If the information as learned applies directly, then it is used directly. However, many times we take what we've learned and modify it or develop something new to suit our individual needs and purposes. This is what is ultimately necessary in military training. Evaluating information is critical, but the last step is to then be able to combine that information within the current circumstance to produce a result. Rarely does the information and environment align like a perfect puzzle. The reality is that new thoughts and ideas are needed, or the ability to put together information learned in an alternate environment. In a military context, solutions are developed in real time based upon what the soldier has been taught during training.

Bloom and later colleagues recognized that the original taxonomy was being used “unexpectedly” by groups other than the authors’ intent which was for university assessments only [74]. The revised taxonomy takes this into account with the other two primary changes. One change addresses a much broader audience and places emphasis on creating a more authentic tool for curriculum planning, instructional delivery, and assessment, as well as the proper alignment of these three [76]. With this shift in emphasis, what appears to be the most dramatic change in the revision to Bloom’s Taxonomy is the structural change. Bloom’s original cognitive taxonomy is one-dimensional. The Revised Bloom’s Taxonomy has a simple one-dimensional form, but actually takes the form of a two-dimensional table, Figure 2.6. One of the dimensions is called “The Knowledge Dimension” and is divided into four levels defined as Factual, Conceptual, Procedural, and Meta-Cognitive. The Knowledge Dimension describes the type of knowledge to be learned. The second dimension is “The Cognitive Process Dimension.” This dimension describes the process of learning and is composed of the six levels of the taxonomy: Remember, Understand, Apply, Analyze, Evaluate, and Create. Each of the four Knowledge Dimension levels is subdivided into either three or four additional categories. For example, Factual is divided into Factual, Knowledge of Terminology, and Knowledge of Specific Details and Elements. The Cognitive Process Dimension levels are also subdivided with the number of sectors in each level ranging from three to eight subcategories. As an example, Remember is sectioned into three categories, Remember, Recognizing, and Recalling. The resulting grid of acute detail is very helpful to teachers in writing objectives and aligning standards with curriculum. The total of 19 subcategories within a two-dimensional organization provides more clarity about the fit of a specific verb or product at a given level [121]. The revised taxonomy presents teachers with a more descriptive tool to use when creating their lesson plans. In general, the taxonomy and structure works well for preparing curriculum and measuring students' cognitive ability. For pre-training program effectiveness evaluations the two-dimensional structure is not entirely useful because the assessment information required is not yet available to assess the meta-cognitive knowledge dimension. For the Factual, Conceptual, and Procedural Knowledge Dimensions this taxonomy could be used in training design. An example is shown in Figure 2.6 of how the *procedural* knowledge dimension and *apply* cognitive process dimension can be used to write a C-130J aircrew-training objective.

Table 2.13: Bloom's Taxonomy of Cognitive Outcomes Revised [72, 76]

Original Bloom's Taxonomy	**Revised Bloom's Taxonomy**
Knowledge: Recall data or information.	**Remembering**: Retrieving, recognizing, and recalling relevant knowledge from long-term memory.
Comprehension: Understand the meaning, translation, interpolation, and interpretation of instructions and problems. State a problem in one's own words.	**Understanding**: Constructing meaning from oral, written, and graphic messages through interpreting, exemplifying, classifying, summarizing, inferring, comparing, and explaining.
Application: Use a concept in a new situation or unprompted use of an abstraction. Applies what was learned in the classroom into novel situations in the work place.	**Applying**: Carrying out or using a procedure through executing, or implementing.
Analysis: Separates material or concepts into component parts so that its organizational structure may be understood. Distinguishes between facts and inferences.	**Analyzing**: Breaking material into constituent parts, determining how the parts relate to one another and to an overall structure or purpose through differentiating, organizing, and attributing.
Synthesis: Builds a structure or pattern from diverse elements. Put parts together to form a whole, with emphasis on creating a new meaning or structure.	**Evaluating**: Making judgments based on criteria and standards through checking and critiquing.
Evaluation: Make judgments about the value of ideas or materials.	**Creating**: Putting elements together to form a coherent or functional whole; reorganizing elements into a new pattern or structure through generating, planning, or producing.



Educational Objective: The student will learn to **apply** the **tanker air-to-air refuel plan** under normal environmental conditions.

The Knowledge Dimension	The Cognitive Process Dimension					
	Remember	Understand	Apply	Analyze	Evaluate	Create
Factual						
Conceptual						
Procedural			★			
Meta-Cognitive						

Figure 2.6: Revised Bloom's Taxonomy C-130J Objective Example

2.4.3 Gagne's Taxonomy

Robert M. Gagne was an instructional psychologist who spent much of his career working at universities and for the U.S. Air Force training personnel, including pilots. He dealt particularly with problems to define what skills and knowledge are required for someone to be an effective performer at a given job [77]. He identified job requirements and then focused on determining how personnel training best fit those requirements. A contemporary of Benjamin Bloom, Gagne published the first edition of *The Conditions of Learning* in 1965 and the fourth edition in 1985. In this book, Gagne proposed not only a new integrated taxonomy of learning outcomes, but also specific leaning conditions for each classification level, and instructional events to activate the learning process.

Gagne was the first to present an integrated taxonomy that included all three domains. He proposed that learning consists of five major outcomes: verbal information, intellectual skills, cognitive strategies, attitudes, and motor skills [91]. A summary definition of each classification and the corresponding learning conditions are provided in

Table 2.14. Verbal information is the category where a person learns declarative or factual knowledge. The ability of a student to remember or recite the NATO phonetic alphabet (Alpha, Bravo, Charlie, etc) is an example of verbal information. In comparison with Bloom's cognitive taxonomy it parallels the knowledge and comprehension levels. Sometimes students memorize information with no association to its meaning or context. Gagne encourages instruction and learning conditions that enable the student to put a context to the information and be able to demonstrate what they have learned by explaining it in their own words or paraphrasing, instead of just repeating after the instructor. Intellectual skills are the second category, and are similar to procedural knowledge from the two-dimension Revised Bloom's taxonomy. According to Gagne, intellectual skills are divided into five hierarchically ordered subcategories: discrimination, concrete concepts, defined concepts, rules, and higher-order rules. Each subcategory is necessary to learn before moving to the next. Discrimination is the ability to distinguish one object from another based on perception (i.e. recognizing the differences between an airplane, helicopter, and high speed jet). The person does not have to know the names of the objects to point out the fact that they are different. For any particular subject matter, once discrimination is acquired, concrete and concept learning can begin. Concrete learning is where a person learns the name and starts grouping classes of like objects or events together. Then definitions and context are provided to the learner. At this point, a student would be able to recognize the types of airplanes that are commercial, versus fighter jets, versus heavy-lift aircraft. The fourth intellectual category is rules. Usually, rule learning involves the use of symbols to represent and interact generally with the environment [122]. But the emphasis here is not so much on

the student's ability to recite a rule, but to apply it appropriately to a class of problems. The final step in Gagne's intellectual skills is higher order rules. This requires the learner to combine simpler rules to solve complex problems. Learners may apply a new combination of rules they have learned individually. An example is giving a pilot an open-ended problem to define a flight plan from destination A to B. He or she must use the knowledge they have gained about the approximate mileage of certain aircrafts, safe airports for landing and refueling if necessary, etc. Higher order rules begins to overlap with the third primary classification in Gagne's taxonomy, which is cognitive strategies. Cognitive strategies are similar to the analysis, synthesis, and evaluation levels of Bloom's Taxonomy. It consists of some meta-cognition where students become aware and monitor their own thinking. They decide which learning techniques work best for them and how they can become successful in the learning process. This part of cognitive strategies is important for reaching higher order thinking, but is very difficult for instructors to assess because tests or exams are about a particular subject matter [77]. A student either does well or does not, but this could be due to poor studying habits and cognitive planning or the student merely did not understand the material. The other part of cognitive strategies, which is easy to assess, is creative thinking and student originality of thought. The desire is for students to not only be able to problem solve, but also to be capable of generating their own problem and then solving it. Take the flight plan example problem above. In order to minimize the vast number of solutions, a professor or instructor may specify a class of aircraft to be used and other constraints, such as the number of stops allowed, number of pilots on the plane etc. If a student generates the problem statement, they are going to do so based on the rules they have learned and their own personal ways of thinking [122]. Some students may include take off and landing weights, while others may only provide the fuel burn rate and specify the aircraft. Some will specify weather conditions, wind patterns, and give a date and cruising altitudes, while others may only recommend using the standard atmospheric table. When students are given the opportunity to generate their own problems within a certain context that has been taught, their cognitive ability (or lack thereof) will stand out. Gagne addressed the affective and psychomotor domains as part of his integrated taxonomy. The fourth classification he calls “attitudes”. Gagne defined attitudes as acquired internal states that influence the choice of personal action [91]. This action could be towards a person, thing, or event. Gagne’s definition of attitudes incorporates the first two levels of Krathwohl’s affective taxonomy, receiving and responding (discussed in detail below in the Affective Domain Learning Section). These two levels highlight information and attitude components of attitude formation [77]. A person must have learned something about a matter before they can have an attitude or feeling that influences their behavior. A pilot choosing to take a detour-flying route in inclement weather is persuaded by his knowledge of the aircraft, training situations, and personal safety concerns. The one area that Gagne does not address with his definition of attitudes is emotions. It’s understood that people will have a certain attitude based on the information they have and that can result in certain behavior, but how to instruct and train a pilot to remain calm and exhibit leadership traits that provide reassurance to passengers and crew on board is not defined. Varying stages of emotional control are lacking for this taxonomy to be useful by itself in training evaluations. The fifth type of learning outcome in Gagne’s taxonomy, “motor skills” corresponds to the psychomotor domain. Motor skills relates to the “precise, smooth and accurately timed execution of performances involving use of the muscles [122].” Gagne recognized that motor skills required to accomplish complex tasks such as landing an aircraft also required intellectual and cognitive skills. This is why it was important for him to develop a taxonomy inclusive of all three learning domains. As with the attitude classification, Gagne explained that there were increasing levels of motor skills required to complete a task, but did not propose specific sub-categories.

The learning taxonomy is only part of Gagne’s proposal for instructional theory. He also iterated that learning conditions (internal and external) are necessary to achieve desired learning outcomes. He stressed the need for instructors to design for learning outcomes rather than designing based on the learning process. He provided nine specific events of instructions, which serve as a guideline for designing instruction. The learning conditions are included in Table 2.14. For more information about the events of discussion the reader is referred to [91].

Table 2.14: Gagne's Taxonomy of Learning Outcomes and Conditions [77]

Primary Classification	**Definition**	**Learning Conditions**
1. Verbal Information	Stating previously learned material such as facts, concepts, principles and procedures	1. Draw attention to important features. 2. Encourage chunking of information. 3. Provide a meaningful context for encoding. 4. Provide cues to stimulate recall and transfer.
2. Intellectual Skills		
Discrimination	Distinguishing objects, features or symbols	1. Draw attention to distinctive features. 2. Stay within the limits of the capacity of working memory. 3. Stimulate the recall of previously learned component skills. 4. Use verbal cues to help order and combine the component skills. 5. Schedule occasions for distributed practice and review. 6. Use a variety of contexts to promote transfer.
Concrete Concepts	Identifying classes of concrete objects, features, or events	
Defined Concepts	Classifying new examples of events or ideas by their definition	
Rules	Applying a single relationship to solve a class of problems	
Higher Order Rules	Applying a new combination of rules to solve a complex problem	
3. Cognitive Strategies	Employing personal ways to guide learning, thinking, acting, and feeling	1. Describe or demonstrate the strategy 2. Provide opportunities to practice the strategy. 3. Provide feedback as to the creativity or originality of the strategy.
4. Attitudes	Choosing personal actions based on internal states of understanding and feeling	1. Associate the attitude with success. 2. Associate the attitude with an admired human model. 3. Arrange for personal action associated with the attitude. 4. Give feedback for successful performance.
5. Motor Skills	Executing performances involving the use of muscles	1. Use verbal guidance for executive routine 2. Arrange repeated practice. 3. Give immediate feedback. 4. Encourage mental as well as physical practice

2.4.4 Component Display Theory (Merrill Taxonomy)

The component display theory (CDT) evolved from attempts by Professor Merrill to clarify the Gagne theory for his students [123]. It is founded on the same assumptions as Gagne's work, but only deals with the cognitive learning domain. It parallels with the first three levels from Gagne: verbal information, intellectual skills, and cognitive strategies but provides more details. It is a two-dimensional classification system.

Varying levels of performance forms one dimension, and the type of content forms the second dimension, creating a performance-content matrix as shown in Figure 2.7. The content dimension is broken into four categories: facts, concepts, procedures, and principles. The levels of performance are: remember, use, and find. CDT also defines a set of primary and secondary presentation forms. The primary presentation forms include: rules, examples, recall, and practice. Secondary presentation forms include: prerequisites, objectives, helps, mnemonics, and feedback. CDT specifies that instruction is most effective when all necessary primary and secondary forms are given. Therefore, a complete lesson would consist of a learning objective followed by some combination of rules, examples, recall, practice, feedback, helps, and mnemonics appropriate to the subject matter and learning task. The theory suggests that for a given objective and learner, there is a unique combination of presentation forms that results in the most effective and efficient acquisition of skills and knowledge available. CDT is designed, primarily, for use by groups of learners. The instructional goals for the students are determined by identifying the elements of the matrix that best meet the desired learning outcome. The simplistic nature of the matrix is adaptable to many different training system designs and evaluations, but for purposes of this research it does not add much value beyond Gagne’s taxonomy. Merrill later advocated a four step problem-centered approach to training that focused more on students demonstrating learned skills [124]. The student must activate prior experiences that are triggered by the problem, demonstrate their skills related to solving the problem, apply those skills to the problem solution, and finally integrate their skills into the real-world application. A taxonomy that merges learning and real world application is a must for training effectiveness.

		Types of Content			
		Fact	**Concept**	**Procedure**	**Principle**
Level of Performance	**Remembering**				
	Using				
	Finding				

Figure 2.7: Performance-Content Matrix [123]

2.4.5 MIL-HDBK-29612-2A Knowledge Taxonomy

In addition to the psychomotor skills learning taxonomy, the DoD also provides a recommended hierarchy for the cognitive learning domain [103]. The five level taxonomy shown in Table 2.15, has characteristics of Gagne's, Merrill's, and Bloom's taxonomies. The first three levels consist of fact learning, rule learning, and following procedures sequentially. They are similar to Gagne's verbal and intellectual skill requirements and Merrrill's remembering and using facts, concepts, and procedures. The fourth level, discrimination learning, identifies a student's ability to match basic knowledge with concepts and applications. The last level is called problem solving and really captures the top three levels from Bloom's taxonomy (analysis, synthesis, evaluation) as well as Gagne's cognitive strategies. The DoD handbook recommended taxonomy for knowledge does not capture innovation or the creation of new ideas based on information the student has learned. However, the final step in the recommended psychomotor taxonomy, origination, involves the creation of complex physical and mental skills to accommodate a new situation. The ability to construct new thoughts is covered when all three DoD handbook taxonomies are used jointly.

Table 2.15: MIL-HDBK-29612-2A Taxonomy of Knowledge Outcomes [103]

Primary Classification	**Key Words (Verbs)**
1. Fact Learning: Verbal or symbolic information (e.g., names, formulas, facts, etc.).	Advise, answer, brief, calculate, define, elaborate, express, identify, inform, instruct, list, name, read, recall, recommend, recount, specify, state, tell
2. Rule Learning: Using two or more facts in a manner that provides regularity of behavior in an infinite variation of situations.	Appraise, compile, compose, compute, encrypt, estimate, evaluate, format, forward, measure, outline, route
3. Procedure Learning: Performing step-by-step actions in the proper sequence.	Check, condense, edit, delete, implement, initiate, pause, resume, set up, start, stop
4. Discrimination Learning: Grouping similar and dissimilar items according to their distinct characteristics.	Allocate, arrange, assign, categorize, classify, collate, correlate, cross-check, discriminated, distribute, eliminate, extract, group, match, organize, rank, realign, schedule, select sort, task
5. Problem Solving: Synthesizing lower levels of knowledge for the resolution of problems.	Analyze, apply, combine, convert, criticize, defend, derive, determine, discover, effect, extend, generalize, generate, illustrate, investigate, modify, predict, resolve, search, solve, synthesize, use

2.4.6 SOLO Taxonomy

The Structure of Observed Learning Outcome (SOLO) taxonomy is another cognitive domain learning scheme that consists of levels similar to Bloom's taxonomy but have a different qualitative approach and details [81]. SOLO describes levels of increasing complexity in student's understanding of any subject, according to its authors [125]. It is used to aid both trainers and learners in understanding the learning process [102]. In comparison with Bloom's taxonomy, which has been very useful because it has extended learning from mere rote learning to more complex cognitive abilities such as analyzing and creating, SOLO adds an additional assertion that the learner can recognize and measure their individual learning stage in a clearer and more precise manner. SOLO is comprised of five levels of increasing order of understanding: pre-structural, uni-structural, multi-structural, relational, and extended abstract, as shown in Figure 2.8 [125]. The heavy emphasis towards the instructor and student being able to clearly identify the learning stage is seen in the descriptions of the levels. In the first learning level, pre-structural, the student may be able to recite answers, but has no understanding of the context and lesson intent. The student’s explanations are simple and unrelated to the subject matter. As the student begins to understand the subject matter context, they move into the uni-structural stage of learning. Here the trainee can correctly relate at least one basic concept to some relevant aspect being taught. As the student or trainee begins to demonstrate an understanding of several instructional concepts they then move into the multi-structural phase. At this point, the learner is still treating ideas independently and cannot see how the lessons all come together. Once the student can integrate different aspects of training into a coherent whole, the student is considered as having mastered the complexity of the subject and is in the relational classification level. At this point, the student is normally rated as having an adequate understanding of the subject. The final level, extended abstract, is similar to Bloom’s levels of synthesis and evaluation because the student can now take what they have learned and relate it to a new or different topic area. The student can create new ideas based on having mastered the subject matter.

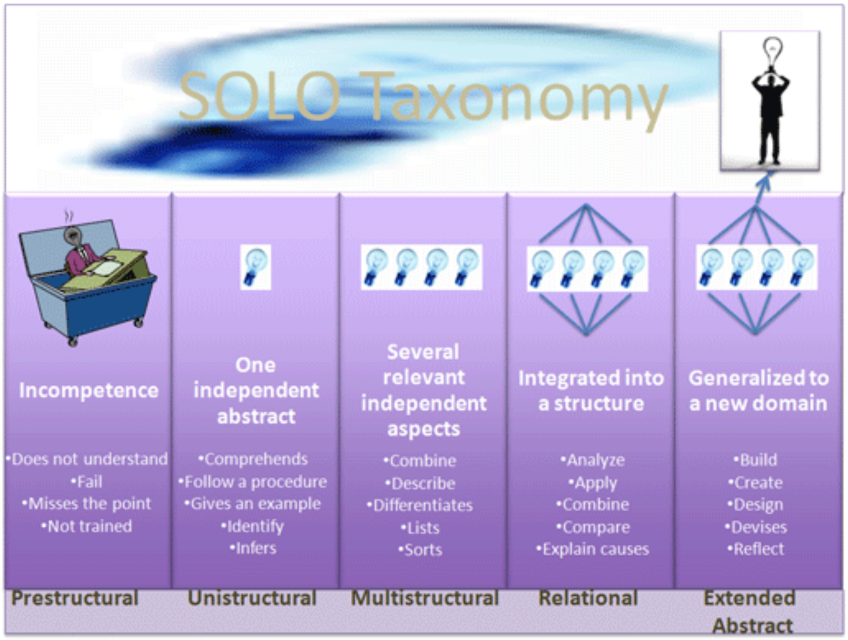


Figure 2.8: SOLO Taxonomy [102]

2.4.7 Cognitive Taxonomies Summary

Presented above are various theories and models of cognitive learning and instruction. Some theories address specific learning tasks while others focus on identifying where students are in the learning process. Table 2.16 summarizes each of the taxonomies, and shows how they compare to Bloom’s taxonomy on the far left since it is the most popular. One theory is not necessarily better than another, but each must be applied based on the learning tasks and objectives fitting a specific training program. When evaluating an overall training program, one should look for a balance between lower and higher order thinking. If using Bloom’s Taxonomy, the lessons should have maybe a third in knowledge and comprehension, a third in application and analysis, and a third in synthesis and evaluation. The correct breakdown may not be thirds but there should certainly be some variation among the levels. If the training program is heavily based in lower order thinking, it is as if the student is being trained as a robot. Military students need to be capable of taking what they have learned and applying it to new situations. It is cost and time prohibitive during training to demonstrate every situation and all environmental conditions that an aircrew member may encounter. However, the aircrew should demonstrate the ability to take what they have learned in training classrooms, simulations, and/or practice flights and apply it to hypothetical mission scenarios to quickly create a plan of action. When any student has this ability, there is greater confidence in accomplishing the mission objective in comparison to training military personnel who are dependent upon a superior for all instructions before they can carry out a task. During time critical situations when a mission may not go exactly as planned and communications need to be kept at a minimum, the aircrew must execute with higher order thinking skills to adapt quickly, and at a high level of affectivity.

Table 2.16: Classification Comparison of Cognitive Learning Objectives [126]

Bloom	**Anderson**	**Merrill**	**Gagne**	**MIL-HDBK-29612-2A**	**SOLO**
Knowledge	Remember	Remember fact verbatim	Verbal Information	Fact Learning	Pre-Structural
Comprehension	Understand	Remember fact by paraphrasing		Rule Learning	Uni-Structural
					Multi-Structural
Application	Apply	Use principles	Intellectual Skills	Procedural Learning	Relational
				Discrimination Learning	
Analysis	Analyze	Find principles	Cognitive Strategy	Problem Solving	
Synthesis	Evaluate				Extended Abstract
Evaluation	Create				

2.5 Affectivity

Affective learning deals with feelings, emotions, acceptance and rejection, qualities of character, and conscience [127]. The objectives are articulated as interests, attitudes, appreciations, values, and emotional sets or biases. This domain deals with how people react emotionally; people doing things because it makes them feel good rather than because the law says so. People choose to participate in certain activities over others because of the influence that these internal states have. The affective domain helps to explain why an individual who knows perfectly well what to do, such as not speeding in a school zone, may choose to break the law and risk getting into an accident or injuring a child because they are worried about being late to work [77]. Affectivity is many times overlooked, or regarded as not important, but in training it is just as important as teaching psychomotor and cognitive skills. When performing skills that involve safety, the trainee’s affectivity is vital. Unlike a lot of other tasks, it is often easier to do something in an unsafe manner, rather than perform it the safe way. For example, it is faster and easier to immediately start operating a piece of equipment rather than to perform the required safety checks beforehand. Teaching someone to act and communicate in a safe manner requires that they not only gain the required knowledge and skills, but that they also change their attitude towards the job they are performing. Otherwise, they will know how to act, but not act appropriately because such things as time and outside pressure. convince them to do it the fast and easy way. Everyone performs calculated risks, which in reality are unsafe acts to various degrees. Someone may never dare to use gasoline to start a barbecue when they are out of lighter fluid, but many people will cross the street outside of the crosswalk when traffic is not busy. Addressing and accounting for the affective abilities of trainees is imperative, but not as easy as categorizing behavioral and cognitive learning. A common problem with affective learning is perceived vagueness or ambiguity about the meaning of the term, and the difficulty to observe and/or measure different stages [83]. It is also intertwined with cognitive and psychomotor abilities, develops slowly, and is personal. This has led to many research evaluations that ignore affective learning or simply assume that resultant emotions will occur naturally or as appropriate. There are other researchers who have experimented and developed independent taxonomies to aid instructors in teaching, observing, and benchmark changes in affective behaviors. In aircrew training, affective learning is primarily addressed as part of crew resource management (CRM) training. The taxonomies presented below provide a broader perspective and application for assessing the affective learning domain.

2.5.1 Krathwohl's Affectivity Taxonomy

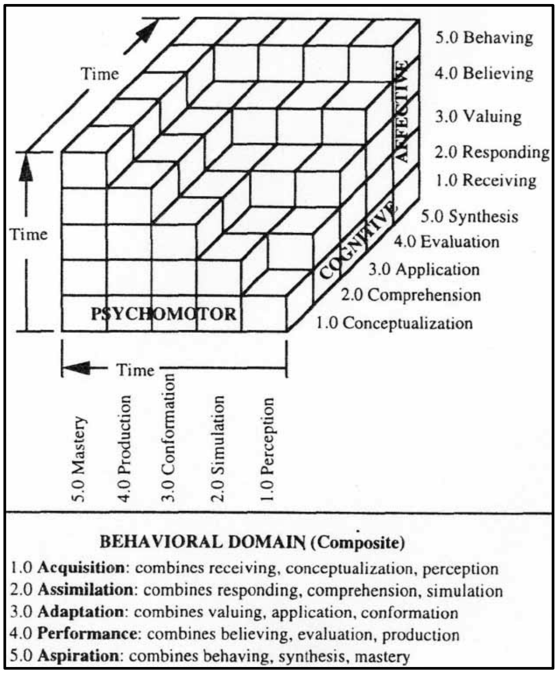
Upon the release of Bloom's cognitive learning taxonomy, there was a recognized need amongst the committee that a similar taxonomy classifying affective learning was necessary [127]. Krathwohl led this committee, which consisted of two committee members who worked on developing the cognitive taxonomy. They published what is commonly referred to as Bloom's Taxonomy for the affective domain trying to follow the same structure used in the cognitive system of simple to more complex abilities. They expected to receive criticism, and they have, but this taxonomy is still being used sixty years later. This affective taxonomy is broken into five categories: receiving, responding, valuing, organization, and characterization by a value, as shown in Table 2.17. The first classification, receiving, is to prepare and focus the student for instruction so that the student is attentive, at least on a semiconscious level. Receiving is divided into three subcategories representing different levels of attention to the teaching experience. They vary from an extremely passive role on part of the learner, referred to as awareness, to the student directing his full attention towards the instructor, called controlled or selected attention. The next level is responding. This is when the student goes from selectively attending to actively participating. This level does not indicate that the student values the instruction, but he or she has decided to commit to the teacher and engages in the subject matter. This level varies from someone complying with rules and regulations but not accepting the necessity in doing so, to feeling good about actually participating. An example is someone who uses their turn signal to change driving lanes because it's the law, versus someone reading a thesis for recreation because the subject matter is of interest. The third category, valuing, represents the point where a person is not motivated by a desire to be obedient or compliance, but is guided by one's own commitment to the individually perceived value when performing a certain behavior. As a learner begins to value different situations, a need arises to organize and determine the proper relationship, and preferences amongst those values. This is what takes place at the organization level. The student is building a value system. In the final category of affective learning, the value system developed controls the behavior of the student habitually. Evoking the behavior does not stimulate any reactions unless someone threatens or challenges the individual. The values and behavior now coincide with the students' routine behavior and views of the world.

Table 2.17: Krathwohl's Taxonomy of Affective Outcomes [127]

Primary Classification	Sub-categories
1. Receiving: Becoming sensitized to or willing to receive certain information	1.1 Awareness 1.2 Willingness to receive 1.3 Controlled or selected attention
2. Responding: Becoming involved or doing something.	2.1 Acquiescence in responding 2.2 Willingness to respond 2.3 Satisfaction in response
3. Valuing: Displaying a commitment to something because of its inherent worth.	3.1 Acceptance of a value 3.2 Preference for a value 3.3 Commitment
4. Organization: Organizing a set of values and determining their relationships, including which should dominate.	4.1 Conceptualization of a value 4.2 Organization of a value system
5. Characterization by value: Integrating values into a total philosophy and acting consistently in accord with that philosophy.	5.1 Generalized Set 5.2 Characterization

2.5.2 Hauenstein's Affective Taxonomy

In 1998, Dean Hauenstein, released a book that attempted to update taxonomies in all three learning domains, and proposed a composite taxonomy that combined the three into one. He used Bloom's cognitive, Krathwohl's affective, and Simpsons' psychomotor taxonomies as a basis for his proposed classification systems. His goal in revising these taxonomies was to help with research, assessments, and curriculum planning while maintaining an emphasis on student learning as a whole person [128]. He believed all three learning taxonomies were necessary and thus should always be considered together, he referred to the unified taxonomy as the "behavioral domain" [129]. The taxonomies Hauenstein created are similar to the originals. The categories of Hauenstein's affective domain are: receiving, responding, valuing, believing, and behaving. The primary and subcategories are organized similar to Krathwohl's break down. There are three subcategories for the first three classification levels and two each for believing and behaving. The top two levels, believing and behaving, replace organization and characterization by value in Krathwohl's taxonomy. The composite behavioral domain and instructional system includes five groupings, as shown in Figure 2.9. The categories are represented as a truncated cube that characterizes how a student thinks, feels, and acts. A student "behaves or acts in relation to what one knows, feels, and can do [129]." In the first level, acquisition, the student acquires basic concepts and ideas. The student then assimilates these ideas based on previous knowledge and experience. During level three, the learners adapt their skills and amend knowledge to solve problems or practice implementing the new ideas or actions. Once the student can analyze, qualify, evaluate, and integrate the new knowledge, values, and beliefs, and effectively use the new skills, then performance can be assessed. The final level of aspiration is obtained once the student operates at a high level of expertise, such that their actions are habitual in terms of knowledge, skill set, and affective qualities. At this point the student demonstrates creativity, wisdom, and sensitivity in their decisive actions.



BEHAVIORAL DOMAIN (Composite)

1.0 **Acquisition**: combines receiving, conceptualization, perception
2.0 **Assimilation**: combines responding, comprehension, simulation
3.0 **Adaptation**: combines valuing, application, conformation
4.0 **Performance**: combines believing, evaluation, production
5.0 **Aspiration**: combines behaving, synthesis, mastery

Figure 2.9: Hauenstein Conceptual Framework [83]

2.5.3 Scientific Attitude Taxonomy

In the area of science education, specifically physics, chemistry, and biology, there have been attempts to measure learner attitudes to investigate why students tend to desert these subjects in high school and college [130]. During the nineteen sixties and seventies a curriculum reform movement occurred and enrollment dropped for secondary pupils in England, as well as for college classes in the USA [131]. In his attempt to measure attitudes in science education, Norman Reid suggested a taxonomy for scientific attitudes. He believes that attitudes are generally important because they can influence subsequent behavior. Attitudes developed at school related to science may well be retained into adulthood, and play a major role in behaviors. Negative attitudes can have potentially harmful effects for people personally and socially and can also affect national issues. His research has focused on physics pupils in Scotland, but the taxonomy is relatable to any subject matter, including that of aerospace military training. The five categories include: directed curiosity, logical methodology, creative ingenuity, objectivity, and integrity. They are defined in Table 2.18. This taxonomy was developed in consultation with twelve scientists from a wide range of disciplines. It covers valuing knowledge but also includes cognitive abilities. The first two levels, directed curiosity and logical methodology, involve student interest and self-commitment to the subject matter. This evaluation begins after the student has decided to pay attention to training or instruction. Level three, creative ingenuity, is where the student begins to truly understand and create new ideas, but notice that the student is doing this not because the instructor requires it, but the student wants to advance their own knowledge. The top two levels, objectivity and integrity, occur when the student can assess their ideas and compare them to others without bias. This taxonomy follows a natural desired learning process. It evaluates students not on what they are forced to learn and memorize, but by how well they seek to understand and make sense of what is being taught.

The fact that the top level is evaluating a student's ability to cooperate and communicate with others about their ideas and work, which may be conflicting to someone else's, is very important in any environment where teamwork is necessary. This is an enabler for effective communication in a military combat situation [21]. There is an increasing awareness that social climate characteristics between people in a dialogue can greatly enhance or degrade communications patterns and that has a strong influence on the efficiency of system performance [132]. Because of the safety risks, researchers involved with aircrew training in the aircraft flight deck have and continue to study this issue in depth [133]. From the analysis of several accidents and near accidents where the copilot or flight engineer, having seen or suspected that the pilot was in error, either failed to call it to the pilot's attention or did so in such a hesitant and polite manner that the error was not corrected. Similar incidents can be imagined between operators such as a surgeon and nurse, pilot and air traffic controller, or the corporate executive and an administrative assistant. In the aerospace community, the establishment of training programs specifically for crew resource management was designed with emphasis on two-way information exchange. Effective cockpit communications have resulted in less error-prone performance in flight simulators and emergency landings [134], [135]. Although Reid developed this taxonomy of attitudes for classifying stages of student growth in science education, it extends well beyond that audience and fits a need in assessing developmental stages of the affective domain for aircraft crew training.

Table 2.18: Scientific Attitude Taxonomy [131]

Primary Classification	Definition
1. Directed Curiosity	A desire to know understand, solve problems and obtain answers
2. Logical Methodology	A knowledge of, and willingness to pursue, a logical and cyclical series of operations in satisfying directed curiosity. This relates to the raising and testing of hypotheses.
3. Creative Ingenuity	A willingness to build mental constructs or models, set up realistic hypothesis, design suitable experimental situations, see beyond set ideas in order to grasp new or create new ideas.
4. Objectivity	A willingness to assess error, control variables, view results objectively, distinguish, description from explanation.
5. Integrity	A willingness to avoid bias, consider details that may appear contradictory, consider implications of one's work, cooperate and communicate with others, respect instruments and materials.

2.5.4 MIL-HDBK-29612-2A Taxonomy for Attitude Learning Outcomes

As with the cognitive and psychomotor learning domains, the DoD handbook for ISD/SAT includes a taxonomy for affective skills, which they refer to as attitudes [136]. This taxonomy is similar to Bloom's affective taxonomy, but is defined from the viewpoint of military job analyses. The attitude learning levels follow the same progression of simple to complex, and are also presented with a set of standardized verbs to aid instructors, as well as students, in understanding the precise meaning of each learning objective, shown in Table 2.19. Notice how the last two levels differ in focus from Bloom's affective taxonomy. Level five is competence, and refers to the student's ability to make sound decisions in all types of situations, whether it is normal, abnormal, or emergency conditions. The final step is innovation. Here the definition language reads almost exact as level four, except instead of the trainees *using* prioritized strategies, they are *generating* or creating new strategies and tactics that they then use appropriately. A sense of organization and characterization of values from Bloom's affective taxonomy is necessary to accomplish these last two skill sets, but in this military handbook, the focus is taken off of the valuing process and instead refers to the visible results after one has mastered valuing.

Table 2.19: MIL-HDBK-29612-2A Taxonomy of Affective Outcomes [136]

Primary Classification	**Key Words (verbs)**
1. Receiving (Perception; Situation Awareness) **Definition:** Demonstrating mental preparedness to perceive the normal, abnormal, and emergency condition cues associated with the performance of an operational procedure.	Attend closely, Listen, Listen attentively, Monitor, Observe, Perceive, Recognize, Reconnoiter, Show awareness, Show sensitivity, Wait
2. Responding (Interpreting) **Definition:** Demonstrating mental preparedness to encode operational cues as indicators of normal, abnormal, and emergency conditions associated with the performance of an operational procedure.	Accomplish, Achieve, Acknowledge, Announce' Ask, Communicate, Complete, Complete assignment, Comply, Demonstrate, Describe, Encode, Execute, Give, Indicate, Interpret, Notify, Obey rules, React, Report, Request, Respond, Resume, Show
3. Valuing (Judgment) **Definition:** Demonstrating the ability to judge the worth or quality of normal, abnormal, and emergency cues associated with the performance of an operational procedure.	Alert, Appreciate, Approve, Assess, Authenticate, Belief, Cancel, Choose, Judge, Justify, Prioritize, Propose, Qualify, Reassess, Review, Share, Study, Validate, Verify
4. Competence (Application of resource management strategies and tactics.) **Definition:** Demonstrating the mental preparedness to make decisions using prioritized strategies and tactics in response to normal, abnormal, and emergency condition cues associated with the performance of operational procedures.	Allow, Alter, Assume, Command, Coordinate, Enforce, Ensure, Influence, Prescribe, Serve
5. Innovation (Generation of new resource management strategies and tactics) **Definition:** Demonstrating the mental preparedness to make decisions by generating the results expected upon completion of prioritized strategies or tactics in response to normal, abnormal, and emergency cues associated with performance of an operational procedure, and generating prioritized strategies and tactics in response to abnormal or emergency cues.	Conceive, Conjecture, Develop, Devise, Formulate, Imagine, Innovate

2.5.5 Affective Taxonomy Summary

Taxonomies describing the affective learning domain process do not seem to vary much from Krathwohl's suggested hierarchy, as shown in Table 2.20. Although presented herein as only a sample of published learning hierarchies, this sample is representative of a thorough literature survey [76]. Krathwohl's taxonomy does well for describing the early process in affective learning. Levels one through three capture the transitions that students go through in their attitude towards learning: 1) learning because it is required, 2) valuing the instruction, and 3) choosing to act in accordance with best practices. The higher two levels focus on the student prioritizing and acting habitually based on their new value system. This is where Reid's scientific attitude taxonomy is more appropriate in demonstrating increased complexity of affective learning. When a student is able to perform with objectivity and integrity in their communications between peers, superiors, and subordinates then the decision making process reflects not only their values but also what is best for all involved. At this point the student has reached a level where they do not react emotionally, but in a controlled and calculated manner. This is what is so important in safety related issues. In military training the students should display characteristics of high affective learning to help them prepare for worse case emergency combat situations where an improper error assessment or inadequate communication can result in death.

Table 2.20: Classification Comparison of Affective Learning Objectives

Levels	**Krathwohl, Bloom, Masia**	**Hauenstein**	**Reid Scientific Attitudes**	**MIL-HDBK-29612-2A**
1	Receiving	Receiving	Directed Curiosity	Receiving
2	Responding	Responding	Logical Methodology	Responding
3	Valuing	Valuing	Creative Ingenuity	Valuing (Judgment)
4	Organization and Conceptualization	Believing	Objectivity	Competence
5	Characterization by Value or Concepts	Behaving	Integrity	Innovation

2.6 Constructivism

Constructivism is the most recently popular position among the education, instructional, and training communities [79]. Notice the use of the word *position* rather than theory. Proponents of constructivism certainly consider this to be a theory, but there are just as many researchers and scholars who disagree. It is beyond the purpose and usefulness of this thesis to debate whether or not constructivism is a learning theory; however, constructivism contains assumptions and ideas about learning that can benefit the design and evaluation of training. It is not considered 'new' because aspects of this theory stem from historical research referred to as discovery learning, generative learning, and situated cognition, to name a few [77]. Constructivism is made up of extremely radical and conservative viewpoints. Instead of explaining in detail the varying philosophies of constructivists, what is presented here are the common assumptions and how they compare to behaviorism and cognitive learning theories from the perspective of instructional design and training.

Constructivist's theory of knowledge is distinctively different from the objectivist tradition. Objectivism is the view that knowledge of the world is a direct result of an individual's experience within the world [77]. It holds as truth that knowledge exists and the world has a certain reality whether or not a person has learned or experienced it. In this case, the process of learning consists of transferring knowledge from outside to within the student. This is the philosophy behind behavioral and cognitive learning. In contrast, constructivism views learning as knowledge construction and considers knowledge as individually constructed by learners, based on their interpretations of experiences in the world. The learners' construed knowledge does not necessarily correspond to external reality. The newly acquired information does not have to be representative of the real world to be useful and viable. This is one of the most argued assumptions about constructivism. It makes sense that a student would form their own ideas about any subject as they go through intentional and unplanned learning experiences. A popular example is the research of children explaining the earth's relationship to the sun [137]. Children typically believe that the earth is flat, and the sun moves across the sky during the day. In constructivist's view, these children have constructed a perfectly viable model of the earth and sun, as it accounts for their own experience in the world. Of course, a person's understanding of the relationship between the sun and earth is corrected in science classes later in elementary school. Most constructivists agree that a limit must exist between someone's perceived reality and how it corresponds to true reality based on the context and subject matter. Although there may be debate between objectivism and individually construing knowledge, the challenging of how perceptions and experiences will affect an individual's beliefs and actions is something that has to be addressed in instructional design.

Based on the principle, through individual experiences and internal knowledge everyone constructs their own viewpoint of the world, constructivism emphasizes the need to prepare students to solve problems in ambiguous situations [79]. This is accomplished by ensuring that the learning environment is a realistic setting. The philosophy of learning in a realistic setting is not original to constructivism; it is very similar to the context of situated cognition theory [138]. What differentiates constructivism is the suggestion among some that problems should not be simplified for novice learners early in the learning process, but only presented in their full complexity [79]. The belief is that by simplifying the problem the student will generate a false impression that the problem is easy and be unprepared for facing the real world event. Because it is many times impractical and cost-prohibitive for learning or training to occur within the actual environment, alternative activities and conditions are used in classrooms, simulators and models. The danger in using these other instructional means is that the students may learn what is required to successfully function in this environment, but not understand or be able to relate to the real world example. An example is with the use of “instructional” computer games. Research findings have shown that students may be actively engaged and enthusiastic but learn nothing more than the rules of the game. I’ve personally witnessed this from watching my young niece play a simple math addition game on my apple iPad. She was getting many of the answers wrong at first (I could tell because of the noise made between a correct and errant response). After a while I heard fewer incorrect noise responses, and as I looked to see that she was doing better I noticed a pattern in the questions and choice of responses. I turned off the game and asked her the same addition problems that were asked during the game and she answered a majority of them wrong. I restarted the game and she correctly picked the answers. She did not learn addition; she had learned to recognize the pattern. This was not the learning objective I had intended for her to acquire. This same situation can happen in training design. The use of more efficient instructional methods must not prevent or reduce the effectiveness of learning and training goals.

2.7 Learning Theory Summary

Presented within this section are several taxonomies, which researchers have proposed will aid instructors in developing effective learning objectives and outcomes. Most of the taxonomies are catered towards one of the three learning domains: psychomotor, cognitive, and affective. There are others such as the taxonomies proposed by Gagne and Ferris, which address all three domains with varying levels of detail. One taxonomy is not necessarily better than another. As early as 1956, psychologist Benjamin Bloom divided what people know and how people learn into separate domains of learning [139]. The cognitive domain focuses on knowledge and the mind. It consists of three practical instructional levels including facts, understanding, and application. As discussed previously, the basic knowledge or rote learning level of the cognitive domain uses verbs such as define, identify, and list. The understanding level adds verbs that include describe, compare, and contrast. The application level uses concepts to synthesize and form new ideas and includes verbs such as explain, apply, analyze, evaluate, and synthesize. Teaching in this domain is typically accomplished by lecture or classroom presentation and will be presented with details in the following section. The psychomotor domain is tactile based and more physical in its outcomes. It is heavily immersed in the student demonstrating actions and producing tangible results. The instructional levels include imitation, practice, and habit in this domain. At the level of imitation, demonstration occurs under the close inspection of the instructor. Practice builds expertise that may be conducted autonomously at the discretion of the teacher. The habit level is reached when the student performs the skill without instructor intervention and without centering all the student's attention to performing the task. The skill has become a habit once it can be conducted quickly, correctly, and while being observant of the surrounding environment. The final province of educational psychology is the affective domain based upon aspects of learning that may be labeled as beliefs, values, or emotions. The three levels in this domain are awareness, distinction, and integration. Action verbs such as display, exhibit, and accept are most commonly used. The first two levels are cognitive (knowledge-based). The remainder of the levels is more affective in nature.

Choosing a taxonomy to use in instructional design or evaluation is based on a synthesis of current thought regarding the forms of knowledge, types of learning activities, importance of each domain, and the effects of the learner's style[4] [82]. A broad variety of learning outcomes should be assessed in accordance with the learning or training goals. The consideration of exact subject matter is critical when selecting taxonomies for training evaluation. There is some generality in learning skills over domains, but having evidence that a pilot can physically fly an aircraft in nominal and off-nominal conditions is not the best indicator for crew resource management, awareness, and communication. There are also training programs that rate importance of speed higher than quality in decision-making; this may result in selecting a more detailed psychomotor classification system and simpler cognitive and affective taxonomies. The

---

[4] Evaluations for the effects of the learner's style are not part of this thesis. These variables are important, and are briefly discussed in paragraph **Error! Reference source not found.**.

[5] Psychometrics is the science and enterprise of using tests to 'measure' psychological traits, abilities, and

context being used for a particular training environment is critical to deciding which taxonomies are best for evaluating training effectiveness. Selecting a specific taxonomy is context dependent, but in general learning taxonomies offer a tool for quantifying training effectiveness based on the LOs used in the design of the training program. Learning taxonomies also classify knowledge, skill, and attitude competency levels used in a training system design. This provides partial answers to RQ2.1 and RQ3. The taxonomies in all three learning domains are not only useful in determining the level of competency for each LO, but also for checking that the LOs are distributed across several levels rather than dominated by lower abilities such as rote memory [63].

2.7.1 Learning Variables Not Considered

In the science of psychology and education there are numerous variables that are not being addressed herein; and it is not because they are deemed any less important in learning and training. This research attempts to develop MPEET, so there are no exact before and after measurements or comparisons that can be made during this evaluation but it is recognized that many other variables influence training effectiveness [65]. It is assumed that to the greatest extent possible the instructional design process has considered the following variables when developing the training program.

Social Dimension. There is another learning dimension that was not discussed above and that is social interactions. A stand-alone social hierarchical system can be derived from Gorman’s research on improving teacher-student and student-student interactions [120]. On the first days of a course, Gorman defined a class as an “aggregate” or random collection of people. Later in the course he describes the same class as having developed into a fully functioning, cooperating, and working group. This taxonomy represents the social learning dimension and helps to identify stages of student progress from individuals to collaborators. In today's work environment teamwork is important and improvement of social skills is required for learners to excel in all other learning domains. Predicting the effectiveness of training in the social dimension, however, is not something that is readily quantifiable due to the subjectivity involved. It is a characteristic that can be tested during and after training, but not from an evaluation of the curriculum design because of the unknown variables about each student such as level of prior knowledge, personality variables, strategies for learning, and demographics discussed below.

Level of prior knowledge. Psychologists can all agree that the level of prior knowledge that the student brings to a training event will vary amongst any class [88]. Experienced learners can deal with larger steps of instruction and more complex learning environments [21]. Novices, on the other hand, require simplification of complex contexts so they don't experience information overload while learning. For example, a classroom full of newly assigned C-130J pilots will have a vast array of flying experience. It may range from some who have never flown a cargo type aircraft to others who have. The only thing the instructor is guaranteed is that the pilots have graduated and completed the basic training per their respective services requirements (ACC or AFSOC). It is part of the instructional design team's responsibility to ensure the material covers all training objectives and begins at a level consistent with expectations. If there is a case where a student's background information is below standard, then the instructor may have to provide remediation.

Cognitive processing variables. Students process information in different ways. For example, some learners prefer to take their learning in a series of logical steps from beginning to end, building to a conclusion (serial learners); others prefer to begin their learning with an overview, the "big picture," and then fill in with the details later (holistic learners) [140]. A strategy where the instructor provides some type of initial overview to give the learners a sense of orientation and to set expectations is normally useful for cognitive processing [88].

Personality Variables. The individual differences or personalities of people have an impact on how they learn, their motivation for learning, and their preferences for receiving information. In terms of how students learn, some learners can look at a whole picture and isolate or abstract individual pieces with ease (field-independent learners); others are strongly influenced by the whole picture and do most of their interpreting of new information in the context in which it occurs (field-sensitive learners) [141]. Abstraction is easier for the former type of learners, and integration is probably easier for the latter. An instructor can include both types of tasks to benefit those students when their preference is being matched, and help them learn to complete assignments that do not match their preferences. Motivation to learn is another personality variable that is sometimes an indicator of how well a trainee will successfully complete a program [65]. When a student has confidence in their ability prior to training, research has shown, they have a better learning outcome.

Some learners are reward based and driven by the competition of "getting a good grade" rather than obtaining a full understanding of the material. Students driven by competition are less likely to remember what they have learned even after a short period of time such as a couple months after a class has ended [142]. In comparison, students who are more interested in understanding what they learn show that they can remember and apply the same material for a much longer time.

The fact that a student may be impulsive or reflective is another personality trait that impacts training effectiveness. An impulsive individual responds quickly, while a reflective person is more thoughtful [143]. This dichotomy is sometimes interpreted as a learner being more willing to take risk versus a more risk adverse student. This can have an influence on students' responsiveness in class, on their test-taking behavior, and even on their choice of assignments [88].

Strategies for learning. Theorists have proposed the concept of learning strategies as an area of individual difference [144]. Learning strategies can be defined as behaviors and thoughts in which a learner engages, and which are intended to influence the learners encoding process. The goal of any learning strategy may be to affect the learner's motivational or affective state. It can also affect the way in which the learner selects, acquires, organizes, or integrates new knowledge. "For example, in preparing for a learning situation, a learner may use positive self-talk to reduce feelings of anxiety; in learning paired-associates, a learner may form a mental image to help associate the objects represented by the members of each pair; in learning from an expository passage, a learner may generate summaries for each section; in learning about a scientific concept, a learner may take notes about the material." Each of these activities: coaching, imaging, summarizing, and note taking is an example of a learning strategy. These strategies are learned rather than part of a learner's basic personality structure. Learned strategies include techniques such as creating visual images to assist with memory, relating new information to previously learned information, and organizing information into an easily remembered outline structure. Based on their past experiences, students express a preference for different types of strategies. For example, some students use surface strategies that mainly focus on memorizing key features to aid in retention. Other students look past the superficial layers and try to understand the fundamental structure of information; they are called deep processors. Many systems of learning strategies have been studied and each system contributes to psychologists' current understanding of how students invest their time during learning [88].

Demographics. Variables such as age, gender, and ethnic background each contribute some special qualities to learners. This is an area of great interests, but is well beyond the limitations of this work. To incorporate demographics into this predictive model, a sample population representing the current and expected trainees would have to be made available. Caution is warranted as this could raise issues regarding personnel privacy protection and social bias.

Skill Decay. Skill decay is real and without scheduled training can be a serious problem [58, 65]. Whether discussing cognitive, psychomotor, or affective skills, all abilities gained in training can be lost without frequent utilization. This is why training is still important, even if a person operates a piece of equipment regularly. Operating under normal conditions can become habitual, and leave a person ill prepared for off-nominal situations. This research is not investigating the frequency of training, but acknowledges that if training is not continued the effectiveness of any training program is diminished.

Learning Transfer. Learning transfer refers to the extent to which learning during training is later applied on the job. When predicting training effectiveness this is an important variable. Without training transfer, an organization will likely not benefit from its training investments. There are now models to help describe the process of training transfer [145], and research that links learning and transfer between the cognitive psychology domain and training context [146].

2.8 Instructional Design

The four fundamental components of instructional design are objectives, methods, learners, and evaluation. The previous discussion of LOs provided options for classifying training effectiveness and competency levels. Although the success of an instructional plan depends largely on the learning level achieved by the trainees, certain characteristics of the student population are important in determining training effectiveness [63]. Identifying which instructional methods are effective in achieving individual learning outcomes is an important step in evaluating training effectiveness. Studies show that the type of learner and instructional environment may affect training effectiveness [65]. This section will examine the effectiveness of common training methods used by the USAF and the results of instructional method experimentation. Also discussed are the differences in preferred instructional methods based upon student populations by age. The effectiveness of instructional methods and media are factors that influence learning in training. Investigating the use of instructional strategies for trainees to achieve each LO and understanding how the effectiveness of instructional strategies impacts different trainee age populations provides answers to RQ.2.1 and RQ3. RQ2.1 asks how to quantify the benefits of soldiers training in terms of effectiveness. RQ3 asks how to quantify increased knowledge, skills and attitudes in training system design.

2.8.1 Instructional Strategies

Even though society continues to change at an increasingly fast pace and within a more global context, there are several key elements that will increase instructional flexibility and effectiveness. The education field has historically and continues to lead the way in preparing global students for government, industry, and business institutions [139]. Instructional strategies determine the approach for achieving desired LOs and are selected during the design of a training system. Learning strategies basically embody the entire gamut of a learning environment, including processes such as media, methods, technologies, and styles [21]. In general, learning objectives point towards instructional strategies, while the instructional strategies will point to the medium that will actually deliver the instruction, such as electronic learning (e-learning), self-study, classroom, or on-the-job training (OJT). Clark provides a general guideline chart, Table 2.21, for selecting a learning strategy [147]. It includes taxonomies for each learning domain: cognitive, affective, and psychomotor based on Bloom's Taxonomy. The taxonomies are listed in ascending order according to skill complexity. The instructional strategies use passive learning methods for lower competency levels, and more active methods involving participation as the skill level complexity increases. This creates a direct correlation in learning:

- Lower levels of performance can normally be taught using the more passive training methods.
- Higher levels of performance usually require some sort of action or involvement by students.

Table 2.21: Instructional Strategy Selection Chart [147]

Instructional Strategy	**Cognitive Domain [72]**	**Affective Domain [127]**	**Psychomotor Domain [94]**
Lecture, reading, audio/visual, demonstration, or guided observations, question and answer period.	1. Knowledge	1. Receiving Phenomena	1. Perception 2. Set
Discussions, multimedia CBT, Socratic didactic method, reflection. Activities such as surveys, role playing, case studies, fishbowls etc.	2. Comprehension 3. Application	2. Responding to Phenomena	3. Guided Response 4. Mechanism
OJT, practice by doing (some direction or coaching is required), simulated job settings (to include CBT simulations)	4. Analysis	3. Valuing	5. Complex Response
Use in real situations. Also may be trained by using several high level activities coupled with OJT.	5. Synthesis	4. Organize Values into Priorities	6. Adaptation
Normally developed on own (informal learning) through self-study or learning through mistakes, but mentoring and coaching can speed the process.	6. Evaluation	5. Internalizing Values	7. Origination

2.8.2 Instructional Methods

Training methods are processes used to deliver instructional content and provide guidance to retain the skills and knowledge communicated [21]. Examples include lectures, demonstrations, case studies, etc. Several factors should be considered when selecting a training method. These factors fall into three major categories: constraints, cost-efficiency, and training effectiveness or considerations [103]. Constraints include the availability and location of students, instructors, facilities, safety, and development and training time. Cost-efficiency tradeoffs occur between the most effective means of imparting knowledge and meeting training requirements. For training that is required over an extended period of time for a larger group of trainees, on-the-job training (OJT) is likely expensive and disruptive for production. In this case, an instructional delivery method such as computer-based training (CBT) may be a better fit. However, if the training content requires frequent updates, CBT maybe less desirable and classroom lecture may be justifiable. A classroom lecture requires an instructor or trainer, which is an additional expense that can be excluded for most CBT. Training should include real application as part of OJT or within a very similar environment. One can see how each training method has pros, cons, and an associated cost to consider. A return on investment (ROI) analysis should be conducted that includes factors such as time spent in training by the student, instructor or facilitator, curriculum and courseware development, maintenance costs, and facilities and equipment costs, and the impact on mission readiness. The effectiveness of each training method must also be well thought out. Considerations for the task's criticality or importance, difficulty, fidelity, and interaction level must be managed. The *Air Force Handbook 36-2235 Information for Designers of Instructional Systems Application to Aircrew Training, Vol. 8* recommends that tasks or lessons that have a high level of difficulty and their performance is critical be taught in a formal classroom or OJT [21]. In some cases a simulated experience may be appropriate in lieu of OJT. For training that requires high fidelity, any training method that uses actual equipment to teach the process and procedures should be considered to ensure that learners are familiar with the actual system performance, characteristics, and environment. Self-study and passive training activities are recommended for learning processes that require low levels of interaction with others. The learning pyramid shown in Figure 2.10 lists several activities involved in training methods [148]. Common training methods defined in Table 2.22 help both the reader and experimenter map learning methods and instructional strategies.

Table 2.22: Training Method Definitions [21]

Method Type	Training Method	Definition
Presentation	Lecture (TM-1)	A formal or semiformal oral presentation of information by a single individual; facts, concepts, problems, relationships, rules or principles presented orally either directly (as by classroom instructor) or indirectly (as by video).
Presentation	Demonstration (TM-2)	Presentation or portrayal of a sequence of events to show a procedure, technique, or operation; frequently combines an oral explanation with the operation or handling of systems equipment or material. May be presented directly (as by a classroom instructor) or indirectly (as by video).
Presentation	Exhibit (TM-3)	A visual or print display used to present information; for example, actual equipment, models, mockups, graphic materials, displays, chalkboard, or projected images.
Student Verbal Interaction	Questioning (TM-4)	An instructor and/or courseware controlled interactive process used to emphasize a point, stimulate thinking, keep students alert, check understanding, or review material. Questioning may be direct, as by a classroom instructor, or may be designed into a film or television presentation.
Student Verbal Interaction	Seminar (TM-5)	A peer-controlled group interactive process in which task- or objective related information and experience are evoked from the students. Questions may be used to evoke student contributions, but the seminar is distinguished from questioning.
Student Verbal Interaction	Discussion (TM-6)	An instructor-controlled interactive process of sharing information and experiences related to achieving a training objective.
Knowledge Application	Performance (TM-7)	A student interaction with things, data, or persons, as is necessary to attain training objectives; includes all forms of simulation (for example, games and interaction with hardware simulators) and interaction with actual equipment or job materials (for example, forms). Performance may be supervised by classroom instructor, tutor, coach, or peer to provide needed feedback.
Knowledge Application	Case Study (TM-8)	A carefully designed description of a problem situation, written specifically to provoke systematic analysis and discussion.

2.8.3 Instructional Media

The means used to present information to learners or trainees is called instructional media. Media is used as the mechanism for presenting instructional material or basic communication stimuli to a student to induce learning [21]. Examples of media are classroom instructors, textbooks, slides, interactive courseware (ICW), and simulators. To meet learning objectives, the use of more than one medium may be required to convey instructional content. Common types of aircrew training media are listed in Table 2.23 along with corresponding definitions, examples, advantages, and limitations of each. Instructional methods and media options are used together to present to students the most effective and cost-efficient training possible. Each situation has to be evaluated for which media or medium is best. Selecting the media delivery format should include considerations for various effects such as, resources, classroom logistics, training schedule, cost etc. The characteristics of each media type listed in Table 2.23 make certain media suitable or unsuitable for particular training settings. Most types of complex skills involve multiple learning objectives that cross learning domains [103]. Media selection for a training skill that involves two or more learning objectives (LOs) from different learning domains typically requires multiple instructional strategies and media formats. The military instructional design handbook, MIL-STD-29612-2A provides some guidelines for proper media selection [103]:

a) Select media that do not conflict with the specific training or job task environment.
b) Select media that effectively supports the LOs at the appropriate learning levels.
c) Select media that supports the training strategy.
d) Select media that allows individualization of training when appropriate.
e) Select media that will support anytime anywhere training.
f) Select media with time and dollar resources in mind.

g) Select media that are effective and cost-efficient.

Advances in technology are prevalent in all aspects of our lives, and the training environment is no exception. Organizations that have been surveyed for current industry practice, show that an increasing number of organizations are implementing technology-based training instead of traditional forms of training [65]. Researchers warn that both traditional and technology-based forms of training can work and fail. Trainees can sit and listen without learning just as easily as they can interact with the computer and make poor decisions that lead to suboptimal learning. Sitzmann et al. conducted a study on self-regulation in both online, work-related training and laboratory settings [149]. The results showed that prompting self-regulation while using technology-driven instruction improved or held constant trainee's performance. The LOs tested included procedural and declarative knowledge and strategic (i.e. tacit) performance. Trainee performance declined over time when they were not prompted to self-regulate. This suggests that implementation of prompts will enhance trainees' ability to remember the key principles presented in training, and their understanding of when, where, why, and how to apply their knowledge and skills [146, 150, 151].

The key to effective training is a well-designed training program that does not depend on the delivery mode and media used, but uses these as enhancements to communicate the learning objectives. This is why evaluating training effectiveness must include an assessment of the learning theory utilized and instructional strategies.

Table 2.23: Common Types of Media [21, 103]

Type of Media	**Definition**	**Example**	**Advantage**	**Limitation**
Instructor/ Tutor	Any individual who presents instruction.	Lecturer Demonstrator Tutor/Coach	1. Immediate feedback about student progress is available and changes to instructional delivery method can be made during the course	1. Traditional classroom instruction requires student and instructor to be in the same location.
Traditional Audio/ Visual Devices	Any delivery device or system, which provides both audio and visual presentations.	Chalkboards Transparencies Overhead projectors Slides Pre-narrated slides	1. Easy to prepare with regular audio equipment. 2. Can provide applications in most subject areas. 3. Equipment is compact, portable, easy to operate. 4. Flexible and adaptable as either individual elements of instruction or in correlation with programmed materials. 5. Duplication easy and economical.	1. Have a tendency for overuse, as lecture or oral textbook reading. 2. Fixed rate of information flow.
Print	Training materials that require reading.	Workbooks Study guides Job aids Training manuals Programmed instruction booklets	1. Include common types of materials. 2. Have wide variety of applications. 3. Simple types quick to prepare.	1. Sophisticated types more costly to prepare. 2. Require suitable reading ability.
ICW	Computer-controlled training designed to allow the student to interact with the learning environment through input devices such as keyboards and light pens. The student's decisions and inputs to the computer determine the level, order, and pace of instructional delivery, and forms of visual and aural outputs.	CBT (traditional) IVD CMI	1. Presents text information and graphic images. 2. Can interact with learners on individual basis through asking questions and judging responses. 3. Can maintain record of responses. 4. Can adapts instruction to needs of learner. 5. Can control other media hardware. 6. Can interface computer and video for learner-controlled programs.	1. Requires computers and programming knowledge. 2. Requires essential hardware and software for development and use. 3. Incompatibility of hardware and software among various systems.
Training Devices and Simulators	Hardware and software designed or modified exclusively for training purposes	Flight training simulators Part-task trainer Computer simulation	1. Imitates operational equipment both physically and functionally 2. Part-task trainers are less expensive to develop and	1. Not all human abilities for the real-world task are simulated 2. Costly to

Type of Media	Definition	Example	Advantage	Limitation
	involving simulation or stimulation in its construction or operation to demonstrate or illustrate a concept or simulate an operational circumstance or environment.	Actual equipment trainers	maintain than a full capability simulator, and multiple units increase the number of trainees who can simultaneously practice. 3. Simulators allow trainees to practice skills in the most realistic artificial environment. They can simulate scenarios that are not possible or practical in the actual setting (e.g. deployment of weapons in combat). Less expensive to operate than most operational equipment.	fabricate and repair 3. Must be constantly revised as operational equipment is upgraded, the theatre of threat is changed, or as the adversary equipment changes. 4. Practice is limited to one person or team at a time per device.

2.8.4 Mapping Learning Objectives and Instructional Strategies

To evaluate training effectiveness or the effectiveness of instructional design one should verify that the strategy (instructional method) used is appropriate for the desired LOs [79]. LOs from the cognitive, affective, and psychomotor domains are taught using multiple training methods. A mapping between instructional strategies and LOs would provide a set of constraints to use when evaluating training system design for effectiveness.

One of the fastest growing areas of science and training is the potential educational benefit of technology-based training. In a survey of organizations in the ASTD's benchmarking service, the percentage of companies using technology-delivered training increased from 8% in 1999 to 27% in 2004, and about 75% of the technology-delivered courses in 2004 were online [34]. Additionally, over 1,100 institutions of higher education in the U.S offer online courses [152]. One branch of the armed services, the Army, uses online instruction as a retention tool, with over 40,000 soldiers in 50 countries pursuing advanced degrees online in 2003 [153]. There is no doubt that technology is shaping how training is delivered in industry, government, and higher education [68]. Organizations still rely heavily on classroom training, but many are implementing technologies such as video conferencing, electronic performance support systems, and on-line Internet/Intranet courses. Advances in technology are also enabling the development of intelligent tutoring systems that have the potential to reduce or eliminate the need for human instructors for certain types of learning tasks. Recognizing the paradigm shift that is taking place in training, MPEET will create a mapping between instructional strategies and LOs based on data collected from the traditional classroom as well as technology-based training.

In 1994, Clark argued that the media type used for instruction doesn't matter [154]. According to Clark delivery media, such as computers, video teleconferencing, and the Internet, are inconsequential in affecting learning outcomes, especially when compared with more powerful influences such as individual differences and instructional methods. Clark's position argues that no instructional medium is uniquely advantageous. Well-designed instruction works irrespective of the delivery mode. Alternatively, pro-technology researchers believe that Web-based instruction (WBI) provides greater flexibility and greater access to multiple instructional methods and may be superior to training media that only use a single instructional method [155-157]. In 2006, Sitzmann, Kraiger, Stewart, and Wisher conducted a meta-analysis to examine the effectiveness of using WBI relative to classroom instruction (CI) for teaching declarative and procedural knowledge. Their hypothesis was that WBI is more effective then CI, for teaching both types of knowledge, thus attempting to reject Clark's claim [153]. They examined 96 studies reporting data, including 65 published studies, 18 dissertations, and 13 unpublished studies. The studies reported data collected from 19,331 trainees who took part in 168 training courses from 1991 to 2005. The topic of training courses ranged from psychology, engineering, computer programming, business, and technical writing. Undergraduate students consisted of 67% of trainees, 18% were graduate students, and 15% were employees. Of the 96 studies that reported demographic information, the average age of participants was 24 years old and 41% of the participants were men. Other meta-analyses have been done comparing WBI and CI, but none cover such a vast variety of employee and college training courses while making a clear distinction between cognitive and physical skill based knowledge [155-167]. Sitzmann et al. concluded that across all 96 studies, on average WBI was slightly more effective than CI for teaching declarative knowledge. However, trainees learned the same amount of declarative knowledge from WBI and CI when the same instructional methods were used to deliver both media types. WBI and CI were equally effective for teaching procedural knowledge. Overall the results supported Clark's argument that instructional methods, rather than delivery media determine learning outcomes. The results of the Sitzman et. al found that the extent to which Web-based trainees learned more than classroom trainees was greatest when Web-based trainees were provided with control, when trainees practiced the training material, when trainees received feedback during training, and in long courses. Under these conditions, the WBI declarative knowledge effect was 19% more effective than CI. In contrast, results showed that it is also possible to design Web-based courses in which learning levels will be inferior to CI. CI was 20% more effective than WBI for teaching declarative knowledge when WBI failed to provide control opportunities to practice, did not give feedback to learners, and in short courses. Thus, attention to course design features is critical for maximizing learning outcomes. Trainees were equally satisfied with the delivery media of both WBI and CI. Sitzmann et al. wanted to assess affective learning, however there was an insufficient number of studies available to determine whether online learning is more or less effective than the classroom for affective objectives. Some of the reports collected also addressed a blended learning environment where WBI was used to supplement face-to-face instruction. Across all the relevant studies, the results indicated that the blended learning environment was more effective than stand-alone CI for teaching trainees job-relevant knowledge and skills. Understanding the best instructional strategies for students to excel in accomplishing each LO provides a set of constraints to use when evaluating training system design for effectiveness

2.8.5 The Impact of Age Differences in Training

The results of Sitzmann et al. meta-analysis showed that CI was more effective than WBI for teaching declarative knowledge when trainees were *randomly assigned* to courses. Normally, trainees are not randomly assigned to a course. So Sitzmann et al. also examined how different age groups responded to WBI and CI training. They found that across all 96 studies, the mean ages of WBI and CI groups accounted for a significant 44.2% of the variance in the effects of declarative knowledge. As the age of Web-based trainees increased and the age of classroom trainees decreased, Web-based trainees learned extensively more. Trainees aged 23–45 learned more declarative knowledge from WBI than CI, while trainees ages 18–22 tended to learn more declarative knowledge from CI. It is possible that, in accordance with andragogical learning theories, slightly older trainees are more adept at dealing with the autonomy and learner control provided by

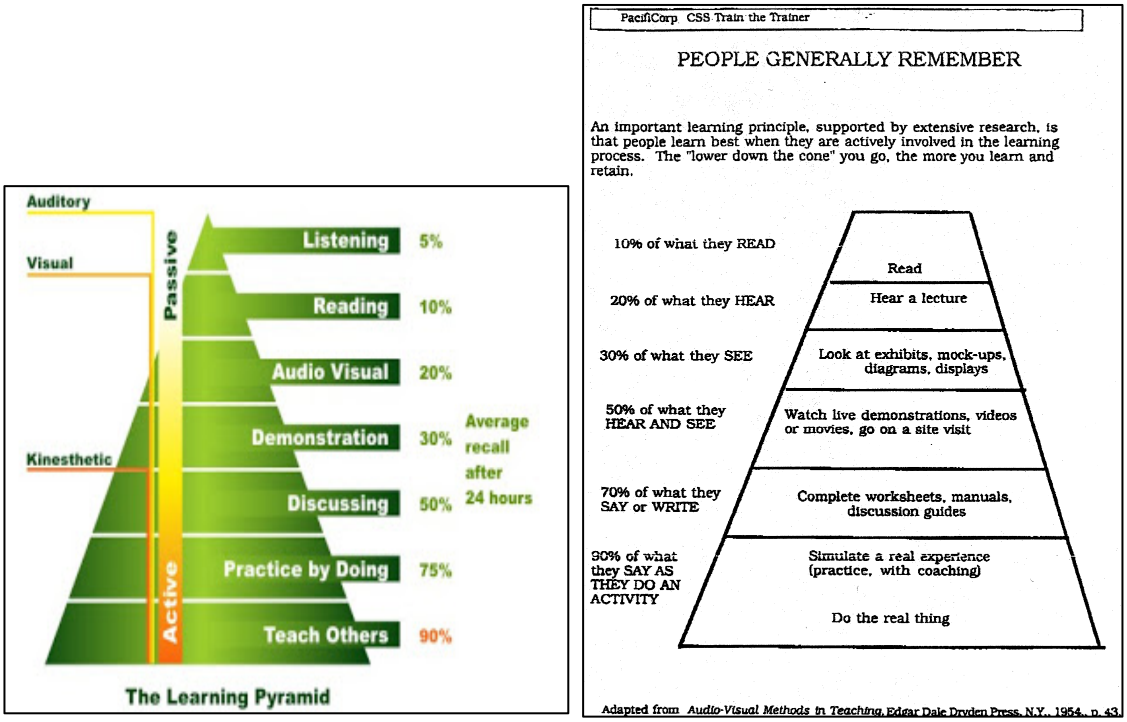
WBI [168]. Younger trainees may be more successful in a structured classroom environment. Differences between younger and older students or trainees have been an area of concern in andragogy for decades.

For a training program to be effective, the design of instructional content should account for characteristics of different types of learners [63, 65, 68]. Andragogy, the study of non-traditional or adult-learners, has observed differences among this student population. Normally, non-traditional or adult-learners have been removed from the academic environment for five years or more and are usually 25 years or older [169, 170]. There is not a commonly accepted definition of an adult-learner, but there are characteristics that adults display which educators use to classify these students [139]. The learning environment for this population may include the following settings: those returning to colleges and universities; enrolling in distance education programs; engaging in community adult education programs; and participating in job training or retraining for new skills in business, industry, health fields, government service, and the military. Traits of adult-learners include engagement in multiple roles such as a spouse, parent, employee, caregiver, or community activist. These roles have a direct impact on the amount and quality of time they can devote to learning [63]. Adult learners tend to bring more life experiences and strong, sometimes unwavering, beliefs to the classroom in comparison to younger students [139]. These experiences and beliefs create a grounding and building block for new knowledge, which could be positive or negative for both the student and instructor. In comparison to traditional students, many adults have some level of fear about the challenges of returning to school or training later in life. In contrast, one advantage they normally have is clear goals and a planned timeline for completing.

Lastly, adult learners are more likely to pay their expenses out of pocket or have their company pay for their courses, many have off-campus activities that require attention, and some may be peers or older than their instructors. Adults who have a higher stake in the cost and time investment required for training tend to be more motivated to learn [171]. They appreciate a program that is structured systematically, with requirements (i.e. objectives) clearly defined. Adults want to know how the course content will benefit them and expect the material to be relevant and practical. Studies have shown that adults respect an instructor who is fully knowledgeable about the subject matter and presents it effectively. Adult students quickly detect an unprepared instructor. Even though adult-learners may lack initial confidence, they are self-directed and independent workers. They prefer that the instructor serve as a facilitator to guide and assist, rather than an authoritarian leader. Adults want to participate in decision-making. They desire to cooperate with the instructor in a mutual assessment of needs and goals, choice of activities, and decisions on how to evaluate learning. These generalizations are widely true of adults, but Morrison et al. believe they apply to all learners [63]. They argue that the degree and specificity of applied instructional strategies may vary among certain groups of learners when the instructional media are designed and instructional activities are carried out. They agree that when instructional content is developed while recognizing and accounting for characteristics of different learners, the training programs are more effective. The results of the WBI versus CI meta-analysis show similarities to other studies involving online and classroom media for adult-learners. Graham compared attitudes toward tasks related to school, motivation, and anxiety levels of traditional and non-traditional aged college students (mean ages 19 vs. 34) [172]. She found that non-traditional students had more positive attitudes, were more motivated, and experienced less anxiety than traditional students. In addition, Tallent-Runnels et al. reviewed the literature on WBI and concluded older trainees in WBI are more focused on achieving specific learning outcomes than younger trainees [173]. Studies on WBI and CI instructional delivery media indicate that WBI is more effective than CI for adult learners. This is a general finding; there are other contributing factors such as prior computer knowledge, online experience, access to quality data connection, and the quality of instructional design. There is enough evidence to conclude that if the student population involves adult learners than the training system design should factor in the cost and benefits of technology-assisted instructional media.

Evaluating pilot training programs requires another distinction in andragogy. Training is required for pilots to become initially certified to fly a particular aircraft, and then throughout a pilot's career they take refresher training or may choose to become certified on a different aircraft. For example, former military pilots who later fly commercially have to go through certification training on commercial planes. This means that there can be multiple generations of students attending a training program and each generation has different learning preferences. Over the last century four generation groups have been defined: Traditionalists, born 1925-1944; Baby Boomers, born 1945-1962; Generation X, born 1963-1979; and Millennials (also referred to Generation Y), born 1980-2000 [174]. Millennials have an appreciation and expectation for the use of technology. Multi-tasking is a way of life for this generation. Unlike the older generations, they believe they can learn complex information while listening to music or engaging in other activities [175]. They dislike traditional lectures where they have to sit and listen for hours; instead they prefer class discussions and stimulating exercises. This has led instructors to modify teaching strategies and invent new learning strategies. From a cognitive learning perspective, Millennials want to learn in a collaborative environment; many of them enjoy the activity of teamwork. They have a preference to learn in their own time and on their own terms. Structured activities that permit creativity are appreciated. They want to be involved with "real life" issues that matter to them. Cognitive psychology research shows that active engagement promotes deeper levels of processing and learning because it creates stronger connections between the subject matter and student. The more connections students make with the material, the more retrieval cues they have to access it later. This helps students build upon and organize what they know. In the case of Millennials there has been an emphasis to use learner- or student-centered strategies [176]. Learner-centered educational methods concentrate on the individual student, allow self-regulation, and engage in student metacognition. These teaching methods empower students with real independence in the learning process. It also means that more of the burden shifts to the student in terms of comprehending and really understanding course material. The effective use of technology is at the core of these learning adaptations for not only Millennials, but also for Generation X. Generation X is familiar with and frequently uses digital and cyber technology, and Millennials are saturated with it [177]. Delivering training with simulation is a popular method for both groups. Simulators are widely used in business, education, and the military, with the military and commercial aviation industry being the largest investors in simulation-based training [68]. These simulators vary in cost, fidelity, and functionality. Many simulation systems (simulators and virtual environments) have the ability to mimic detailed terrain, equipment failures, motion, vibration, and visual cues giving very realistic experience to students. In some cases simulation exercises allow students to experience scenarios that are too dangerous or costly to rehearse live. Low fidelity simulators have less sophistication, but can represent the knowledge, skills and attitudes to be trained very well [178]. In the 1990s there was a trend to uses more of the low-fidelity devices to train complex skills. Studies are ongoing to determine the viability of computer games for training complex skills [68]. There is a concern that simulation and games are being used for training, but the skills are not transferring to the real environment. Nevertheless, Millennials will arrive with an expectation that training involves the use of electronic systems. CBT and simulation-based training provide a means to facilitate the transfer of information to Millennials in a format that meets those expectations.

Research studies claimed by the National Training Laboratories (NTL) have also shown that lower behavioral expectations can be met with passive learning methods, while higher learning performance requires active training methods. Their studies were done in the context of knowledge retention. Based on the method of instruction for learning and training, how much does a student retain 24 hours later? The results, referred to as the learning pyramid, are commonly presented as a triangular image mapping a range of teaching methods and learning activities in proportion to their effectiveness in promoting student retention of the material taught, Figure 2.10 [148]. The research base for the pyramid is difficult to establish conclusively. It was developed and used by the National Teaching Laboratory Institute at their Bethel, Maine, campus in the early nineteen sixties, when that organization was part of the National Education Association’s Adult Education Division. NTL believes it to be accurate, but says that it can no longer trace the original research that supports the numbers. NTL acknowledges that, in 1954, a similar pyramid with slightly different numbers appeared on p. 43 of a book called *Audio-Visual Methods in Teaching*, by the Edgar Dale, shown in Figure 2.10. The currently used learning pyramid seems to have been modified, but has always been attributed to the NTL Institute. NTL allows free use of the pyramid and gives specific instructions for citations. Although there remains a level of discomfort and disagreement in academia around the use of an instrument with such a tenuous research base, NTL is a reputable organization that requests to have its name – and reputation –associated with the pyramid [179]. After his 1954 publication, Edgar Dale continued his research and advised instructors that the learning pyramids and cone of experience are not an exact or flawless representation of everything that takes place in the learning process [180]. These models are useful, but one should not think that the method for teaching all objectives should be an active form of practice just because it has a higher retention rate. Varied types of sensory experiences should be provided to students based on the learning objectives. Lalley and Miller give a good example of the importance of diverse teaching methods based on a scenario of a heart surgeon [181]. The surgeon cannot learn only by hands-on experience, or the patient will likely die. Likewise, the surgeon cannot simply learn through reading, since reading is not a substitute for real life practice and experience. A variety of teaching methods are required to effectively teach a surgeon and maximize retention. Information must be presented sequentially and with the most appropriate method for the current learning stage of the student. The same applies for any other instructional field or training program. Essentially, the training methods have a built in additive effect when used properly. With the surgeon example, the practice by doing would not be 75% effective if the surgeon had never read about heart surgery (10% effective), watched demonstrations (35% effective), and had the opportunity to ask questions and get clarification during discussions (50% effective). As Lalley described, if the first training method introduced to a medical student was hands on application the effectiveness is zero, because the student has not received other instruction to prepare him or her for the on-the-job tasks. Therefore, learning pyramids and taxonomies "serve as a guide to (1) the uses of certain print and non-print material in teaching, (2) the progression or stages of various forms of group discussion, (3) ordering of particular forms and techniques of activity, and in general the sequencing of class work and assignments in a lesson, a unit, a course, or even an entire curriculum" [120]. Trainees learn new skills by being informed of the learning objectives, watching others perform the action correctly and improperly through audio and/or video media, practicing targeted behaviors, receiving feedback, and being given an opportunity to translate their abilities into new environments [65]. To accomplish all these steps a variety of learning strategies, methods, and media devices are necessary.



PacifiCorp CSS Train the Trainer

PEOPLE GENERALLY REMEMBER

An important learning principle, supported by extensive research, is that people learn best when they are actively involved in the learning process. The "lower down the cone" you go, the more you learn and retain.

10% of what they READ	Read
20% of what they HEAR	Hear a lecture
30% of what they SEE	Look at exhibits, mock-ups, diagrams, displays
50% of what they HEAR AND SEE	Watch live demonstrations, videos or movies, go on a site visit
70% of what they SAY or WRITE	Complete worksheets, manuals, discussion guides
90% of what they SAY AS THEY DO AN ACTIVITY	Simulate a real experience (practice, with coaching) Do the real thing

Adapted from *Audio-Visual Methods in Teaching*, Edgar Dale Dryden Press, N.Y., 1954., p. 43.

Figure 2.10: Learning Pyramid [148]

When evaluating a training program for effectiveness, especially for hands-on military soldiers, the training program should build up from passive methods to more active student participation exercises. There is not a one-size fits all mathematical equation or graphical curve to represent how training should vary with teaching methods, but a visible progression should be evident. For the most accurate predictability model, a physical experiment should be conducted within the specific training environment being evaluated to determine the true knowledge retention rates and the pace of student progression necessary to ensure skill development at each level of the hierarchies. This experiment should include a large enough sample population to validate the bounds for each training method so an expected value and variance could be used in future evaluations. Because of the associated costs, length of time, and availability of C-130J aircrew, an actual experiment was not being performed as part of this research. The methodology proposed herein uses the learning pyramid and learning domain taxonomies in conjunction with SME input to predict training effectiveness by testing whether or not the curriculum lessons generally follow the suggested order of the training industry accepted models. Verifying that a training program is administered through a sequence of passive to active instructional methods aids in quantifying the effectiveness of training, RQ 2.1.

In summary, MPEET can now fully assess two of the four primary components in training system design, objectives, and methods. A compatibility matrix will be developed to create a mapping between the LO competency levels for the cognitive, affective, and psychomotor taxonomies and the training methods that can be administered during training. Knowing the effectiveness or percentage of knowledge recall for each training method is an attribute that will be used in the cost-utility analysis. The third fundamental component of instructional design is a focus on the learner. MPEET does not assess all the variables that impact learners and the effectiveness of training. MPEET does take into consideration the impact training will have on different generations of students through the effectiveness values assigned to each method. Generalities for age differences between adult learners, Generation X, and Millennials are the only learner attributes included in MPEET. As previously discussed in section 2.7.1, there are many other variables that influence training effectiveness that require pre- and post-training comparisons. Although these variables are not included during this initial development of MPEET, it is acknowledged that they exist and inclusion may result in more accurate predictability, if the data is available for input. The final principle component of instructional design is ensuring the evaluations used to verify that trainees have learned the course objectives are adequate and well designed and also that the course itself is properly evaluated for the design and implantation of the training and instructor. Because MPEET is assessing training system design before training is administered, this fourth component does not fit within the primary purpose of this research. LOs in three learning domains, instructional methods, and media have been discovered as factors that influence training effectiveness. The research thus far reveals how instructional strategies are used to accomplish LOs, and how they impact different age populations. This provides partial answers to understating how to quantify the benefits of soldiers' training in effectiveness terms, RQ 2.1, and how to quantify increased knowledge, skills, and attitudes in training system design, RQ 3.

2.9 DIFE Analysis

2.9.1 Design

When an instructor or evaluator wants to decide between 'need to know' and 'nice to know' training content a difficulty, importance, frequency, and consequence of error (DIFE) analysis is appropriate [182]. DIFE analysis is used to help decide what training subtasks are required. It is also used to determine the length or intensity of a training task to ensure a student is adequately prepared [103]. Figure 2.11 provides an example of a DIF analysis. Shown in this diagram are three criteria – the level of difficulty of a task, the importance that is placed on the task, and the frequency with which it is performed – all used to decide if training for a particular task is necessary and to what level in general. An organization or instructional designer can create a hierarchy based on the information available about the job. In this particular example, a task that is difficult, important, and performed frequently requires training. A task that is difficult, not important, and not performed frequently requires no training because it is assumed the person can learn this skill via on-the-job experience. However a task that is difficult and important, but not performed frequently requires over training. Over training does not refer to a repetitious or unnecessary amount of training, but the trainee must be trained to such a level in terms of skill or knowledge retention that there is a minimal chance of underperformance when the event occurs. Built into the decisions to train, not to train, or to over train in this diagram is an assumption regarding the consequence of error and possible immediacy of response time being low. By not training personnel on how to perform a task that is difficult, unimportant, and infrequently performed assumes that the employee has time to seek assistance because they will not know what to do in this situation; and if the employee does act without assistance and incorrectly performs this task, there is little to no consequence of safety, danger, production time, etc. For situations such as emergency procedures, personnel are usually over trained to minimize the catastrophic risks associated with tasks that are difficult and important, but hopefully infrequent. Not every task needs to be measured against DIFE criteria for training design purposes. It is a useful tool when decisions are difficult to make regarding what training must be covered in a limited amount of time.

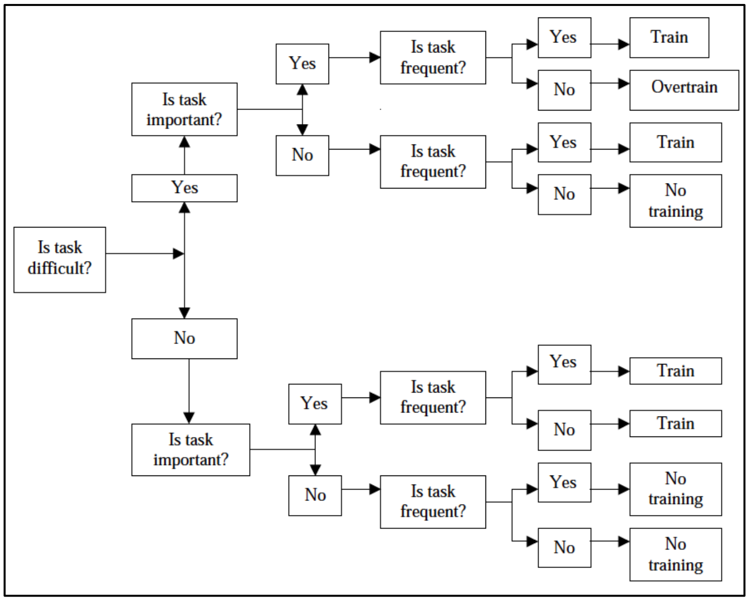


Figure 2.11: Example of DIF Analysis [182]

The military training guidebook also provides several training selection models to help instructional designers in selecting the proper tasks for training [103]. One of these models is called the criticality, difficulty, frequency model, shown in Figure 2.12, and is simple, yet more descriptive than Figure 2.11. Difficulty is rated as low, average, and high. Criticality of performance, like importance, is scaled as yes or no, but the frequencies of performance options are low, average, and high. It has the advantage of being straightforward, easy to administer, it can be used with for a small or large training program, and requires inputs from all stakeholders. The disadvantage is that this is still a crude tool for analysis and very subjective.

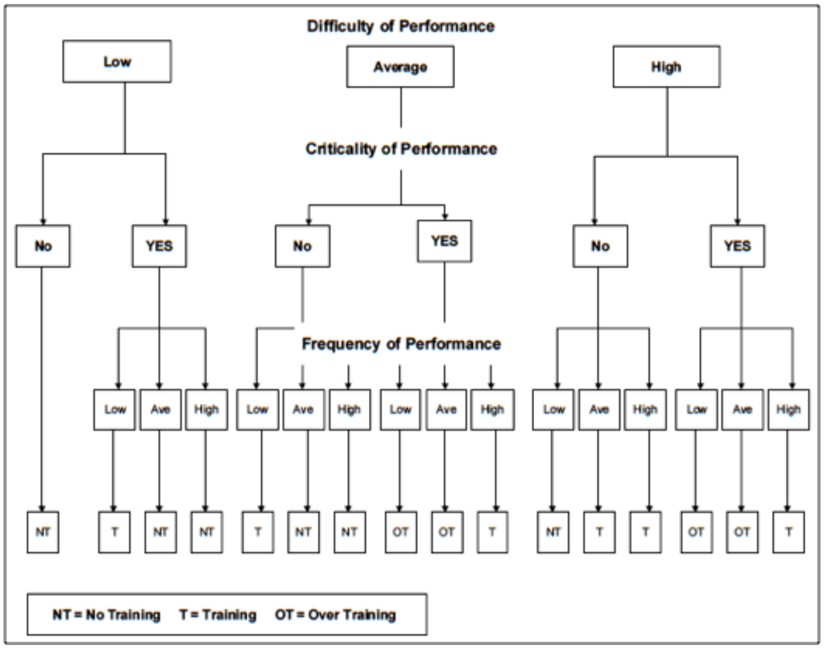
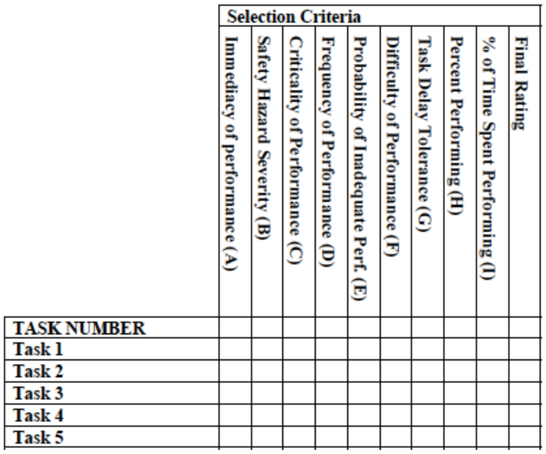


Figure 2.12: Criticality, Difficulty, Frequency Task Selection Model [103]

A more extensive multiple factor model, the all critical learning (ACL) model, allows the training system designer to select the training tasks based on weighted criteria. Using ACL, the instructional designer can select the criteria, choose the scaling ranges, and assign weight factors based on the most important criteria. Figure 2.13 shows a sample analysis sheet. Once the criteria ratings are determined for each training task, the rating is multiplied by the corresponding criteria weight factor, and the sum of all the evaluations is added per task to calculate the final rating. The tasks with the highest overall final ratings are selected for training. This method provides a fairly comprehensive set of data for each task, and will certainly aid future instructors, designers, and evaluators in understanding how the training tasks were selected. The documentation and decision making process is very easy to follow. The disadvantage in using this method is that the weighting factors can be subjective and it could become time consuming. Because the evaluation is mathematically based, trade studies can be performed by changing the weighting factors or criteria ratings to determine the sensitivities in selecting training tasks based on various stakeholder inputs.



	Selection Criteria									
	Immediacy of performance (A)	Safety Hazard Severity (B)	Criticality of Performance (C)	Frequency of Performance (D)	Probability of Inadequate Perf. (E)	Difficulty of Performance (F)	Task Delay Tolerance (G)	Percent Performing (H)	% of Time Spent Performing (I)	Final Rating
TASK NUMBER										
Task 1										
Task 2										
Task 3										
Task 4										
Task 5										

Figure 2.13: ACL Training Task Selection Model [103]

2.9.2 Evaluation

In terms of evaluation, DIFE analysis can be used to determine the relative weight of training lessons [183]. By building in different degrees of difficulty, importance, and frequency for a given task, the DIFE analysis technique can be enhanced. Two approaches are presented here, the first by Buckley and Caple. They propose a technique, as shown in Figure 2.14, that includes five levels of training. Each task difficulty remains a yes or no value, as in Figure 2.11, but the importance of each task is rated as not very important, moderately important, or very important. Frequency is also rated as infrequent, moderately frequent, or very frequent. Instead of the simple decision of to train, not to train, or to over train as shown in Figure 2.11, the rating of DIFE corresponds to specific levels of training that will be given to each student for that task. The levels of training are on a scale of one to five, where five is over training and level one is no training. The advantage to having this information in predicting training effectiveness is the ability to see how many training lessons fall into each level of training. This allows an evaluator to assess how much training time and cost is being spent on average for high priority tasks versus the time and money spent on low priority tasks. Of course it is important to have inputs on the relative training level definitions because jobholders, supervisors, and customers can all have different perspectives when it comes to the degree of importance and difficulty for a particular job or task. Frequency and consequence of error are normally easy to get stakeholders to agree upon because measurable criteria can be set. For example, instead of scaling frequency in terms of ‘not very’, ‘moderate’, and ‘very’, a more objective time variable scale can be used such as ‘daily’, ‘weekly’, and ‘monthly’. Error consequence can be rated as harmful to employee, dangerous to life, costs of failure to the organization, etc. Scaling difficulty is normally easy, as long as long-term workers or supervisors assess the level of difficulty for a new person and not in light of their numerous years of experience. Instead of a simple yes or no, as shown in Figure 2.11 and 2.14, difficulty may be rated as not difficult, average, or very difficult. Here are the descriptions for the five levels of training corresponding to Figure 2.14 [182]:

- Level 1 indicates a very high priority for training to a standard, which will ensure that a high level of skill and knowledge is retained without the job being done frequently. In effect this is ‘over training’.
- Level 2 sets a high priority for training to a standard of competence that will ensure that the task can be done without further training.

- Level 3, being the midpoint of the scale, sets the priority level at average and to a standard that will ensure that the task is done efficiently. Further training or practice would be required to enhance performance.
- Level 4 sets a low priority for training at a standard, which provides no more than a basis for on-job training and practice.
- Level 5 indicates that no formal training is required and that the task should be easy to learn while doing the job.

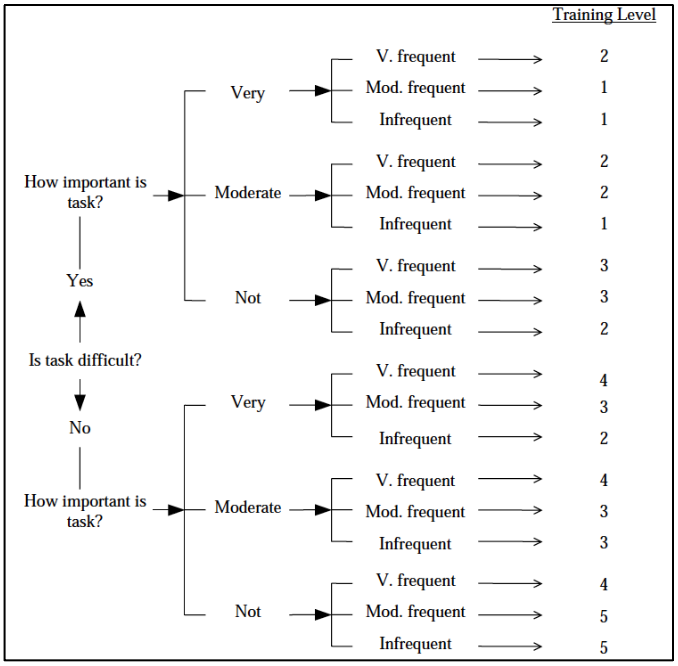


Figure 2.14: Example of DIF Analysis and Levels of Training [182]

The second method for using DIFE analysis in training evaluation can be shown from an experiment conducted for Bell during the 1980s by Cascio and Ramos. They used a DIFE analysis to calculate the relative weight of managerial activities and later used these weightings to calculate the dollar value and worth of the principal tasks performed by each manager within the company [183]. In their approach time/frequency

(F) was rated on a zero to one hundred percent scale. For all the activities performed by the manager, each was ranked based on the time spent performing that task, but the total time required of all tasks must add to 100. Importance (I), level of difficulty (D), and the consequence of an error (CoE) in performing a task were all rated on a scale from zero to seven. Descriptions of each rating level for difficulty, importance, and consequence of error were provided to the raters. The scales for all four dimensions were multiplied together under the assumption that each scale relates directly to job performance. The relative weight was calculated by multiplying the values for F x I x CoE x D, and dividing by the total of all the activities. Table 2.24 provides an example of this method.

Table 2.24: Sample Data Illustrating DIFE Analysis and Conversion to Relative Weightings [183]

Principal Activity	**Time/Frequency (F)**	**Importance (I)**	**Consequence of Error (CoE)**	**Level of Difficulty (D)**	**Total**	**Relative Weight**
1	30	3	5	3	1,350	0.29
2	20	5	3	5	1,500	0.32
3	40	2	1	2	160	0.03
4	10	7	6	4	1,680	0.36
Total	100	-	-	-	4,690	1.0

Being able to view the relative weightings of the training lessons included in a training program enables designers and decision makers to compare how training costs are being dispersed for various training activities. If someone looks at plots of training time versus relative weightings or lesson costs versus relative weightings, there is an expectation of positive correlation. Negative correlation suggests that time and money are being spent on training activities that may be unimportant, infrequently performed, not difficult, or have a low CoE. This is why it is essential to have agreement on the scaling used for each rating category. Also, in this experiment conducted for Bell, the researchers choose to combine DIFE into one number multiplicatively, under the assumption that each scale relates directly (or independently) to overall job performance. Depending on the context and task this assumption may be invalid. For example, reading a gauge once a day in a nuclear power plant may not be difficult and takes very little time to complete; however, the consequences of failure may be disastrous and thus the importance is very high. Frequency may be low or high (depending on the scale being used), difficulty is very low, but importance and CoE are very high. One must exercise caution when combining the four scales into one mathematical value. Research done in the field of psychometrics[5] suggests that the scale scores should be combined multiplicatively if they are significantly intercorrelated; however the scores should be combined additively if they are independent of each other [185]. If using this technique of combining DIFE ratings into one number for comparisons and decision-making, one must first examine the correlations between each dimension. In Bell's collected data the ratings of DIFE were all significantly correlated, so the approach by Cascio and Buckley is valid.

Information provided from DIFE analyses can benefit MPEET in the assessment of training system designs. The advantage to having this information in predicting training effectiveness is the ability to see how many training lessons fall into each level of training. This allows an evaluator to assess how much training time and cost is being spent on average for low and high priority tasks, and aid in deciding appropriate resource allocations, which relates to RQ 2.2.

---

[5] Psychometrics is the science and enterprise of using tests to 'measure' psychological traits, abilities, and learning 184. Curren, R., *Connected Learning and the Foundations of Psychometrics: A Rejoinder.* Journal of Philosophy of Education, 2006. **40**(1): p. 17-29.

2.10 Cost Analysis

Regarding the allocation of monetary resources to gain maximum training effectiveness, RQ2.2, training systems must be strategically designed, delivered, and evaluated with clear documentation of training benefits [35]. Managers and other decision makers prefer information and data in terms of business-related results (cost, time, productivity etc.) to make decisions about how to allocate resources, including those for training activities [186]. In today’s highly competitive environment, with budgetary cuts and constraints and market-driven economic philosophies, the most cost efficient means of training is being used in organizations without always properly considering how effective the training program actually is. Value based assessment cost analyses can provide objective analysis on the balance of training effectiveness and efficiency.

Performing value based assessments and alternative analysis is not unique to the military and defense industry. Firms across various industries are constantly performing cost analyses prior to committing investments into new products, services, or research and development (R&D) efforts. These resource allocation decisions are essential to companies remaining competitive in their respective markets. For-profit firms in highly competitive markets typically have an advantage when conducting cost analysis because the market communicates the relative value of a good, service, or investment in terms of a monetary price through the relationship of supply and demand [187]. Thus, the commercial firm is more objectively equipped to compare the costs it will incur as well as the required rate of return it must achieve from an investment of resources when assessing different alternatives. In comparison, it is sometimes difficult when assessing military training programs to determine the appropriate balance between training cost investments and an effectiveness benefit. Compounding this problem is the fact that this research conducts the cost analysis before training actually occurs, and the benefits of increased military training effectiveness are not as easily interpreted as a cost benefit compared to a commercial good or service sold. This prompts non-traditional products and systems to adopt Cost-Benefit (CB) Analysis, Cost-Effectiveness (CE), or Cost Utility (CU) analysis methods to evaluate different investment alternatives. In any serious evaluation a proper assessment of both costs and effectiveness is necessary, and the two should not be considered separately because the independent results can be misleading [188].

The need for understanding the effectiveness of warfighter training and the associated costs has already been established in regards to executive level decision making among DOTMLPF-P solutions [7]. But there is also a general benefit in performing cost-effectiveness analysis for training. At any level of decision-making, cost-effectiveness analysis can aid in determining the most efficient use of training resources. Applying the appropriate resource constraints and criteria weightings, cost-effectiveness analysis can objectively and/or subjectively identify the cost necessary to reach a particular objective(s), and/or identify which objectives can be met within a financial threshold. It is important to properly assess both the cost and effectiveness elements in any training evaluation. Making decisions based on cost or effectiveness as independent variables can lead to unexpectedly expensive training programs or low quality trainee outcome. For example, project managers may identify a performance gap and choose training as the means of filling that gap, but there is always an allocated training budget that may or may not be sufficient to meet the expectations of the management team. It is very common for a manager to want maximum effectiveness for a given budget, or conversely, to achieve a certain level of effectiveness at a minimal cost [188]. Decision makers are unlikely to accomplish either goal if higher effectiveness and lower costs are pursued as independent goals. Thus, it is wise to consider the cost and effectiveness of training dependently throughout any problem-solving process. There are a number of cost analysis techniques to use in evaluation and decision-making. Three common analysis methods are cost-effectiveness, cost-benefit, and cost-utility analyses, summarized in Table 2.25. Each of these methods are related and have some similarities; however, there are some distinct differences that should be considered when selecting a cost-analysis method. Certain approaches will have more strengths than others depending on the type and validity of the data available, as well as the types of goals being evaluated.

Cost-Benefit analysis involves determining the Benefit-Cost Ratio (BCR) of a project or alternative. A general definition of BCR is the ratio of the equivalent worth of benefits to the equivalent worth of costs [189]. Using the word *worth* to describe both benefits and costs denotes that each alternative is measured in monetary terms [187, 188]. However, the benefit itself does not have to be, and many times it is not financial. Benefits can range from reduced potential for losses of life, a societal increase in confidence in the defense capability of the U.S. military, increased probability of mission success, etc. These benefits are translated into monetary values via experimentations, correlation studies, surveys, and observed behavior. Once the monetary value of benefits is determined the BCR is calculated by dividing the benefit (B) by the cost (C) to achieve it:

$$BCR = \frac{B}{C}$$

BCR is interpreted as the number of monetary units of benefit for each unit of costs. A ratio greater than one implies that benefits outweigh costs and that program is desirable.

The attractive features of CB analysis is the ability to compare many alternatives with widely disparate objectives, as long as their costs and benefits can be expressed monetarily [188]. This method is suited for certain evaluations of the DOTMLPF-P alternatives because the immediate objectives of each alternative will have varying effects on the nations strategic capabilities. Being able to compare costs of benefits versus physical investment costs brings objectivity into the comparison analysis. Any alternative where benefits to do not exceed costs can be removed from the selection process, and a decision can be made based on which alternative has the highest ratio of benefits to costs. The drawback to CB analysis is that benefits have to be assessed in pecuniary terms. Researchers have found that determining equivalent monetary values of some benefits can be a flawed approach in some cases or not possible in others [190], [188]. For example, how does one put a precise dollar value on the amount of freedom and safety a C-130J rescue mission provides? Safety and freedom are two important objectives, but turning those benefits into a dollar value will undoubtedly require subjectivity. CB analysis is prohibitive when benefits cannot be readily converted to monetary value. In this case CE Analysis is a more commonly utilized technique [188].

Using cost-effective analysis, alternatives are evaluated in accordance with both their costs and effects to produce some outcome. Like CB analysis it involves determining a ratio, but the use of *effectiveness* instead of *benefits* means that native units of evaluation can be used. In this case, the Cost-Effectiveness Ratio (CER) merely requires combining cost data with the criteria or measures of effectiveness (MoEs) that have been carefully considered and evaluated. For example, instructional strategies can be evaluated on the basis of their cost for increasing knowledge recall or physical skill ability by a given amount. From a decision-oriented perspective, the most preferable alternatives are those that show the lowest cost for any given or required increase in effectiveness. By choosing the most cost-effective alternative, resources are made available for other investments. In a modeling and simulation (M&S) environment the CER can be used to eliminate non-viable alternatives. The remaining alternatives can then be compared and the most cost-effective solution(s) selected for recommendation and development. Instead of showing CER results as a single value, the defense acquisition community traditionally utilizes scatter plots, such as the one shown in Figure 2.15 from the Defense Acquisition Guidebook.

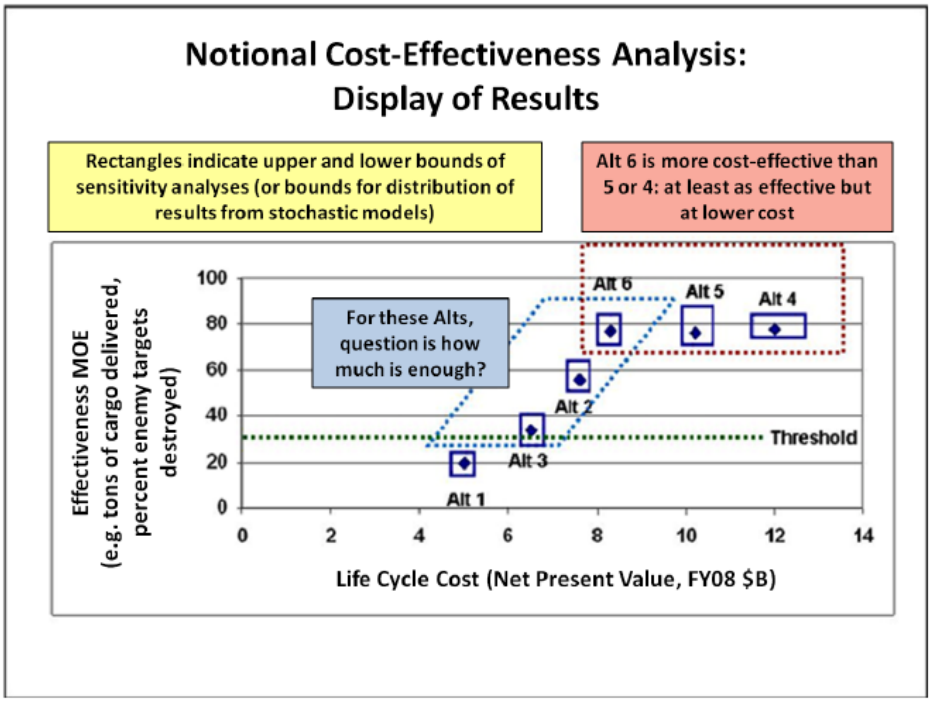


Figure 2.15: Notional Scatter Plot of Effectiveness vs. Cost [191]

Presenting CE analysis in a scatter plot avoids the use of cost-to-effectiveness or effectiveness-to-cost ratios that are more commonly seen in other applications of CE analysis [188]. The rationale for this is provided in the following [191]:

> Note that the notional sample display shown. . . does not make use of ratios (of effectiveness to cost) for comparing alternatives. Usually, ratios are regarded as potentially misleading because they mask important information. The advantage to the approach in the figure above is that it reduces the original set of alternatives to a small set of viable alternatives for decision makers to consider.

Implementation of a CE Analysis (scatter plot or CER) allows the relative comparison of alternatives for fulfilling a particular set of mission capabilities [189]. Sensitivity analyses are usually included in order to quantify the amount of uncertainty present in the cost and effectiveness estimates, and they provide an added dimension to decision making. This can be seen in the boxes surrounding each alternative in Figure 2.15. The decision makers can safely conclude that Alternative 1 would be a poor selection because it does not meet the minimum required threshold for effectiveness. The same argument could be made for Alternative 3, whose measured uncertainty crosses the threshold boundary. However, the issue is not as clear for the remaining alternatives. Alternative 6 would be chosen over Alternative 2 if the increase in effectiveness were assumed to be worth the additional cost. Also, while Alternative 6 is deemed more cost-effective than Alternatives 4 & 5, this cost-effectiveness may come with some types of programmatic risk not captured in the displayed uncertainty estimates [192]. Levin addresses this when describing one of the significant limitations of CE analysis [188]:

> That is, we can state whether a given alternative is relatively more cost effective than other alternatives, but we cannot state whether its total benefits exceed its total costs. That can only be ascertained through a cost-benefit analysis.

To compute a CER, the cost of a given alternative (C) is divided by its effectiveness (E):

$$CER = \frac{C}{E}$$

The ratio is interpreted as the cost required to obtain a single unit of effectiveness. The evaluator defines the units of effectiveness. Once the CER is calculated for each alternative, the alternative that exhibits the lowest costs per unit of effectiveness is selected. Rank ordering the alternatives from smallest to largest CER helps to identify the best alternative. An important caveat when interpreting CER is paying attention to the ratio scales. If the cost and effectiveness values between alternatives have varying scales, the results from applying one alternative to the entire program may not come out as expected. For example, using a simulation environment for training pilots to perform low-visibility soldier extractions may have a cost of $1000 and have a knowledge recall of 85% (a CER or 11.8). Performing this training on an actual aircraft may cost $10,000 and the knowledge recall is 99% (a CER of 101). Based on the decision rule the simulated environment would always appear to be most effective, but if the pilot only trains in a simulated environment they would not be fully prepared in the actual aircraft. This is another case where a utility measure that can weigh effectiveness relative to cost is a better analysis technique because the importance of each factor can be included in the evaluation.

In some cases, the CER is inverted, and presented as an effectiveness-cost ratio (ECR). It is an effectiveness ratio that some evaluators like to use, where effectiveness (E) is divided by cost (C):

$$ECR = \frac{E}{C}$$

ECR is interpreted as the units of effectiveness that are obtained by incurring a single unit cost (generally the dollar value or multiple dollars per unit of effectiveness). When using ECR, the alternatives should be rank orders from largest to smallest. The alternative that provides the greatest effectiveness per unit of cost is the best. If properly interpreted, there is no difference between conclusions drawn based on CER or ECR.

The CE analysis approach has a number of strengths. The simplicity of combining cost data with effectiveness data that is ordinarily available in training evaluations is a major advantage. This method works well for evaluations of alternatives that are being considered for accomplishing a particular training goal. The disadvantage is that it is limited to comparing one criterion at a time, and it does not allow calculation of the overall worth of a program. The conclusion from a CE analysis is that one alternative is *relatively* more cost-effective, but not that the total benefits exceed total costs. Knowing the relative CE of a group of alternatives does not guarantee that the most CE alternative will justify the investment of resources in all situations, but when comparing alternatives that are similar, and when benefits are difficult to put into pecuniary terms, CE analysis does provide some strong objective evidence for decision makers to consider.

Cost-utility (CU) analysis is very similar to CE analysis, but it makes careful attempts to consider individual preferences of one or more criteria [188]. It is rare that a single MoE fully describes the outcome of a training program. For example, training method and resources used may improve the learning outcome in cognitive, affective, or psychomotor skills. Further analysis may uncover that having a live instructor may be the most cost-effective in increasing affective learning, but a simulated environment without an instructor is more cost-effective for building psychomotor ability. CU analysis provides a solution to determine which training alternative is better based on the utility values or preferences assigned, in this example, to the affective and psychomotor LOs. CU analysis refers to the evaluation of alternatives based on their costs and value or utility. When performing CU analysis, researchers solicit the preferences of stakeholders in order to express their overall satisfaction with a single or multiple MoEs. A CE analysis focuses on one MoE at a time, while CU analysis encompasses information on all the MoEs. The common method of combing multiple MoEs into a single estimate is to weigh each MoE using an “importance weighting.” The weightings represent the contribution a MoE has to the overall utility of the decision-maker. Once the overall measures of utility are obtained, the analysis process is executed in the same manner as a

CE analysis. The cost (C) of each alternative is divided by its utility (U) to yield a cost-utility ratio (CUR):

$$CUR = \frac{C}{U}$$

CUR is interpreted as the cost of obtaining a single unit of utility. Rank ordering CURs from smallest to largest allows the decision-maker to choose the lowest cost alternative at a specified utility level or choose the alternative that provides the greatest utility for a given cost. Similarly to CER and BCR, analysts need to consider the scale of each alternative when comparing ratios. The disadvantage to CU analysis is that it can result in two evaluators following the same methodology, yet having drastically different conclusions based on differing stakeholder and decision-maker preferences. Even worse are the situations where stakeholders and decision-makers have varying opinions about the importance weightings of the criteria.

There is not a cost evaluation technique that perfectly forecasts costs and effectiveness. There is always some degree of uncertainty and unknowns will arise later in the design process. With any type of estimation, cost models are usually tailored to a specific discipline or problem domain. There are many cost models and methods for cost estimation. The most common examples of different cost estimating methods are: analogy, engineering build up, and parametric analysis [193]. Less frequently used methods for cost estimation include relying on subject matter expert (SME) opinion, extrapolating actual costs and data from prototypes to predict future costs, or extrapolating actuals costs from learning curves. The engineering build up methodology requires detailed work break down structures and cost data of the various engineering tasks to be completed, and is generally more time-consuming. The parametric method is usually used in the early stages of a program and involves collecting relevant historical data at an aggregated level of detail and relating it to the area to be estimated through cost estimating relationships (CER). These CERs are based on actual program cost history, but are at a very high level so that most detail is lost. Lastly, the analogy method uses actual costs from a similar program with adjustments based on the level of difficulty and other differences the new program may have. This method is normally used when there are few data points available, such as in the early phase of a program. For this research effort the cost estimating method will be primarily used, and where details are available, the parametric method. Table 2.25, provides a quick reference summary of the three cost analysis methods discussed.

Table 2.25: A Summary of Three Approaches to Cost Analysis [188]

Types of Analysis	**Analytical Question(s)**	**Measure of Cost**	**Measures of Outcome**	**Strengths of Approach**	**Weakness of Approach**
Cost-Benefit Analysis	• Which alternative yields a given level of benefits for the lowest cost (or the highest level of benefits for a given cost)? • Are the benefits of a single alternative larger than its cost?	Monetary value of resources	Monetary value of benefits	• Can be used to judge absolute worth of a project. • Can compare CB results across a wide variety of projects in training or other areas (e.g. facilities, materiel)	• Often difficult to place monetary values on all relevant system benefits
Cost-Effectiveness Analysis	• Which alternative yields a given level of effectiveness for the lowest cost (or the highest level of effectiveness for a given cost)?	Monetary value of resources	Units of effectiveness	• Easy to incorporate standard evaluations of effectiveness • Useful for alternatives with a single or small number of objectives	• Difficult to interpret results when there are multiple measures of effectiveness • Cannot judge overall worth of a single alternative; only useful for comparing two or more alternatives
Cost-Utility Analysis	• Which alternative yields a given level of utility at the lowest cost (or the highest level of utility at a given cost)?	Monetary value of resources	Units of utility	• Incorporates individual preferences for units of effectiveness • Can incorporate multiple measures of effectiveness into a single measure of utility • Promotes stakeholder participation in decision-making	• Sometimes difficult to arrive at consistent and accurate measures of individual preferences • Cannot judge overall worth of a single alternative; only useful for comparing two or more alternatives

2.10.1 Sensitivity and Uncertainty Analysis

As a best practice, sensitivity analysis should be included in all cost estimates, because all estimates have some uncertainty [193]. Sensitivity analysis attempts to isolate the effect of changing one variable at a time, and helps determine the amount of risk for a particular solution. The results provide a range of costs including best and worst case approximations. To examine the effect of changing more than one variable at a time, uncertainty analysis is necessary. Uncertainty is added to cost data because variations in cost data occur due to errors in historical data and CERs, variations associated with input parameters, errors with analogies and data limitations, data extrapolation errors, and optimistic learning and rate curve assumptions. Past data is not always relevant in the future, and even recent data from one training activity will not necessarily translate exactly into a new training session, so cost estimates can also contain a vast amount of uncertainty. There may not be enough information available this early in the acquisition process to create a frequency distribution, but as part of the uncertainty and sensitivity analysis a range of costs with a specified confidence level can be developed and that is recommended as part of this methodology. This will add robustness and greatly aid the decision maker.

Summary cost analyses can be used to determine the best resource allocations to maximize training effectiveness, answering RQ 2.2. The previous review of literature to determine how to quantify training benefits in terms of effectiveness and increased knowledge, skills and attitudes (RQ2.1 and 3) has found that following variables influence learning in training: LOs in three learning domains, instructional methods and media, and DIFE analysis. Incorporating decision maker preferences towards the relative importance of these factors is desired. Therefore MPEET uses CU analysis to evaluate the cost and effectiveness of training alternatives. CU analysis allows for the combining of multiple criteria and decision maker preferences into a single utility measure or criterion. The failure of CU analysis to determine the overall worth of a training alternative is not a concern when performing assessments using MPEET, because MPEET is used to evaluate the cost and effectiveness of a training program and generate robust training alternatives. A CB analysis is more appropriate when trading training as an alternative to materiel and non-materiel solutions.

2.11 Measurement Criterion

A criterion, in the most simplistic form, is a standard by which something is measured [21]. In terms of modeling and simulation (M&S), it can be defined as the standard against which test instruments are correlated to indicate the accuracy with which they predict human performance in some specified area. For evaluation purposes a criterion is used as a measure to determine the adequacy of a product, process, behavior, and other conditions. When there are multiple metrics being evaluated, each criteria can be independently assessed or combined into an overall evaluation criterion (OEC). McCabe & Butler developed measurement techniques to quantify the architectural complexity of different software designs. Their recommendation was presented in the context of architectural complexity and included criteria that can be used to judge the applicability of any proposed measure of a complex system [194]. Their criteria, listed below, works not only for complex software design, but also for developing a criterion to evaluate training system designs:

1. *The metric intuitively correlates with the difficulty of comprehending a design i.e.,* when we view large complicated designs, the metric should yield a high number. Designs we intuitively deem as simple should have a relatively low number.

2. *The metric is objective and mathematically rigorous*. The same design viewed at two different times or by two people should yield the same complexity.

3. *The metric should be related to the effort to integrate the design*. The proposed metric should correlate directly with the cost and effort experienced in the integration phase.

5. *The metric and associated process should be automatable.*

These requirements they may seem obvious or common sense; however, ensuring that these criteria are adequately addressed is important to confirming the overall utility measure.

2.12 Multi-Attribute Utility Theory

A training program produces outcomes in a multitude of learning domains: cognitive, affective, and psychomotor abilities. Within each category, there are a variety of subcategories or competency levels. For example, cognitive learning can be divided into knowledge and comprehension, application and analysis, and synthesis and evaluation using Bloom's taxonomy. Literature on utility theory refers to each of these MoEs as an "attribute." Stakeholders may derive utility from – or have preference for – each attribute. Multi-attribute utility theory provides a set of techniques for quantifying the utility derived from individual attributes, and combining the utility from each attribute into an overall measure of utility [188]. The general tool for carrying out these two tasks is called the multi-attribute utility function. Utility represents a numerical measure of "goodness" or relative preference [195]. The utility function assigns a numerical value to each attribute based on stakeholder inputs. It provides a structured and systematic method for identifying and analyzing multiple attributes to derive a common basis or criterion for CU analyses and decision-making.

Assume each attribute of a training system design program is simply referred to as x_1, x_2, x_3, and so on until the final attribute, x_m. The attributes would all be measured in their "natural" units. For example, increased knowledge recall might be expressed in percentage, asset cost in dollars, and time in months etc. Expressing each attribute on a new scale, a common "utility" scale, is a way to describe the strength of preferences for a given increase in competency levels, improvement of knowledge recall, or for a change in any of the attributes. A single-attribute utility function is estimated in the form of $U_1(x_1)$, $U_2(x_2)$, $U_3(x_3)$, through $U_m(x_m)$. The notation represents the utility, U_m, produced by the attribute, x_m. How to convert each attribute to a utility scale is discussed shortly. Once single-attribute utility functions are determined, they can be combined into an overall measure of utility or criterion. The mathematical tool for combining the criteria is called the multi-attribute utility function. Prior to summing the single-attribute utility functions, each is multiplied by an "importance weight" ($w_1, w_2, w_3, \ldots. w_m$). These weights reflect the relative importance of each attribute to the stakeholders and/or decision-makers. The importance weights for all attributes should sum to one. The overall additive utility or criterion from a alternative (and its *m* attributes) is expressed as [188]:

$$U(x_1, \ldots, x_m) = \sum_{i=1}^{m} w_i U_i(x_i)$$

This type of multi-attribute utility function is referred to as additive because it involves simply adding up the weighted utilities of individual criteria. It can be easily applied in a wide variety of analyses and it makes intuitive sense to most people. The major disadvantage of an additive multi-attribute utility function is that is assumes the preferences for each attribute is independent of the preferences of the other attributes. For example, assume x_1 is cognitive achievement, and x_2 is psychomotor ability. Overall utility increases with increasing amounts of either attribute. The additive utility function implies that the amount of utility produced by enhanced cognitive ability is independent of the level of psychomotor skills achieved. Meaning whether the cognitive competency level is low or high, psychomotor skills will still yield the same amount of utility. Conversely, the amount of utility produced by greater psychomotor ability is independent of the level of cognition. If assessing a training program that has several dependent attribute preferences, the analyst may use other forms of multi-attribute utility functions that are of increased complexity. Keeney and Raiffa provide industry standard references on different forms of utility functions [196, 197].

2.12.1 Single-Attribute Utility Function Assessment Methods

A common mistake engineers make when using utility functions is a failure to convert the attribute value into a utility value [198].There are three common approaches for converting attributes into a single-attribute utility scale. They are proportional scoring, the direct method, and the variable probability method. Proportional scoring is a simple dimensionless linear rescaling of each attribute. This rescaling can be done via graphical or mathematical means [188]. To graphically rescale an attribute, plot each attribute value along the x-axis, ranging from the lowest value to the highest. The utility scale is then plotted on the y-axis, ranging from zero to 100. The lowest value on the x-axis becomes zero utility, and the highest value is set to a utility of 100. Using these two points, all other utility values can be interpolated. The utility scale does not have to range from zero to 100. The lowest and highest attribute values can be set to any value as long as these same values are used to interpolate the attribute values for the in between points. It is important that the attribute values all be assessed on the same utility scale. Once all the attributes have a calculated utility value, a line can be drawn connecting each of the points. From left to right, an increasingly straight line implies that increasing the attribute value results in increases in utility, as shown on the left of Figure 2.16. In some cases, as the attribute value increases the utility may increase and then plateau implying that at higher attribute values, the gains in utility are smaller. An example is shown in the right plot of Figure 2.16.

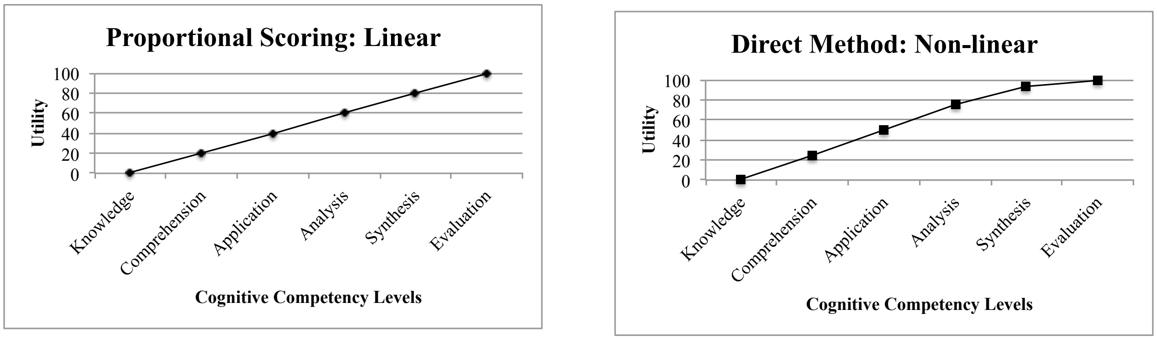


Figure 2.16: Example Utility Functions for Cognitive Competency Levels

The same utility scored can be derived mathematically, without graphical analysis. The formula for a proportional scoring single-attribute utility function is:

$$U_i(x_i) = \left(\frac{x - Lowest}{Highest - Lowest}\right)_i * 100$$

Proportional scoring does not rely on the expressed preferences of stakeholders or decision-makers. It assumes that increasing amounts of an attribute have a linear or non-linear relationship with utility.

To include direct input from individual stakeholders on the utility curve from varying amounts of an attribute, the direct method can be used. To apply the direct method, one identifies the low and high values on the relevant attribute scale. The lowest and highest values are assigned utility values of zero and 100, respectively. The stakeholder is then directly asked to rate their preferences for the middle attribute points relative to the endpoints of zero and 100. A regression analysis of the data points can then be run to find a line of best fit. The line of best fit then becomes the single-attribute utility function.

Another assessment method that incorporates stakeholder preferences for varying amounts of an attribute is the variable probability method. Unlike the direct method, it uses a decision tree to choose between different options. The variable probability method highlights the amount of risk the stakeholder or decision-maker is willing to take in order to attain the highest probability of the best attribute value. Displayed in Figure 2.17, the decision-maker is presented with two choices. They can choose the risky option that has a probability of p to obtain the highest attribute value, but also has a probability of $1-p$ to get the worst value. Or they can choose the less risky and certain option, with a probability equal to one, of obtaining the middle attribute value. To assess the utility of the middle score, stakeholders are asked to choose the probability (p) that makes them indifferent between the risky alternative and the riskless alternative. For example, assume the attribute being assessed is cognitive ability with an initial probability of 0.99. This means there is a 0.99 probability that trainee's will obtain a competency level of synthesis and evaluation (best case). Conversely there is a 0.01 probability that the trainee's only reach the knowledge and comprehension level (worst case). The decision-maker can choose to gamble based on the 0.99 probability of achieving the best attribute value, or they can choose to accept that the trainee's will reach a competency level of application and analysis (middle level) for certain. The risky option is likely attractive to most stakeholders, but if the initial probability was reversed to 0.1 for best case, and 0.99 for worst case, most decision makers may prefer the middle score. Between 0.99 and 0.1 is a probability value that stakeholders would feel indifferent choosing either the certain middle score or taking a risk to obtain the best attribute value. Whatever the probability value is that makes individuals feel indifferent, that probability is interpreted as the utility value for that attribute. With the endpoints set to zero and one, the probability can be multiplied by 100 to use the zero to 100 utility scale. The same process of finding the probability value that makes the decision-maker indifferent to another middle value is repeated for each of the attribute values between the lowest and highest numbers. This creates pairs of attribute values and their associated utilities, while taking into account the amount of risk and uncertainty the stakeholders are comfortable with. These data points can be plotted and connected similar to the proportional scaling method to view the relationship between the attribute values and the associated utility.

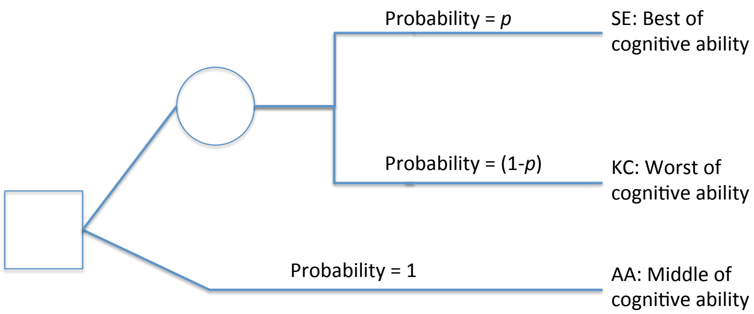


Figure 2.17: Sample Utility Function with Variable Probability Method

Proportional scaling, the direct method, and the variable probability method are three options for determining the single-attribute utility function. Proportional scaling is the easiest to implement and does not consider stakeholder preferences. The direct method and variable probability both incorporate decision-maker preferences, and the variable probability method also includes risk and uncertainty. MPEET will use the proportional scaling method to find the utility function for each attribute. In training system design the desire is for each lesson is to result in maximum student performance (highest attribute value) for the corresponding learning objective. The utility for each attribute can be proportionally scaled based on where each attribute value falls within the range of lowest to highest values.

2.12.2 Importance Weightings Assessment Methods

Estimating the relative weight or *importance* of each attribute in other overall utility function is best accomplished by directly involving the stakeholders. The direct method and variable probability method can be used as described above. Using the direct method, ask the individuals involved to allocate a total of 100 points across each of the attributes. The attributes that are relatively most important are given a larger weighting, and all the weights must sum to 100. Importance weights are applied on a zero to one scale. Normalize the stakeholder weightings by dividing each by 100.

A variation of the direct method asks the decision-makers to rank order the attributes by relative importance. The most important attribute is assigned a value of 100, and the remaining attributes are assigned values in relation to their ranked importance level. A value of 50 implies that the attribute is half as important as the most important attribute. To normalize these values, each value is divided by the sum of all the values. If there were only two attributes with a value of 100 and 50, the normalized weightings would be 0.67 (100/150) and 0.33 (50/150).

The variable probability method can also be used to determine the importance weights from decision-makers. As described above, the probability that makes individuals indifferent between the certain and risky options is determined for an attribute. The lower branch of the decision tree is then replaced for each of the attributes until the resulting probabilities are found for all attributes. These probabilities are used as the importance weights, but not before verifying that they sum to one. If their sum is close to one but not exact, then dividing each value by the total sum can normalize them. If their sum is not close to one, it is an indication that the additive utility function does not adequately represent the stakeholder's preferences [199]. A more complex version of the utility function that includes interactions between attributes may be necessary [197].

Variations of the direct method, and the variable probability method both provide means for determining the importance weightings from stakeholders. Direct method techniques are straightforward in explanation to decision-makers and analysts. Their results are easily understood and applied [188]. Sometimes stakeholders have a hard time with the probabilistic choices and make contradictory decisions when faced with the risky options. Either method can be prone to error and disagreement when there are multiple individuals providing input, therefore scholars to not make general recommendations in favor of any method. When possible, it is recommended to present multiple methods to the decision-makers and look for any inconsistencies of results. The process may take several iterations before the true preferences are discovered. The drawback in using multiple techniques is the time required of the stakeholders and analysts in ensuring that all individuals understand their own preferences. In this research the direct method is used to find the importance weightings. The decision-maker is requested to rate each attribute on a zero to 100 scale (the sum totaling 100) based on the relative importance of each attribute during the design of the training program. A sensitivity study is conducted to determine how important the weightings are in predicting the most effective training alternatives.

In summary, determining the CUR requires an estimate of utility for each alternative. A single-attribute utility function requires that the attribute value be converted to a common utility scale using proportional scaling, direct method, or variable probability method. Importance weightings are gathered from decision-makers based on their preferences of the importance of each attribute. The importance weightings are incorporated into the utility function by multiplying the importance weight by the utility value. Each single-attribute function is then combined into an overall utility measure or criterion, called the multi-attribute utility function. The additive multi-attribute utility function presented is one of several techniques for combining single-attribute utility functions. It is the simplest method and works very well for independent attributes. If the attributes are dependent, a more complex multi-attribute utility technique is necessary. When dealing with cost analyses it is important to assess whether the ranking of alternatives is sensitive to assumptions made during the analysis. In a CU analysis the importance weighting can have a significant impact on the resulting *best* alternative. Human decision-making and preference selection is not certain; therefore, a sensitivity analysis of importance weightings is included as a step in MPEET to assess how robust the training method alternatives are to subjective rankings. Cost analyses can be used to determine the best resource allocations to maximize training effectiveness, answering RQ 2.2.

2.13 Summary

In response to RQ1, what is an appropriate method of measuring training effectiveness during early phase defense acquisition, a literature survey of existing training effectiveness models was conducted. The literature review provided insight and partially answered this question. Five criteria were determined necessary for a method to be used to evaluate training effectiveness during the defense acquisition CBA analysis phase. Specifically, the method must: 1) connect training results to mission specific goals, 2) be based primarily on objective data (can be supported by subjective data), 3) account for variation of skill levels, 4) include uncertainty analysis, and most importantly, 5) can be used to estimate, rather than simply evaluate, performance results after training is complete. The evaluation methods of Deitchman, as well as Bahlis and Tourville, include the majority of the criteria necessary to meet the primary research objective, but both were missing at least one criterion, either predicting training effectiveness and/or the use of uncertainty analysis. The research focus thus concentrated addressing the gaps in predicting training effectiveness by assessing training system design, and determining the associated cost and effectiveness via andragogy principles.

The second research question, RQ2, contained two parts regarding the measurement of training effectiveness. The first part, RQ2.1, asked how to quantify the benefits of soldiers training in terms of effectiveness. The second part, RQ2.2 questioned how resources should be allocated to gain maximum training effectiveness. These questions were partially answered through the literature reviewed from education, training, and psychology fields. Taxonomies for describing the stages of learning and competency were reviewed and recommendations were made regarding which taxonomies are best for use in MPEET. Instructional strategies were examined to determine which training methods and media resources resulted in maximum trainee knowledge retention and ability. This literature review also provided answers to research question three, how to quantify increased knowledge, skills, and attitudes in training system design.

Based upon the literature review, an overall evaluation criterion was created that enveloped the various attributes of training system design and importance weightings were assigned from the decision-maker preferences. Ten attributes were discovered as necessary to predict the effectiveness of a training program: learning objectives in the cognitive, affective, and psychomotor domains, instructional methods, instructional media, use of an instructor as a resource, and difficulty, importance, frequency, and consequence of error ratings. These attributes can be grouped into three categories as shown in Figure 2.18: Learning Objectives, Instructional Strategies, and Criticality Ratings.

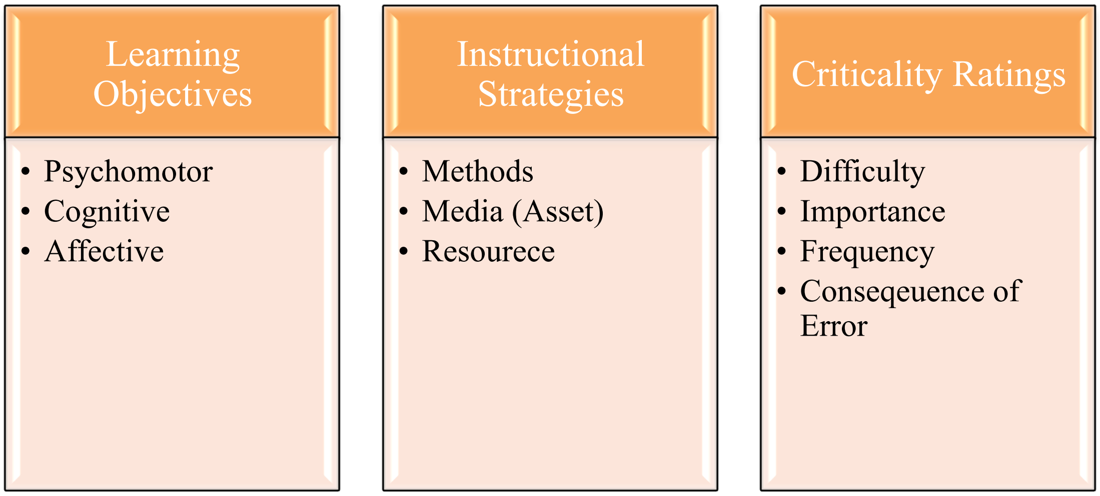


Figure 2.18: Predicting Training Effectiveness OEC Criteria

CHAPTER 3

APPROACH

The necessary criteria identified in the literature review for predicting training effectiveness consist of variables from multiple industries and fields of study. To integrate these variables in a cohesive manner the author used processes and tools gained from professional experience in systems engineering. The International Council on Systems Engineering (INCOSE) provides three definitions to represent systems engineering. The first definition describes the approach used in the development of a new training effectiveness evaluation method that will meet the requirements necessary for use during the JCIDS acquisition process:

> *"Systems engineering is a discipline that concentrates on the design and application of the whole (system) as distinct from the parts. It involves looking at a problem in its entirety, taking into account all the facets and all the variables and relating the social to the technical aspect [200]."*

The above systems engineering definition is applicable in this research because the fields that are being combined involve technical or mathematical based practices with psychological and educational theory. The instructional system design and development that occurs during the design process of a training system is typically performed by human factors engineers, cognitive scientists, psychologists, and SMEs in instruction and training. They are responsible for ensuring training is designed to impart knowledge and skills in the most effective, efficient, and engaging manner. To evaluate the effectiveness of training early in the defense acquisition process, the training system design must be tested using predictive analysis and probability theory. System engineers commonly use modeling and simulation (M&S) to perform these analysis techniques when evaluating processes and products. System engineers use modeling and simulation to determine system requirements; predict system performance; calculate process input, outputs, and throughput rates; support trade studies; estimate cost and schedules; and optimize processes. M&S are two valuable tools of the systems engineering trade or discipline that can reduce the cost of a project, improve the efficiency of a process, and provide a safe mechanism and environment for experimenting. “A model is a physical, mathematical, or logical representation of a system, phenomenon, or process. There are many classifications of models. Models may be predictive or interpretive, physical or mathematical, numerical or analytical, and continuous or discrete [201].” System engineers have used tools such as Excel, Mathcad, Analytica, and common programming languages to develop both physical and mathematical models. “A simulation is the implementation of a model over time. A simulation is an imitation of a system based on knowledge or assumptions about the behavior of the parts of that system, with the purpose of obtaining insight into the behavior of the whole system.” Simulations help bring insight to models and show how a particular system, object, or phenomenon will behave. Like the models they represent, simulations can be continuous or discrete. Tools such as Matlab, Maple, and Simulink are commonly used for their simulation capability. M&S has been used in traditional system engineering analysis for over 40 years, but the recent advances in technology in the last two decades have improved capabilities and results [200].

A new paradigm has emerged within the DoD regarding the use of M&S in the acquisition process [202]. Previously M&S was considered a tool just to be used in the design of products such as weapon systems. With the increasing advances in M&S tools and the decreased availability for resources the acquisition community has begun integrating the use of M&S throughout all phases of the acquisition cycle. Models created during the acquisition phase, when properly incorporated in to a program, tend to evolve as the program progresses. Using M&S has yielded benefits of reduced risk in cost, schedule, and performance. M&S are currently used for a number of applications in the DoD, particularly to help support arguments presented in analysis of alternatives to justify proceeding with system development [203]. It is used to augment operational test, design, and evaluation, and to provide insight into data collection. When information is limited and resources are constrained M&S provides the means for conducting "what if" drills when exploring new concepts or stressing a system's performance. It can also be used to identify design flaws, thus reducing and delaying the need for physical prototypes. Critics of modeling and simulation highlight that if not used with care and a proper understanding of the limitations of experimentation in the M&S environment the results can lead to ineffective or unreliable systems [204].

The predictive capability and the ability to combine and interpret integrated variables make M&S an ideal tool to use for the development of new methodology to evaluate training effectiveness during the JCIDS acquisition process. With the incorporation and acceptance of M&S based results by the defense acquisition community using these systems engineering tools aligns with the current practice and therefore will reduce the resistance of incorporating a new method to the CBA process. When processes and tools change simultaneously it is typically harder and takes longer for people to adapt. Using common tools and software will facilitate acceptance, integration, and application of this new methodology into CBA process guidance.

Another valuable systems engineering tool that assists with the collaboration of various disciplines during all phases of a product life cycle is the integrated product and process development (IPPD) process. IPPD in a design context can be defined as a "management methodology [or strategy] that incorporates a systematic approach to the early integration and concurrent application of all the disciplines that play a part throughout the system's life cycle [205]." IPPD is a key enabler to obtain producible and affordable products. At the core of the IPPD concept is the focus on the customer and meeting the customer needs. Although no single implementation strategy exists for IPPD, the generic IPPD process is a disciplined, systems engineering approach that entails an iterative scheme between customer requirements, products, and associated processes. In the Department of Defense's *Guide to IPPD*, key tenets were identified to effectively implement IPPD and include: customer focus, concurrent development of products and processes, multidisciplinary teamwork, robust design and improved process quality, and proactive identification and management of risk [206]. Although the DoD guidance does not provide a structured approach on how to implement IPPD, other researchers and industry experts have. To implement the IPPD strategy, Schrage and Mavris proposed four elements to guide the development of a product within the IPPD framework as evolved out of Concurrent Engineering principles [207]. The elements are quality engineering methods, systems engineering methods, a computer integrated environment, and top-down design decision support processes as shown in Figure 3.1.

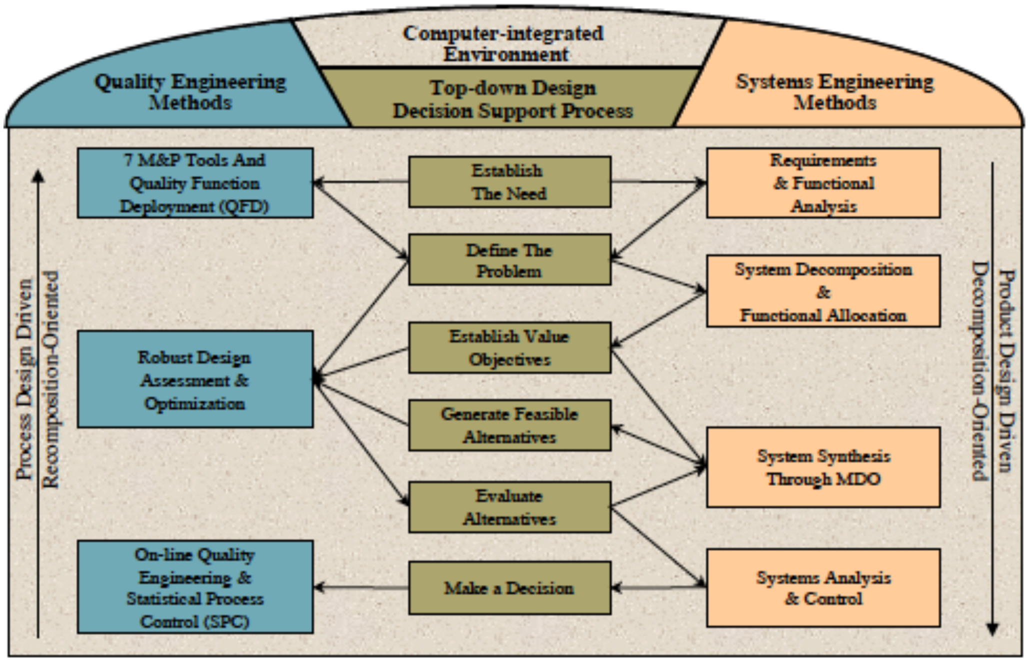


Figure 3.1: IPPD Implementation [207]

At the center of the IPPD implentation approach developed by Schrage and Mavris is a top-down design decision support process. Decision support is an essential element that can support a trade-off process and can be used to focus efforts on design goals. It provides a logical and balanced means for including factors that must be considered when making a decision. The steps to execute their approach, as depicted in Figure 3.1, begins with “Establishing the Need” and concludes with “Make a Decision”. The techniques and methods required to execute each step are listed under the quality and systems engineering methods. The arrows into the top-down design decision support represent the trade-off assessments and information flow to accomplish each step. The primary design iteration loop in the IPPD approach consists of generating feasible alternatives, performing a robust design assessment, evaluating the alternatives, and then applying Multidisciplinary Design Optimization (MDO) techniques to identify the most robust design alternative. Robust design is defined as the “systematic approach to finding optimum values of design factors, which result in economical designs with low variability [208].” The most robust design alternative in the newly developed methodology for assessing training is the most cost-effective training alternative. The IPPD approach presented by Schrage and Mavris has been applied to numerous aerospace vehicle concepts in graduate courses involving, introduction to Concurrent Engineering, fixed wing design, and advanced design methods in the School of Aerospace Engineering at the Georgia Institute of Technology [207]. The IPPD implementation framework by Schrage and Mavris offers a generic top-down design decision support process that is used as the framework for the proposed methodology to evaluate training effectiveness. The core iterative design loop of generating and evaluating feasible alternatives is beneficial in creating a robust set of training alternatives that will result in the most cost-effective training system design.

Having the proper tools to generate and evaluate alternatives is important, but being able to communicate analysis results by providing the decision-maker a visual means by which informed decisions can be made is just as imperative. The National Visualization and Analytics Center (NVAC) established in 2004 by the U.S. Department of Homeland Security defines visual analytics as “the science of analytical reasoning facilitated by interactive visual interfaces [204]." When generating and evaluating alternatives with today’s improved computer technology and M&S software, the data output can become overwhelming. Visual analytics combats this problem with tools and techniques to help analyst make sense of information and uncover key insights such as patterns and data trends used to draw conclusions from analysis results. NVAC and software companies are working to make visual analytics tools and techniques the 21st century's answer to information overload. One of the major goals of visual analytics is to facilitate analytical reasoning. Analytical reasoning techniques are defined by the NVAC to be "the method by which users obtain deep insights that directly support situation assessment, planning, and decision making [204]." By taking advantage of a broad range of visual representations and interaction techniques the analyst view data in multiple formats and interact with the data in real-time. According to the NVAC, "Interaction techniques are required to support the dialogue between the analyst and the data. Visual analytics facilitates high-quality human judgment and requires a limited investment of an analyst's time." Visual analytics has been applied to many industries and organizations. In addition to homeland security and the defense industry, businesses such as healthcare, telecommunications, marketing and education can all potentially benefit from the use of visual analytics. There are an increasing number of software packages being developed to fulfill the needs of visual analytics. As full exploration of visual analytics software is outside the scope of this work, the statistical analysis package JMP® will be used here to aid in visual analytic analysis. JMP® has built in modeling and simulation functions, including probability theory and robust design techniques, that are can be tailored through program codes to meet user specific needs. Due to its capability, availability to the author, and the author's familiarity using JMP® allows the methodology programming, analysis, and visual graphics to be completed within one software tool.

CHAPTER 4

A METHODOLOGY TO PREDICT AND EVALUATE THE EFFECTIVENESS OF TRAINING (MPEET)

The Methodology to Predict and Evaluate The Effectiveness of Training, MPEET, addresses two missing criteria in existing training effectiveness models, the ability to predict training effectiveness and the quantification of uncertainty in training evaluations. MPEET evaluates the cost and effectiveness of an existing training program, and creates an alternative training design that is based on decision-maker importance weightings for each of the input variables. The decision-maker is provided with the option to select the original training program, incorporate specific changes from the alternative program, or implement the recommended alternative in its entirety.

MPEET has been formulated to evaluate the effectiveness and cost of a training program, test the legitimacy of the findings from the literature survey in Chapter 2, and yield new observations that can enhance training system design and evaluations. MPEET is most useful when post-training evaluation information is not available and the cost to collect actual data is infeasible. This is a common situation when conducting CBA analysis during the early phases of the defense acquisition process. However, MPEET is not limited in application to the DoD acquisition process. MPEET can be used in any situation where the assumptions (below) are met, and there is a desire to enhance the instructional design process by adding a verification step to objectively (considering the weightings placed on cost versus effectiveness) determine how well the instructional strategies used in the training program design meet the required learning objectives. The development of MPEET combines two of the primary elements of instructional design, learning objectives (LOs) and methods, with the systems engineering decision-making process. The instructional design elements of learning objectives and methods form the basic components for ensuring and increasing human performance during training. MPEET should not be used in place of, or as an alternative to the instructional design process. The goal of instructional design is to make learning more efficient, more effective, and less difficult [63]. MPEET predicts how well the design of a training program meets those goals. MPEET uses only a fraction of the information that is generated from any comprehensive training system design analysis. As discussed in the assumptions below and in section 2.7.1, there are numerous learning variables MPEET does include. These variables have an impact on training effectiveness and are included as part of the instructional design tasks/needs and learner analyses. Figure 4.1 illustrates where MPEET fits within the instructional design process.

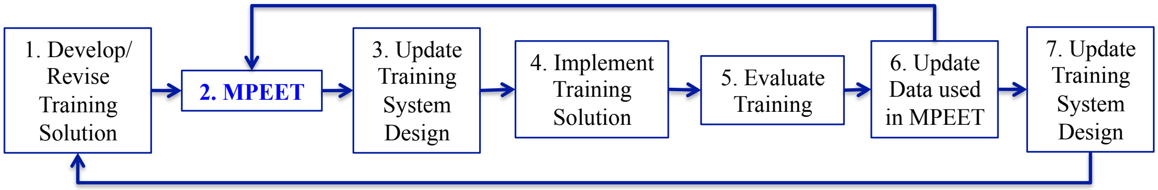


Figure 4.1: The Instructional Design Process with MPEET

4.1 MPEET Assumptions

MPEET provides a means for understanding and showing the training alternative space. It evaluates the effectiveness and efficiency of a baseline training system design, and develops strong alternatives to training tasks. It evaluates the cost and effectiveness of the training alternatives early in the design phase and compares the alternatives to the baseline. It includes the preferences of decision-makers by weighting the metric criteria. There are limitations on the use and applicability of MPEET, and these are characterized by the following assumptions. These assumptions are also listed as MPEET is presented in the next section.

1. MPEET assumes that time is constant. The time allotted for a student to accomplish a particular LO can vary based on the instructional method used and personal learner style. In the future, MPEET may evolve to include variances in time for each training task. One can then observe how changes in time correspond to training effectiveness. First, however, a demonstration is required as to how well MPEET works. For this research effort, the total length of training and the daily hours spent in training are defaulted because the focus is on the ability to predict the effectiveness of a training system design, not to design the system. It is assumed that the instructional designer knows the correct number of hours required to gain the skills for each LO. The hours input for any lesson are used without modifications or variance, to determine the cost and effectiveness of that task. MPEET does not evaluate if the hours for each lesson are valid. That analysis is performed during the instructional design process.
2. MPEET assumes that an hour spent using one training method equals the same amount of time (an hour) of any other type of method. There are studies that claim, if a training lesson takes one hour via lecture and discussion, the time required to teach that same lesson via interactive courseware (ICW) is reduced significantly [209]. Other research claims that students will repeat all or some parts of the lesson using ICW resulting in the student spending the same or even

more time in training. This does not mean that the learning effect is the same between methods. If using MPEET, and there is data available that proves there are time differences between training methods for the training program being assessed, then use the real data. Otherwise, include this assumption.

3. MPEET selects only one instructional method type for each training task. The lesson goals may require multiple lessons, methods, and pre-requisites, but only one instructional method can be selected per lesson title. If a task requires the use of multiple training methods, then the task can be broken into multiple tasks. Each of these subtasks can be created based on the time spent using each instructional method. For example, if a pilot is being trained to perform an aerial refuel using a combination of lecture and discussion then simulation in two hours, this can be broken into two tasks. The first being mission preparation for aerial refuel via lecture and discussion, one hour. The second task would be performing aerial refuel in simulator, one hour. MPEET would correctly analyze the two subtasks. With the tasks combined, MPEET would evaluate based on the first method input and would ignore the second method.
4. MPEET will not select a less effective training method than what is already included in the baseline. The philosophy behind the development of MPEET is maximum training effectiveness. Therefore, a constraint is placed on the selection of training methods to ensure the alternative training program is always equal to or more effective than the baseline. For example, if a pilot is learning the procedure for landing an aircraft using the learning pyramid for effectiveness evaluation, a lesson taught via computer based training (CBT) is approximately

20% effective. Assuming all lesson pre-requisites are met, if the same lesson is taught using a high fidelity simulator the effectiveness increases to 75%. If the lesson was taught via lecture the effectiveness decreases to 5%. MPEET will reject the lecture option because it is less than the baseline method of 20%.

5. MPEET assumes that the desire for the training program being assessed is to reach maximum competency levels for all learning objectives across all learning domains.
6. MPEET assumes that the developmental costs of a resource or asset are included in the hourly cost value. This is an important assumption to remember. If new software has to be developed, existing software modified, or new equipment purchased and/or modified and these costs are not included in the resource cost, then the affordability prediction will be incorrect. This can be addressed in the implementation of MPEET for a particular case study by including a penalty or additional cost for new or modified resources.
7. The use of MPEET must follow the creation of a training system design that used the instructional design process. Three primary elements of the instructional design process are available pre-training: learning objectives, instructional methods, and learner specific variables. MPEET only considers learning objectives and instructional methods. Learner specific variables are equally important as discussed above and in detail in section 2.7.1. Decisions made within MPEET in terms of which instructional methods are most effective require an understanding of the trainee population. Some information and references are provided about student age and learning preference but MPEET does not include

any assessment of individual learning styles, motivations, previous trainee knowledge etc. Because of this limitation in the capability of MPEET, for the most accurate results when evaluating and predicting any training program the user must ensure that the training program was developed following best practices for instructional design. If not, the results from MPEET must be applied with consideration that particular information about the trainee population is not included.

4.2 MPEET

The MPEET process consists of five major steps similar to the systems engineering decision-making process and is summarized in Figure 4.2. The first step involves defining the training program requirements in terms of learning objectives and competency levels for each training lesson. In step 2 the training strategies are defined. A mapping of the instructional methods, media, and resources that correspond to each learning outcome is developed. This creates a portfolio of possible training alternatives for each objective. The effectiveness of each instructional method is also determined in step 2 by means of experimentation or SME consultation. In step 3 feasible training alternatives are generated by translating all the cost and effectiveness variables into a utility scale, collecting decision-maker preferences for each attribute, and determining the utility of each of the training alternatives. The training alternative with the lowest cost-utility ratio is selected and assigned for the recommended training program. Step 4 evaluates the baseline and alternative training programs. The sensitivity of each of the cost and effectiveness variables, as well as the probabilistic variable sensitivity is determined. Based on the sensitivity evaluation results modifications to the OEC criteria and/or importance weightings used in step 3 may be required. The last step is to present the evaluation results from the original and recommended training programs to the decision-maker. The decision-maker is provided with the option to select the original training program, incorporate specific changes from the alternative program, or implement the recommended alternative.

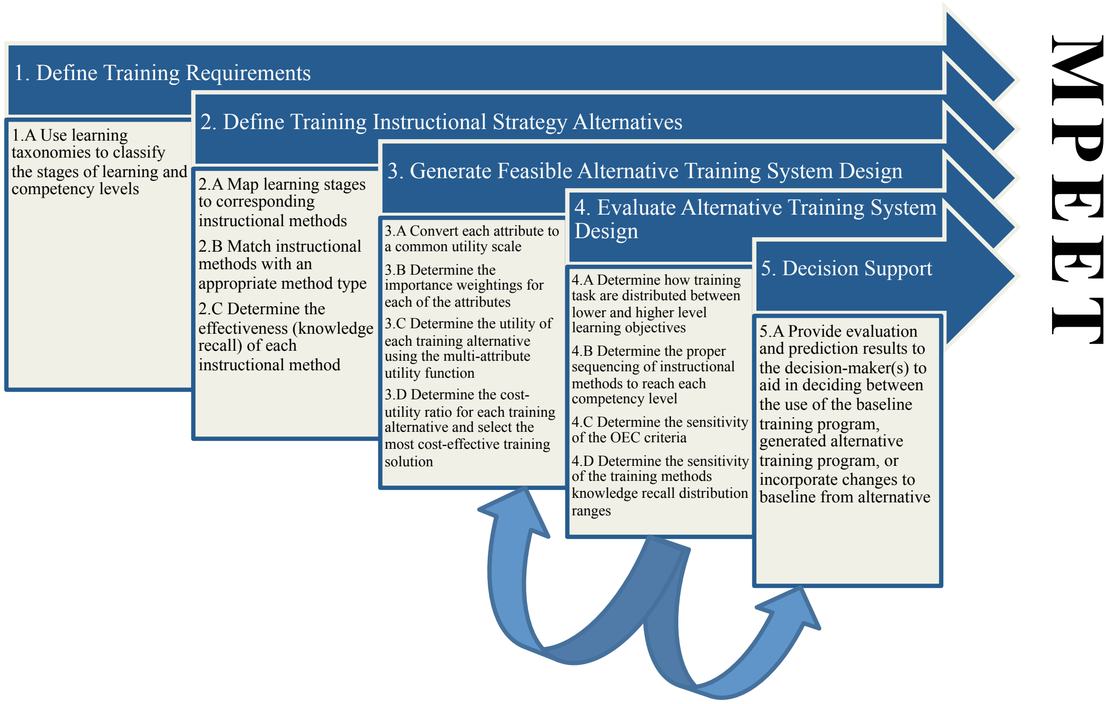


Figure 4.2: MPEET

4.2.1 Step 1: Define Training Requirements

Use learning taxonomies to classify the stages of learning and competency levels (1.A)

For each lesson within the training program identify the learning objective domain and competency level. Fifteen taxonomies have been presented that cover cognitive, psychomotor, and affective learning objectives. Based on the information available for the training program, the reader can select which taxonomies best describe the learning process under evaluation. The author recommends the following taxonomies for general use. To classify psychomotor LOs, use Ferris' taxonomy: Recognition, Handling, Basic Operation, Competent Operation, Expert Operation, Planning, and Evaluation [96]. It logically progresses through the motor skill development process and can be easily applied to any crew position. Testing using this taxonomy will emphasize the ability of the crew to perform procedural tasks and plan and improve upon what has been taught. Using this taxonomy will also highlight whether or not the training program is solely focused on teaching mechanical skills and lacks higher order thinking. To classify cognitive LOs, use Bloom's taxonomy: Knowledge, Comprehension, Application, Analysis, Synthesis, and Evaluation [114]. It is cost and time prohibitive to subject every trainee to every possible mission scenario. That is why it is important that trainees demonstrate the ability to apply what has been learned to new situations. Problem solving skills and decision-making capabilities are taught as part of higher order cognitive training [114]. Applying this taxonomy to a training program evaluation will show if the students are simply remembering and applying what is taught, or generalizing the information to create and devise solutions for scenarios that will arise after training. To classify affective LOs, use Reid's scientific attitude taxonomy: Directed Curiosity, Logical Methodology, Creative Ingenuity, Objectivity, and Integrity [131]. This affective taxonomy not only identifies the trainee's commitment and value based judgment of the LOs, but it also assesses their ability to communicate effectively. The affective classification level of objectivity measures the student's ability to assess error, control variables and view results objectively. Integrity is shown by the student's willingness to avoid bias, consider details that may appear contradictory, consider implications of their own work, cooperation and communication with others. Many accidents can be prevented if the managers and subordinates or peers know what to say and how to bring one another's attention to an errant situation in a non-offensive, yet stressing manner.

4.2.2 Step 2: Define Training Instructional Strategy Alternatives

In this step the reader will determine the availability, cost, and effectiveness of the instructional methods and media for the training program being evaluated. A mapping will be created to identify the compatibility between instructional methods and the learning objectives identified for the training program in step 1. The process for this step does not require a lesson-by-lesson mapping. For each of the learning taxonomies (cognitive, affective, psychomotor) used in step 1, the instructional methods that can help accomplish those objectives will be identified, step 2.a. The instructional methods will then be matched with corresponding method types (media and resources), step 2.b. Lastly, the effectiveness of each instructional method will be determined in step 2.c.

Map learning stages to corresponding instructional methods (2.A)

Develop a compatibility matrix to map each learning objective domain competency level to the instructional methods available in the training program. Part of the research presented in Chapter 2 discussed essential elements of instructional design that are needed to evaluate and predict training effectiveness. Two of the essential elements are learning objectives and instructional strategies. Each training lesson is comprised of learning objectives in the cognitive, affective, and/or psychomotor learning domain. Each training lesson is taught using an instructional method and media type. A physical instructor may also be required. Figure 4.3a depicts each of these elements as a part of the training lesson. Each one of the elements in Figure 4.3a contains subcategories. For example, cognitive learning objectives may be represented by Bloom's Taxonomy with three subcategories of 1) Knowledge and Comprehension, 2) Application and Analysis, and 3) Synthesis and Evaluation [114]. Training Media devices may contain the following four options 1) classroom, 2) computer, 3) simulator, and 4) aircraft. To define the possibilities of how each of the three levels of Bloom's Taxonomy are associated with all four media devices the compatibility between each pair must be determined. Rationalization and SME input may be required to determine if each pair is compatible, and the results can be shown in a compatibility matrix. A by-product of the compatibility matrix is that it reduces the number of alternatives for training lesson evaluation. With the number of LO competency levels and training methods the possible combinations of alternatives grows exponentially. Any reduction in alternatives saves time during the modeling and simulation process. Below are two examples of how to develop a compatibility matrix. The first example, shown in Figure 4.5, only matches two of the elements in Figure 4.3a and is brief in order to help the reader understand the compatibility matrix generation process. The second example shown in Table 4.3 is representative of the effort required to create a compatibility matrix. It shows how and where information can be collected to determine compatibility. If the training context is similar, the reader for evaluation of his or her own training program may use the compatibility matrix shown in Table 4.3. This compatibility matrix is developed using instructional methods recommended by the USAF.

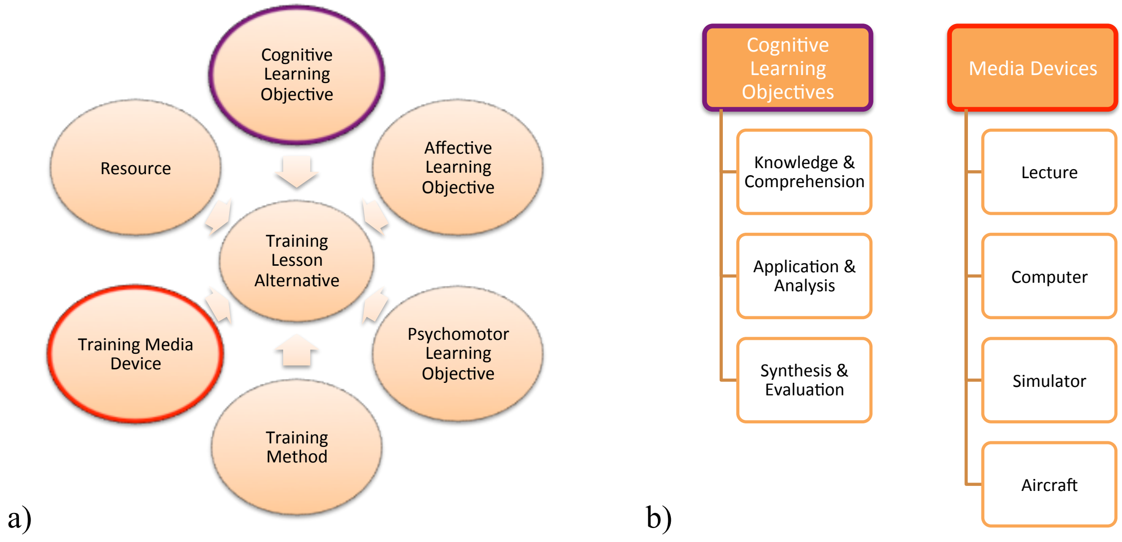


Figure 4.3: Training Lesson Alternative Criteria Elements

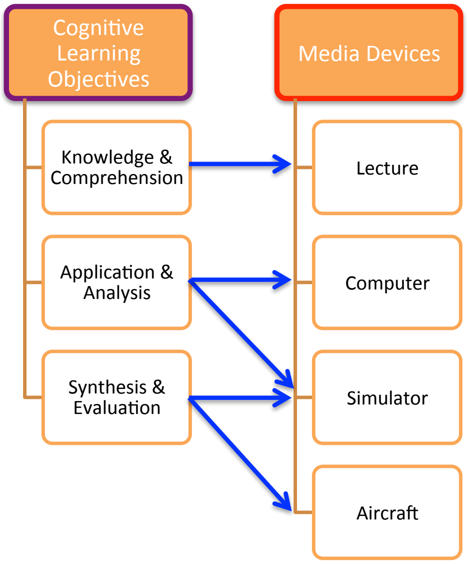


Figure 4.4: Criteria Elements Example Compatibility

		Cognitive Learning Objectives			Media Devices			
		Knowledge & Comprehension	Application & Analysis	Synthesis & Evaluation	Lecture	Computer-Based Training	Simulator	Aircraft
Cognitive Learning Objectives	Knowledge & Comprehension	▨	0	0	1	0	0	0
	Application & Analysis		▨	0	0	1	1	0
	Synthesis & Evaluation			▨	0	0	1	1
Media Devices	Lecture				▨	0	0	0
	Computer-Based Training					▨	0	0
	Simulator						▨	0
	Aircraft							▨

Figure 4.5: Example Compatibility Matrix

The first example creates the compatibility matrix shown in Figure 4.5. It shows example compatibility for two of the training lesson alternative criteria in Figure 4.3a: cognitive learning objective (circled in purple at the 12 o'clock position) and training media device (circled in red at the 8 o'clock position). Assume the subcategories for the cognitive learning objectives are represented by Bloom's Taxonomy as listed in Figure 4.3b. Assume the available training devices for this example are as also listed in Figure 4.3b: classroom, computer, simulator, and aircraft. Assume the only compatibilities are those represented by the blue arrows between the cognitive learning objectives and training media devices shown in Figure 4.4. Knowledge and Comprehension can only be taught via lecture. Application and Analysis is taught by computer-based training or in the simulator. Synthesis and Evaluation only pairs with the simulator or aircraft. Then the symmetric compatibility matrix, shown in Figure 4.5, will provide a mapping of these exact relationships. In the compatibility matrix a "1" implies that the combination of learning objective and training media is compatible and a "0" implies that the combination is not compatible. At the intersection of Knowledge and Comprehension versus lecture is a "1". A "1" is input where Application and Analysis intersects with computer-based training and simulator. A "1" is input where Synthesis and Evaluation intersects with simulator and aircraft. All other intersections, such as Knowledge and Comprehension and aircraft, contain a "0" because they are incompatible. The training media devices are not compatible with each other based on MPEET assumption number three. Although a training lesson and objective may require multiple instructional methods and devices to achieve, only one method and device is analyzed at a time. Within MPEET the training lesson must be broken into multiple parts based on the time spent using each instructional method and media device. Therefore, a "0" is placed at the intersections of each training media device as shown in Figure 4.5. This first example assumes the compatibility between each training lesson criteria element is known. The reader will likely have to research or solicit SME input to determine compatibility. The next example discusses how to determine initially unknown elemental compatibility for each of the criteria shown in Figure 4.3a. Other than this research step, the rest of the process for the second example is the same as this first example.

References for information on the compatibility between learning objectives and instructional strategies are contained within instructional design books and published articles and experiments as previously discussed in sections 2.8.4 and 2.8.5. Clark provides general guidelines for mapping different instructional strategies (methods and media) to LOs, as listed in Table 2.21 [147]. Findings from the Sitzmann et al. meta-analysis provide scientific justifications for mapping learning objectives to instructional strategies [153]. These findings have been used to generate the second example compatibility matrix shown in Table 4.3, and are recommend for use as part of the MPEET process. The sample compatibility matrix, Table 4.3, maps the recommended LOs (presented from previous sections 2.3 through 2.5) to the instructional methods recommended by the USAF that were previously listed in Table 2.22 [21]. This matrix formalizes which LOs and training methods are compatible. As in example 1, in the compatibility matrix a “1” implies that the combination of LO and training method is compatible and a “0” implies that the combination is not compatible. If the reader has not already, it is suggested that the reader examine Table 4.3. Each area will be discussed in detail, however a visual review before reading the details may aid in understanding.

In step 1, learning taxonomies were selected to describe the learning objectives and competency levels for each training lesson. The taxonomies recommend by the author are Bloom’s Taxonomy for the cognitive learning domain, Reid’s Scientific Attitude Taxonomy for the affective learning domain, and Ferris’ Taxonomy for the psychomotor learning domain [96, 114, 131]. Each taxonomy and its corresponding subcategory levels are included in Columns 2 – 16 in the sample compatibility matrix in Table 4.3. The rows of the compatibility matrix correspond to each column number to make the matrix symmetric just as in the first example. To make the table easier to read, the names of the subcategory levels have been abbreviated as shown in Figure 4.6. The recommended LO for cognitive learning is Bloom’s Taxonomy represented in three competency levels: 1) KC for knowledge and comprehension, 2) AA for Application and Analysis, and 3) SE for Synthesis and Evaluation [72]. Affective learning is represented in terms of Reid’s taxonomy consisting of five competency levels: 1) D for Directed Curiosity, 2) L for Logical Methodology, 3) C for Creative Ingenuity, 4) O for Objectivity, and 5) I for Integrity [131]. Psychomotor LOs are represented by Ferris’ seven level taxonomy: 1) R for Recognition, 2) H for Handling, 3) B for Basic Operation,

4) CO for Competent Operation, 5) EO for Expert Operation, 6) P for Planning, and 7) E for Evaluation [96].

Level	**1**	**2**	**3**	**4**	**5**	**6**	**7**
Cog - Bloom's Cognitive LO	**KC**: Knowledge & Comprehension	**AA**: Application & Analysis	**SE**: Synthesis & Evaluation				
Aff - Reid's Affective LO	**D**: Directed Curiosity	**L**: Logical Methodology	**C**: Creativity	**O**: Objectivity	**I**: Integrity		
Psy - Ferris' Psychomotor LO	**R**: Recognition	**H**: Handling	**B**: Basic Operation	**CO**: Competent Operation	**EO**: Expert Operation	**P**: Planning	**E**: Evaluation

Figure 4.6: Learning Taxonomy Abbreviations

Columns 17-24 of Table 4.3 list the USAF recommended training methods [21]. To determine which training methods (Columns 17-24) are compatible to each of the LO competency levels in Rows 2-16 the author referred back to the results from the 2006 Siztmann et al. meta-analysis [153]. For details of the Sitzmann et al 2006 meta-analysis refer back to section 2.8.4. Presented here are applicable conclusions from the study. A summary of the research purpose, analysis data set, application and conclusions are included in section 2.8.4. In terms of the performance or practice training method (Column 23 TM-7), web-based instruction (WBI) and classroom instruction (CI) were both more effective when training included practice for teaching declarative knowledge. WBI was more effective than CI when both delivery media incorporated practice, but WBI was less effective than CI when both failed to include practice during training. When teaching through WBI, practice should be incorporated into training in order to achieve the same effect as teaching in the classroom. Procedural knowledge obtained was equal for WBI and CI based on the applicable studies. With this information, the matrix is populated with compatibility between performance (TM-7) and all cognitive (Cog-1, Cog-2, Cog-3) and psychomotor (Psy-1 through Psy-7) LO competency levels.

The collaborative learning environment, which corresponds to the seminar training method (Column 21 TM-5), became more effective using WBI as the length of training increased. For shorter training programs CI was more effective; for training facilitated over a longer time, WBI was more effective. Using either method students are able to engage in peer-to-peer interactions. With this information the matrix is populated with compatibility between seminar (TM-5) and all cognitive (Cog-1, Cog-2, Cog-3) and psychomotor (Psy-1 – Psy-7) LO competency levels.

Providing feedback to trainees was beneficial during both WBI and CI. WBI was more effective than CI for providing feedback when teaching declarative knowledge. WBI and CI were equally effective for procedural knowledge. Feedback can and is given as part of multiple training methods. Based on the training method description listed in Table 2.22, the methods involving feedback are questioning, discussions, and case studies. With this information, the matrix is populated with the use of questioning (Column 20 TM-4), discussion (Column 22 TM-6), and case study (Column 24 TM-8) for all three cognitive competency levels (Cog-1, Cog-2, Cog-3) and the seven psychomotor competency levels (Psy-1 – Psy-7). Two statistically significant correlations between feedback and practice were found. Courses that incorporated practice also tended to provide feedback to trainees, and college students were more likely to receive the opportunity to practice during training compared to employees.

In terms of human instructor interaction, the meta-analysis found that human interaction did not affect learning from WBI relative to CI for declarative information.

There was evidence that synchronous communication facilitated learning more than asynchronous communication in WBI for declarative knowledge. Of the studies available to assess procedural knowledge, WBI and CI were equally effective in regards to human interaction. With this information the compatibility matrix is populated with the potential to use or not use an instructor (Columns 25-26, R-0 and R-1) for all cognitive LOs (Cog-1, Cog-2, Cog-3) and psychomotor LOs (Psy-1 – Psy-7).

The meta-analysis results presented by Sitzmann et al. do not provide any indicators to help determine the compatibility of the remaining training methods: lecture (Column 17 TM-1), demonstration (Column 18 TM-2) and exhibits (Column 19 TM-3). Fortunately, there are numerous ISD textbooks that directly map these training methods to all LOs, including the affective domain. *Designing Effective Instruction* by Morrison et al. was chosen to complete compatibility of the remaining training methods and the affective LOs competency levels because the recommendations made in this book are based on academic and professional scientific studies of education and training. The authors have more than 100 years of combined practice in instructional design, have published hundreds of journal articles in educational technology, instructional design, and human performance, authored and co-authored several textbooks, and have received scholarly awards and endowments. Additionally, their prescribed instructional methods for TM-4 through TM-8 for the cognitive and psychomotor LOs were compared to those found in the meta-analysis conducted by Sitzmann et al. as a verification that the two sources do not provide conflicting information. Table 4.1 shows the training methods mapped to the cognitive and psychomotor LO competency levels found in Morrison's et al. ISD reference book versus the results from Sitzmann's et al. meta-analysis. Each of the prescribed instructional methods by Morrison et al. that are included in the Sitmannn et al. meta-analysis results are equivalent. Based on the scholarly aptitude of Morrison et al. and the fact that their recommended training methods for TM-4 through TM-8 in the cognitive and psychomotor domain correspond to the meta-analysis results of Sitzmann et al., their prescribed training methods are used to complete the mappings between TM-1, TM-2, and TM-3 for the cognitive and psychomotor competency levels and all eight training methods related to affective LOs. Table 4.2 summarizes these mappings.

A combination of presentation (TM-1), demonstration (TM-2), and exhibits (TM-3) is used by Morrison et al. to teach cognitive ability. From teaching facts, concepts, principles, rules and complex procedures, they provide different strategies for using all three methods, and the most effective sequence order. The strategies included instructor led training and the use of video and multi-media devices. They prescribe methods for teaching interpersonal skills and attitudes separately. The authors describe attitudes as beliefs and associated behavior or responses. Interpersonal skills deal with the development of communication skills. Attitudes correspond to the first two levels of Reid's scientific attitude taxonomy: directed curiosity and logical methodology [131]. Interpersonal skills encompass the next three levels: creative ingenuity, objectivity, and integrity. All eight of the instructional methods (Column 17-24 TM-1 through TM-8) are given as possibilities to teach both interpersonal skills and attitudes.

After using the meta-analysis results from the Sitzmann et al. study and the prescribed training methods for each LO competency level defined by Morrison et al., the highest two psychomotor competency levels, planning (Psy-6) and evaluating work instructions (Psy-7) are not directly addressed by either reference. Previously discussed in the introduction of Ferris' psychomotor learning taxonomy, these top two competency levels overlap with higher order cognitive ability [96]. Ferris believed that mastery of physical skills was accomplished by not only becoming an expert in motor development, but also by being capable of critically assessing the effectiveness of tools and how they can be enhanced or the process be modified using the same object. These two psychomotor levels are very similar to synthesis and evaluation in the cognitive learning domain. The compatibility matrix is populated for psy-6 and psy-7 the same as cog-3. This completes the mapping process of the training methods to each of the stages of learning in terms of cognitive, affective, and psychomotor abilities.

Table 4.1: Comparison of TM to LOs Mappings from Sitzmann and Morrison

	Sitzmann et al.		**Morrison et al.**		
	LO Competency Levels	**Training Methods**	**LO Competency Levels**	**Training Methods**	**Page Reference[6]**
1	Cog-1 (KC)	TM-4 (Questioning)	Cog-1 (Remember)	TM-4 (Questioning)	Pg. 144
2	Cog-1 (KC)	TM-5 (Seminar)			
3	Cog-1 (KC)	TM-6 (Discussion)			
4	Cog-1 (KC)	TM-7 (Performance)	Cog-1 (Remember)	TM-7 (Practice)	Pg. 146
5	Cog-1 (KC)	TM-8 (Case Study)	Cog-1 (Remember)	TM-8 (Case Study)	Pg. 144
6	Cog-2 (AA)	TM-4 (Questioning)	Cog-2 (AA)	TM-4 (Questioning)	Pg. 145-147
7	Cog-2 (AA)	TM-5 (Seminar)	Cog-2 (AA)	TM-5 (Seminar)	Pg. 145-146
8	Cog-2 (AA)	TM-6 (Discussion)	Cog-2 (AA)	TM-6 (Discussion)	Pg. 145-147
9	Cog-2 (AA)	TM-7 (Performance)	Cog-2 (AA)	TM-7 (Practice)	Pg. 144-146
10	Cog-2 (AA)	TM-8 (Case Study)			
11	Cog-3 (SE)	TM-4 (Questioning)			
12	Cog-3 (SE)	TM-5 (Seminar)			
13	Cog-3 (SE)	TM-6 (Discussion)	Cog-3 (Evaluation)	TM-6 (Discussion)	Pg. 149
14	Cog-3 (SE)	TM-7 (Performance)	Cog-3 (Synthesis)	TM-7 (Practice)	Pg. 148
15	Cog-3 (SE)	TM-8 (Case Study)	Cog-3 (Synthesis)	TM-8 (Case Study)	Pg. 149
16	Psy-1 thru Psy-7	TM-4 (Questioning)			
17	Psy-1 thru Psy-7	TM-5 (Seminar)			
18	Psy-1 thru Psy-7	TM-6 (Discussion)	Psy-1 thru Psy-5	TM-6 (Discussion)	Pg. 150
19	Psy-1 thru Psy-7	TM-7 (Performance)	Psy-1 thru Psy-5	TM-7 (Practice)	Pg. 149
20	Psy-1 thru Psy-7	TM-8 (Case Study)			

---

[6] G.R. Morrison, S.M. Ross, H.K. Kalma, and J.E. Kemp, *Designing Effective Instruction*, 7th ed. Hoboken, NJ: John Wiley & Sons, Inc., 2013.

Table 4.2: Training Method Compatibility for TM-1 thru TM-3, and Affective LOs

LO Competency Level	**Training Method**	**Page Reference**[7]
Cog-1 (Remember)	TM-1 (Presentation)	Pg. 146
Cog-1 (Remember)	TM-2 (Demonstration)	Pg. 146-147
Cog-1 (Remember)	TM-3 (Exhibit)	Pg. 147
Cog-2 (Application and Analysis)	TM-1 (Presentation)	Pg. 145-146
Cog-2 (Application and Analysis)	TM-2 (Demonstration)	Pg. 144-147
Cog-2 (Application)	TM-3 (Exhibit)	Pg. 144, 147
Cog-3 (Synthesis)	TM-1 (Presentation)	Pg. 148
Cog-3 (Synthesis)	TM-2 (Demonstration)	Pg. 149
Cog-3 (Synthesis)	TM-3 (Exhibit)	Pg. 149
Psy-1 thru Psy-5	TM-1 (Presentation)	Pg. 149
Psy-1 thru Psy-5	TM-2 (Demonstration)	Pg. 149
Psy-1 thru Psy-5	TM-3 (Exhibit)	Pg. 149
Aff-1 thru Aff-5	TM-1 (Presentation)	Pg. 150
Aff-1 thru Aff-5	TM-2 (Demonstration)	Pg. 150
Aff-1 thru Aff-5	TM-3 (Exhibit)	Pg. 150
Aff-1 thru Aff-5	TM-4 (Questioning)	Pg. 150
Aff-1 thru Aff-5	TM-5 (Seminar)	Pg. 150
Aff-1 thru Aff-5	TM-6 (Discussion)	Pg. 150
Aff-1 thru Aff-5	TM-7 (Practice)	Pg. 150
Aff-1 thru Aff-5	TM-8 (Case Study)	Pg. 150

In addition to mapping the LOs (Rows 2-16) to training methods (Columns 17-24), LOs must also be mapped between and within each learning domain; mapping Rows 2-16 to Columns 2-16. The competency levels of Bloom's cognitive taxonomy are incompatible with each other. They each describe increasing ability to reason, process, and create new information. The competency levels of Reid's affective taxonomy are incompatible with each other. They describe increasing levels of commitment, value, and attitude control towards the information being taught. Each affective competency level is compatible with each level of cognition. Trainees' can reach any cognitive ability level and place very little value or appreciation towards the knowledge gained. The first five competency levels of Ferris' psychomotor taxonomy are incompatible with each other [96]. They describe different stages of physical skills from recognizing a tool or material, being able to handle it properly, conducting basic operations with the item, competently operating with the tool, and expertly handling and working with it. Levels six and seven, planning and evaluation, describe a trainees' ability to abstractly define the use and effectiveness of the tool during the work process. A trainee can be theoretically capable of planning and evaluating the work process with a basic, competent, or expert operational ability. This is common in production facilities where supervisors and managers can recognize and handle the tools, but because of skill decay, may no longer be certified to operate the machinery. In this case, planning (Psy-6) and evaluation (Psy-7) are incompatible with recognition (Psy-1) and handling (Psy-2), but are compatible with basic operation (Psy-3), competent operation (Psy-4), and expert operation (Psy-5). The first five skill levels of Ferris' psychomotor taxonomy are compatible with all cognitive and affective competency levels. The planning and evaluation of work operations (Psy-6 and Psy-7) requires higher cognitive abilities. To plan the work operations the trainee must know how to apply the use of a tool or material in proper sequence, therefore planning (Psy-6) is incompatible with the lowest level of cognition, KC (Cog-1), but is compatible with AA (Cog-2) and SE (Cog-3). To evaluate the work process (Psy-7) the trainee must be capable of creating improved work methods to increase effectiveness. In relationship to cognitive ability, this corresponds to only the highest cognitive competency level, SE (Cog-3). Therefore, psychomotor evaluation (Psy-7) is not compatible with the two lower cognitive levels, KC (Cog-1) and AA (Cog-2).

As stated in the third MPEET assumption, one training method is used at time. Therefore, the training methods are incompatible with each other- TM-1 is not compatible with TM-2, TM-3, TM-4, TM-5, TM-6, TM-7, and TM-8. The same philosophy is applied for the other training methods. Multiple training methods may be used to teach any LO, such as a discussion following practice or a video review. When analyzing the effect of each training alternative, each method is assessed individually with consideration given for improved effectiveness when training methods are utilized in proper sequence (this process is upcoming in step 4).

The development of the compatibility matrix requires an understanding of which training methods can be used to accomplish the learning objectives in each learning domain. As in the case of example two, industry standard references can be used to gain insight into best instructional practices for the methods available in the training program that the reader is evaluating. The compatibility matrix mapping, as shown in Table 4.3, was generated based on meta-analysis and instructional evaluation studies that tested the effectiveness of various instructional methods and resources. Assuming the training program under evaluation has the same instructional methods available and used the same taxonomies, then Table 4.3 can be used as presented. If other taxonomies or training methods are used, then create a compatibility matrix that maps the specific data available. The described development process of Table 4.3 can be used as a guide. If one is unfamiliar with the training system, solicit input from SMEs such as instructors and the instructional design team, or from student questionnaires and experimentation.

Table 4.3: Sample Compatibility Matrix for LOs, Training Methods, Resources

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26
		Cognitive			**Affective**					**Psychomotor**							**Training Method**								**Resource**	
		KC	**AA**	**SE**	**D**	**L**	**C**	**O**	**I**	**R**	**H**	**B**	**CO**	**EO**	**P**	**E**	**Lecture**	**Demo**	**Exhibit**	**Questioning**	**Seminar**	**Discussion**	**Performance**	**Case Study**	**No**	**Yes**
1	**Level**	**1**	**2**	**3**	**1**	**2**	**3**	**4**	**5**	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**TM-1**	**TM-2**	**TM-3**	**TM-4**	**TM-5**	**TM-6**	**TM-7**	**TM-8**	**0**	**1**
2	**Cog-1**		0	0	1	1	1	1	1	1	1	1	1	1	0	0	1	1	1	1	1	1	1	1	1	1
3	**Cog-2**			0	1	1	1	1	1	1	1	1	1	1	1	0	1	1	1	1	1	1	1	1	1	1
4	**Cog-3**				1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
5	**Aff-1**					0	0	0	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
6	**Aff-2**						0	0	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
7	**Aff-3**							0	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
8	**Aff-4**								0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
9	**Aff-5**									1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
10	**Psy-1**										0	0	0	0	0	0	1	1	1	1	1	1	1	1	1	1
11	**Psy-2**											0	0	0	0	0	1	1	1	1	1	1	1	1	1	1
12	**Psy-3**												0	0	1	1	1	1	1	1	1	1	1	1	1	1
13	**Psy-4**													0	1	1	1	1	1	1	1	1	1	1	1	1
14	**Psy-5**														1	1	1	1	1	1	1	1	1	1	1	1
15	**Psy-6**															1	1	1	1	1	1	1	1	1	1	1
16	**Psy-7**																1	1	1	1	1	1	1	1	1	1
17	**TM-1**																	0	0	0	0	0	0	1	1	1
18	**TM-2**																		0	0	0	0	0	0	1	1
19	**TM-3**																			0	0	0	0	0	1	1
20	**TM-4**																				0	0	0	0	1	1
21	**TM-5**																					0	0	0	1	1
22	**TM-6**																						0	0	1	1
23	**TM-7**																							0	1	1
24	**TM-8**																								1	1
25	**R-0**																									0
26	**R-1**																									

Level	**1**	**2**	**3**	**4**	**5**	**6**	**7**
Cog - Bloom's Cognitive LO	**KC**: Knowledge & Comprehension	**AA**: Application & Analysis	**SE**: Synthesis & Evaluation				
Aff - Reid's Affective LO	**D**: Directed Curiosity	**L**: Logical Methodology	**C**: Creativity	**O**: Objectivity	**I**: Integrity		
Psy - Ferris' Psychomotor LO	**R**: Recognition	**H**: Handling	**B**: Basic Operation	**CO**: Competent Operation	**EO**: Expert Operation	**P**: Planning	**E**: Evaluation

Match instructional methods with an appropriate method type(s) (2.B)

Expand the compatibility matrix produced in step 2.A to include the relationship between instructional methods, media type, and resources. The example from step 2.A is continued here to aid the reader. Instructional media are the mechanisms used for presenting material to trainees. Common media types used by the USAF listed in Table 2.23 included audio/visual devices, print materials, computers, simulators, and actual equipment trainers [20]. These media types can be classified by instructional methods given in a classroom (CR), by computer (CMP), part-task trainers (PTT), simulators

(SIM), and in the aircraft (AC) as listed in Columns 27-31 of Table 4.4. PTT represents mock-ups or varying fidelity level aircraft hardware that allows students to practice training exercises. Table 4.4 shows the appended sample compatibility matrix, which now includes the LO competency levels for all three learning domains, training methods, resources, and media types. Similar to the assumption used with instructional methods, it is assumed that each media type can be used to obtain any LO competency level. This does not consider differences in effectiveness, which is a part of upcoming step 2.C. The media types are not compatible with each training method.

A lecture (Row 17 TM-1), questioning (Row 20 TM-4), and discussion (Row 22 TM-6) are not suitable when using PTT (Column 29) or SIM (Column 30) devices. Therefore a "0" is placed at the intersection of lecture and PTT, questioning and PTT, and discussion and PTT representing incompatibility. Likewise, a "0" is placed at the SIM intersections with lecture, questioning, and discussion. The methods of lecture, questioning, and discussion are compatible with the CR (Column 27), CMP (Column 28), and AC (Column 31) environment. A "1" is place at the intersection of lecture (Row 17 TM-1) and CR (Column 27), lecture (Row 17 TM-1) and CMP (Column 28), and lecture (Row 17 TM-1) and AC (Column 31) showing their compatibility. A "1" also place where question (Row 20 TM-4) and discussion (Row 22 TM-6) intersect with CR, CMP, and AC.

Interactive web discussions and forums allow the use of CMP for TM-1, TM-4, and TM-6. This also enables training for distance learners. Using the AC for these three training methods is expensive, but possible. Cost of a method and media device is factored into the OEC and CU analysis in step 3. At this point in the process, compatibility is completely based on mapping possibilities. Seminar (TM-5) and case study (TM-8) are not compatible with PTT, SIM, or AC devices. The CR or interactive CMP environment is needed for these methods due to the amount of lecture and discussion between students and instructors. With these methods students are synthesizing information, evaluating, and/or planning, but not engaged in physical training. Performance or practice (TM-7) is not compatible with the CR environment for aircrew training. Performance here refers to the student interacting with equipment, which requires the use of a CMP, PTT, SIM, or AC. All of the media devices are compatible with demonstration (TM-2) and exhibit (TM-3). A live instructor, referred to as a resource, can be used for any training method or media type. Training content can be designed and implemented with or without an instructor, as is the case with many online or computer based courses.

Table 4.4: Sample Compatibility Matrix for Instructional Methods and Media Types

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31
		Cognitive			**Affective**					**Psychomotor**							**Training Method**								**Resource**		**Media Type**				
		KC	**AA**	**SE**	**D**	**L**	**C**	**O**	**I**	**R**	**H**	**B**	**CO**	**EO**	**P**	**E**	**LE**	**DE**	**EX**	**Q**	**S**	**DI**	**PE**	**CS**	**No**	**Yes**	**CR**	**CMP**	**PTT**	**SIM**	**AC**
1	**Level**	**1**	**2**	**3**	**1**	**2**	**3**	**4**	**5**	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**TM-1**	**TM-2**	**TM-3**	**TM-4**	**TM-5**	**TM-6**	**TM-7**	**TM-8**	**0**	**1**	**MT-1**	**MT-2**	**MT-3**	**MT-4**	**MT-5**
2	**Cog-1**		0	0	1	1	1	1	1	1	1	1	1	1	0	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
3	**Cog-2**			0	1	1	1	1	1	1	1	1	1	1	1	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
4	**Cog-3**				1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
5	**Aff-1**					0	0	0	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
6	**Aff-2**						0	0	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
7	**Aff-3**							0	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
8	**Aff-4**								0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
9	**Aff-5**									1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
10	**Psy-1**										0	0	0	0	0	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
11	**Psy-2**											0	0	0	0	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
12	**Psy-3**												0	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
13	**Psy-4**													0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
14	**Psy-5**														1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
15	**Psy-6**															1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
16	**Psy-7**																1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
17	**TM-1**																	0	0	0	0	0	0	1	1	1	1	1	0	0	1
18	**TM-2**																		0	0	0	0	0	0	1	1	1	1	1	1	1
19	**TM-3**																			0	0	0	0	0	1	1	1	1	1	1	1
20	**TM-4**																				0	0	0	0	1	1	1	1	0	0	1
21	**TM-5**																					0	0	0	1	1	1	1	0	0	0
22	**TM-6**																						0	0	1	1	1	1	0	0	1
23	**TM-7**																							0	1	1	0	1	1	1	1
24	**TM-8**																								1	1	1	1	0	0	0
25	**R-0**																									0	1	1	1	1	1
26	**R-1**																										1	1	1	1	1
27	**MT-1**																											0	0	0	0
28	**MT-2**																												0	0	0
29	**MT-3**																													0	0
30	**MT-4**																														0
31	**MT-5**																														

Level	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**
Cog - Bloom's Cognitive LO	**KC**: Knowledge & Comprehension	**AA**: Application & Analysis	**SE**: Synthesis & Evaluation					
Aff - Reid's Affective LO	**D**: Directed Curiosity	**L**: Logical Methodology	**C**: Creativity	**O**: Objectivity	**I**: Integrity			
Psy - Ferris' Psychomotor LO	**R**: Recognition	**H**: Handling	**B**: Basic Operation	**CO**: Competent Operation	**EO**: Expert Operation	**P**: Planning	**E**: Evaluation	
TM - USAF Training Methods	**LE**: Lecture	**DE**: Demonstration	**EX**: Exhibit	**Q**: Questioning	**S**: Seminar	**DI**: Discussion	**PE**: Performance	**CS**: Case Study
MT - USAF Media Types	**CR**: Classroom	**CMP**: Computer	**PTT**: Part-Task Trainer	**SIM**: Simulator	**AC**: Aircraft			

Determine the effectiveness (knowledge recall) of each instructional method (2.C)

As part of the instructional design process, the effectiveness of each instructional method is typically determined. If this was done and the information is available prior to the reader performing the training evaluation, then collect the information from the ISD team and move to step 3. If the information is not available the most accurate means of determining the effectiveness of the instructional methods is to perform an experiment with a small sample group of the trainee population. Subject each trainee to the instructional techniques available in the training program. Administer examinations twenty-four hours after completion of the training objective and record results. Use the mean and standard deviations from the sample population to create the expected range of knowledge recall in the modeling environment. Budget constraints may prohibit this type of experiment. The next option is to use data available from a similar training program. If representative data is not available, use The NTL Learning Pyramid and SME input to generate an expected value for effectiveness with lower and upper bounds [152]. Each instructional method could be modeled as a deterministic or probabilistic variable. As discussed within section 2.8.4, there are learner specific traits that cause training results to vary among students irrespective of the andragogical principles used. Knowing the training results will vary, the effectiveness of each training methods needs to be modeled as a distribution permitting probabilistic results. If an experiment is conducted on a sample population, the distribution can be determined from the data collection. If using the Learning Pyramid [152], shown in Figure 2.10, and SME input, use a triangular distribution for each method. The triangular distribution is defined by three values: the minimum value “a”, the maximum value “b”, and the most likely value “c”. This distribution is beneficial when the mean and standard deviation are unknown, but the minimum and maximum values can be estimated. The minimum and maximum values are definite lower and upper bounds in the triangular distribution. When triangular distributions are summed together the exact bound is retained, which avoids undesirable extreme values. Shown below is the equation for the triangle distribution probability density function.

$$f(x|a,b,c) = \begin{cases} 0 & ; \ x < a, x > b \\ \dfrac{2(x-a)}{(b-a)(c-a)} & ; \ a \leq x \leq c \\ \dfrac{2(b-x)}{(b-a)(b-c)} & ; \ c \leq x \leq b \end{cases}$$

4.2.3 Step 3: Generate Feasible Alternative Training System Design

The literature review contained in background Chapter 2 revealed ten attributes grouped into three categories of attributes that may effect training: Learning Objectives, Instructional Strategies, and Criticality Ratings, as shown in Figure 2.18. These attributes provide the basic measure by which each training alternative will be judged for effectiveness, and are the desired attributes for predicting training effectiveness. If all ten criteria are not available, solicit SME input for the missing data. If SMEs or similar historical data is unavailable use the accurate data that is available. Crafting false data or assigning a zero utility value because data is lacking is not beneficial in a predictive analysis. Ideally, the data needs to come from credible references that are familiar with the training system design being evaluated.

Chapter 3 discussed several modeling and simulation approaches and tools for analyzing and visualizing large sets of data. Steps 1 and 2 can easily be accomplished using Microsoft (MS) Excel or even creating tables in MS Word or Power Point. To complete Steps 3, 4, and 5 it is recommended that the user select a modeling environment capable of coding multiple equations to avoid repetitious calculations, performs random sampling, Monte Carlo analysis, probabilistic analysis (standard and variations of the t-test), 2D and 3D graphing, allow input of constraints and performs data searches. The author recommends using JMP® based on its included built-in function capabilities.

Convert each attribute to a common utility scale (3.A)

Define the functions: $U_1(x_1)$, $U_2(x_2)$, $U_3(x_3)$, through $U_{10}(x_{10})$ for all ten variables that are necessary to predict the effectiveness of a training program: learning objectives (LOs) in the cognitive, affective, and psychomotor domains, instructional methods, instructional media, use of an instructor as a resource, and difficulty, importance, frequency, and consequence of error (DIFE) ratings. As described in section 2.12.1, the simplest method for converting an attributes to a utility scale is the proportional scoring formula [188]:

$$U_i(x_i) = \left(\frac{x - Lowest}{Highest - Lowest}\right) * 100$$

If there is a desire to include stakeholder judgment, use the direct method. If risk and uncertainty need to be assessed for the training method selection (not for cost here), then use the variable probability method. Both the direct and variable probability methods are presented in section 2.12.1. When converting each attribute value into a utility value it is important to consider the relationship between each variable and the OEC. To maximize the OEC each variable needs to be converted such that the best value corresponds to the highest utility, and the worst value corresponds to the lowest utility. For example, the most active training method, TM-8, should have a utility value of 1.0 because the most effective training method is desired over any other alternative (when evaluating effectiveness only), per assumption #5. In the same manner, TM-1 has a utility value of 0 because it is the least effective training method. In terms of cost, the resources and media devices that have the highest expense will have the lowest utility. Cost variables will be negatively correlated to their utility values when maximizing the OEC.

Determine the importance weightings for each of the attributes (3.B)

Obtain the relative importance, w_i, of each of the ten criteria from the stakeholders and decision-makers. There are various techniques for accomplishing this, as discussed in 2.12.2. The direct method is recommended. Ask the decision-makers to allocate a total of 100 points among the attributes according to their relative importance. This is best accomplished in a group setting so a consensus can be reached on the final allocation. If it cannot be done in a group setting, collect the individual preferences and use statistical analysis to create an overall set of preferences. Check the mean, median, and standard deviation of the group's preferences. Look for weightings that are extreme outliers. When the group cannot be brought together to discuss outliers, several iterations of data collection may be required before the combined preference values are representative of all stakeholders.

Determine the utility of each training alternative using the multi-attribute utility function (3.C)

As discussed in section 2.12, the multi-attribute utility function provides a framework for determining the overall utility of an alternative and builds upon the components of a sound cost-effectiveness (CE) analysis [188]. The ability of each alternative to alter the attributes must be established using a cause-and-effect relationship between each alternative and the MoE. Use the compatibility matrix created in step 2.B to identify cause-and-effect relationships. This matrix does not provide the impact of each relationship, but it does establish a correlation between the training method alternatives and MoEs. Assume that each training lesson can be administered using methods that are compatible with the baseline LOs and DIFE ratings. Each training method uses a certain type of media that does or does not require an instructor and has an associated effectiveness value (knowledge recall percentage) as discovered in step 2.C. Calculate the overall utility for each training alternative method and store this information, it will be used in the next step 3.D. The multi-attribute utility function is calculated based upon all ten criteria. The multi-attribute utility function with weightings is:

$$Overall\ Utility\ (Criterion) = U(x_1, \dots, x_{10}) = \sum_{i=1}^{10} w_i U_i(x_i)$$

Determine the cost-utility ratio for each training alternative and select the most cost-effective training solution (3.D)

Determine the cost-utility ratio (CUR) for each of the possible training alternatives within the training program; this requires three steps. To do this first calculate the CUR for each training lesson in the original training program. Calculate the cost of each training lesson by multiplying the amount of time spent in training times the cost for the training asset used during the corresponding lesson.

$$Training\ Lesson\ Cost = (Time * Asset\ Cost) + (Time * Resource\ Cost)$$

If a training lesson uses multiple instructional methods or training media devices (assets), per MPEET assumption number three, that lesson needs to be broken into multiple lessons. Training lessons using more that one instructional method must be separated to properly select a training alternative for each training lesson. Part of the upcoming evaluation process in step 4, is to compare the effectiveness of the original versus alternative training solution programs and the knowledge recall of each instructional method is used in this comparison. Currently, MPEET considers one method at a time, and does not calculate effectiveness for a combined set of instructional methods. For each lesson or sub-lesson instructional method and media device used the overall utility, training lesson cost, and CUR must be calculated. The overall utility is comprised of the utility of the ten attributes multiplies times the decision-maker's importance weightings, step 3.C.

$$CU\ Ratio = \frac{Training\ Lesson\ Cost}{Overall\ Utility}$$

Once the CUR is calculated for each training lesson in the original training program store this value.

The second step to calculate the CUR for each of the possible training alternatives within the training program is to determine which alternative instructional methods are feasible. For each original training lesson compare the effectiveness of the instructional method (step 2.C) used to the effectiveness of all other training methods available in the training program. If the knowledge recall is less than 100% or the highest probabilistic range available in the training program, then there is an opportunity to use an alternative instructional method. Per MPEET assumption number four, create a constraint to eliminate any alternative instructional methods that have a knowledge recall less than the original training lesson method. MPEET philosophy is maximum training effectiveness; therefore anything less effective than the original is not a feasible alternative. To determine which remaining alternative instructional methods are feasible the compatibility matrix (step 2.B) can be used. For each alternative instructional method that has a higher effectiveness than the instructional method used in the original training program, check that it is compatible with the learning objectives (cognitive, affective, and psychomotor) required by the original training program lesson. If the alternative instructional method is not compatible with the learning objectives in all three domains, then it is not a feasible alternative and should be discarded. The alternative instructional methods that are compatible with the same learning objectives as the original training program are valid alternatives for consideration. The instructional method for each training lesson must be individually evaluated because each lesson will have varying learning objectives. For each training lesson, a set of alternative instructional methods now exists.

The third and final step to calculate the CUR for each of the possible training alternatives within the training program is to now calculate the utility and cost for each of the alternative instructional methods based on the information from each individual training lesson. Each training lesson has an associated overall utility based on the ten criteria, and this was calculated in step 3.C. The alternative instructional method cost can be calculated based on the time required for the original training lesson. Per MPEET assumption number two, the time spent using one training method equals the same amount of time of any other method type. Calculate the cost of each feasible training alternative by multiplying the amount of time spent in training times the cost for the training asset used during the corresponding lesson. If an instructor is required for the alternative instructional method, then add in the instructor hours multiplied by the time spent administering the lesson.

$$\mathit{Alternative\ Training\ Lesson\ Cost} = (\mathit{Time} * \mathit{Asset\ Cost}) + (\mathit{Time} * \mathit{Resource\ Cost})$$

Using the calculated utility (step 3.C) for each training alternative and the associated cost, the cost-utility ratio (CUR) for each feasible training alternative can be calculated. The CUR is the cost of the training lesson divided by the overall utility score:

$$\mathit{Alternative\ CU\ Ratio} = \frac{\mathit{Alternative\ Training\ Lesson\ Cost}}{\mathit{Overall\ Utility}}$$

Compare the CUR from the original training lessons to the corresponding alternative CURs of the feasible alternatives. Whichever is the smallest CUR should be selected as the most cost-effective solution. This process must be repeated for each training lesson or sub-lesson in the original training program. Once all training lessons have been evaluated a new training program is created that consists of the most cost-effective training solutions based upon the instructional methods available in the training system design, the decision-maker importance weightings for each attribute, and the ten criteria enveloped in the multi-attribute utility function.

4.2.4 Step 4: Evaluate Alternative Training System Design

To evaluate the alternative training system design created in step 3, information from both the original and alternative training programs is required. Predict the effectiveness of the original training system design by comparing the alternative training program, created from selecting the training methods with the smallest CUR values, to the original training program based upon the number of original training lessons that are statistically equivalent or different. The paired t-test is a statistical hypothesis test commonly used to determine if there is a statistical difference between two data sets. Specifically, the matched pairs t-test compares the expected value between two or more correlated attributes and assesses the differences. For each of the training lessons, the possible training method alternatives have the same attribute values for LO competency and DIFE levels. The training instructional method and associated costs are changing in the OEC and CUR calculations. Therefore, there will be correlation between the original and recommended alternative training program. Determine if the recommended alternative training system design is statistically different than the original using the matched pairs t-test. If the original and alternative training programs are statistically different, one can only conclude with confidence in the overall cost-effectiveness of the number of training lessons administered using the most-cost effective training methods. For example, if 80% of the original training program matches the alternative training program, and the matched pairs t-test reports that the original and alternative training programs are statistically different at a 95% confidence level, then the conclusion is that there is 95% confidence that 80% of the original training program is cost-effective. The cost-effectiveness of the remaining 20% of the original training program can be improved. This conclusion is based upon the importance weightings assigned to each attribute (LOs, cost, resource (instructor), method type, DIFE rating).

A method for comparing the cost-effectiveness for the original and alternative training programs in their entirety is to compare the expected value, standard deviation, and minimum and maximum values of the CUR. If the alternative training program has a smaller CUR mean and standard deviation over the same bounds, then it is generally more cost-effective and desirable than the original training system design. If the alternative training program has an equivalent or smaller CUR mean and wider variance, then a conclusion as to which training program is better cannot be made without decision-maker input. Any alternative with the same mean but wider variance is less desirable because of the greater uncertainty. In step 5, the decision-maker is presented with the predicted effectiveness of the original training program as well as the alternative. He or she may choose a training program that is less cost-effective, if it has a smaller variance depending on the amount of risk he or she is willing to accept. If the alternative training program has a higher CUR mean and a very narrow variance, decision-maker input is also required before a choice is made between the original and alternative solutions. The decision-maker may prefer the higher certainty that results from a narrow variance to the decrease in cost-effectiveness. Sharing these summary statistics for the original and alternative training programs with the decision-maker allows him and/or her to make a more informed decision compared to only testing the mean difference.

In addition to comparing the predicted effectiveness of the original and alternative training programs, the reader should also evaluate how well each training program follows the instructional design recommendations discovered in Chapter 2, and how sensitive the results are to changes in the ten effectiveness criteria and probabilistic variables. Steps 4.A and 4.B permit a comparison of how well distributed the learning objectives are throughout the training program, and if they are sequenced per ISD standard practices. Steps 4.C – 4.E will determine the sensitivity of the variables comprised in the CUR. This will provide indications as to how future changes may affect the feasible training alternatives. At the bottom of step 4 in the MPEET process summary depicted in Figure 4.2 there are two arrows. One proceeds to step 5 and the other shows the option of returning and repeating step 3. Based on the sensitivity evaluation results from steps 4.C – 4.E the reader may choose to make modifications to the OEC criteria and/or importance weightings used in step 3. If this action is taken, then repeat all of step 3 and 4 before proceeding to step 5.

Determine how training tasks are distributed between lower and higher level learning objectives (4.A)

When the fifteen taxonomies from the cognitive, affective, and psychomotor learning domains were reviewed in Chapter 2 one of the commonalities is that they all began with simple learning objectives and increased in difficulty. Each author stressed the importance of reaching the highest level of complexity or difficulty. There was no consensus found regarding how much time should be spent administering lessons at each learning objective competency level, but it is clear that a training program should consist of a distribution among all levels. The author recommends viewing this distribution based on two variables. The first variable is the total number of training tasks contained in the curriculum design. The second is the number of hours spent performing each training activity. Plots showing how many training tasks fall into each competency level will show if the training program addresses the full spectrum of LOs. If there are no tasks with higher competency level LOs, the training program may be deemed ineffective because the crew is only prepared for the specific mission scenarios rehearsed during training. Using Reid's scientific affective taxonomy, objectivity and integrity are a must because this is where the crew learns how to communicate effectively. Without these top two competency levels the crew can physically handle the aircraft and create a solution to any mission problem, but they may lack the skills to effectively communicate the plan of action.

Hypothesis 1: The LOs of an effective training system design are distributed across lower and higher competency levels.

Success Criteria 1: At least one percent of the training activity requires the highest level of competency for cognitive, affective, and psychomotor LOs.

If the tasks vary between all competency levels, the conclusion is that the training program maybe effective. Further analysis is needed. There is not enough information at this step to draw any conclusions regarding how evenly distributed the training program should be between the LO competency levels. If the tasks do not vary among all learning objective competency levels this is an issue that needs to be raised with the instructional design team. Notify the design team that not all the training requirements are being met, and request an update to the design of the training program. Recommend that the original training program be revised. Otherwise, a training program may be implemented without the identified training requirements being met. Scenarios discovered by the GAO in regards to failure to link training needs and requirements to the actual training program, see section 1.3, can lead to the cancellation of training programs. Worse results can be envisioned where the training program is implemented and the trainees are not properly instructed and lack critical complex problem solving skills or effective communication. These are the skills necessary to perform in complex military missions and without proper training can result in serious injury or even loss of lives.

Determine the proper sequencing of instructional methods to reach each competency level (4.B)

The instructional design background research of Chapter 2 discovered a need to use a variety of instructional methods that range from passive to active student participation. Having a student engage in hands-on exercises before explaining the tools being used, their purpose, safety considerations, etc., can be harmful to the students, and is not very effective. For example, if a training objective is for the pilot to perform a C-130J takeoff, there should be several training tasks leading up to the pilot's actual first flight in the C-130J. This may include familiarization with the aircraft and instrumentation, step-by-step procedural guidance of pre-flight checks and communication required before takeoff, practice in the simulator, and written and simulated practical exams. Each of these subtasks can be administered with various instructional methods. To accomplish the goal of familiarizing the pilot with the instrumentation, a classroom lecture using a textbook showing cockpit pictures, a computer based training module, a lecture around a low or high fidelity cockpit mock-up, or a live lecture on the aircraft are all possibilities of instructional method alternatives. The competency level expected from each of these subtasks would progress from knowledge and comprehension to application and analysis. In the psychomotor domain, the expectancy would move from recognition and handling to competent and expert operation. The instructional design team is responsible for developing and sequencing training lessons for student safety and effectiveness. MPEET can assess the overall training program by verifying that the design has a general trend of using passive then active instructional methods and that the learning objective competency levels increase over time.

Plot the learning objective competency levels for each of the training lessons in chronological order. If a positive trend of increasing competency levels over time is observed, then that indicates the training program does not start off at too high of a complexity level. If a negative trend is observed or no trend (flat line), then the instructional design team should be made aware that the training program may not follow recommended instructional design principles and should be re-evaluated. A negative trend is not recommended because the students may not be capable of understanding the advanced training concepts before being introduced to the basic information and having time to build up to complex learning objectives. A flat line, or no general trend, indicates that the training program is centered around one specific competency level and likely fails hypothesis 1.

Next plot the training instructional methods for each individual training lesson in chronological order. Observe where the use of passive and active instructional methods occurs. Examples of passive instructional methods from the learning pyramid are auditory and visual lectures as well as watching demonstrations [148]. Examples of active instructional methods are class and group discussions where students are engaged in the discussion, practice, and eventually peer-to-peer teaching. In effective training system design passive instructional methods are used during the introductory phase of training to allow students to become familiar with the concepts and tools being used. Student engagement increases throughout a normal training program. In this plot verify that a positive trend, moving from passive to active instructional methods, is observed when viewing the training program in chronological order. If a negative trend is observed or no trend (flat line), then the instructional design team should be made aware that the training program may not follow recommended instructional design principles and should be re-evaluated. A negative trend may result in increased risk and safety issues if the trainees are being asked to perform hands-on activities before an understanding of the operations is obtained. A flat line represents the use of the same instructional method throughout the entire training program. Depending on if this line is in the passive or active range will determine the effectiveness of the training program. If the flat line represents the excessive use of passive instructional methods then the trainee population will not gain much more than factual knowledge, education about an operation, and likely will not be able to actually perform the task because no active participation or practice occurs during training. If the flat line represents excessive use of active instructional methods then precautions are necessary for safety, and consideration should be given that the trainees may learn how to perform but lack the understanding of why they perform certain actions. This can prohibit trainee's ability to relate what is learned in training to on-the-job or real-life scenarios. If there is no trend or a negative trend, the information should be given to the instructional design team for explanation as to why the instructional design best practices were not used and possible modification.

Identified in the introduction of instructional strategies in section 2.8.1 is a direct correlation between learning objectives and instructional methods. Lower levels of performance are usually taught using passive training methods. Higher levels of competency require some sort of action or involvement by students. These literary findings can be tested to determine the effectiveness of the training program.

Hypothesis 2: If a training program is effective, then passive training methods are used to teach non-complex learning objectives and complex or high levels of performance activities are taught using active instructional methods.

Success Criteria 2: The lower competency levels correspond to instructional strategies of 49% or less knowledge recall. The high competency levels map to instructional strategies of 50% or greater knowledge recall.

Plot learning objectives in the cognitive, affective, and psychomotor domains versus the instructional methods for each training lesson. Observe how the learning objectives are paired with the instructional methods for the training program. Identify any training lesson that uses a passive instructional method when an active method should have been use or vice versa. For any training lesson that fails hypothesis 2, identify an alternative instructional method to use and provide these findings to the instructional design team for updates to the training program. The decision-maker should also be made aware of any training lessons that fail this hypothesis. This information should be considered when deciding between the original and recommended alternative training programs.

Determine the sensitivity of the OEC criteria (4.C)

The resultant most cost-effective training alternatives found in step 3 are based upon the attribute importance weightings given by the stakeholders, the accuracy of the resource and asset cost estimates, and the multi-attribute utility function created by the author. The multi-attribute utility function is comprised of variables discovered in the literature search contained in Chapter 2 that attempted to answer RQ2.1, RQ2.2, and RQ3. Chapter 2 discusses the authors findings in trying to identify appropriate measures for quantifying the benefits of soldiers' training in term of effectiveness (RQ2.1), quantifying increases in knowledge, skills, and attitudes in training system design (RQ3), and determining how to allocate resources to maximize training effectiveness (RQ2.2). Ten attributes were discovered as necessary to predict the effectiveness of a training program: learning objectives in the cognitive, affective, and psychomotor domains, instructional methods, instructional media, use of an instructor as a resource, and difficulty, importance, frequency, and consequence of error ratings. All ten have been included in the overall utility function and CUR calculations. An implied assumption is that each of the ten attributes included in the OEC have significance at the importance weighting value assigned by the decision-makers. If this assumption is valid, then a sensitivity analysis will objectively show the importance of each criteria at the assigned weighting values and across a range of values.

Hypothesis 3: If the ten attributes identified as necessary to predict training effectiveness are statistically significant irrespective of the decision-maker importance weightings, then the multi-attribute utility function created in step 3.C is valid for predicting the cost-effectiveness of a training system design.

Success Criteria: Changes to the assigned importance weighting values for each individual criteria result in statistically significant changes to the training system design.

To test this hypothesis, a set of importance weightings must be generated as part of the sensitivity analysis. Use the assigned values of w_1 through w_{10}, determined in step 3.B, as the baseline set. Based on the increments of the baseline set, develop a range of values for assessing each attribute. A recommended range that covers 0 through 100% importance is: [0, 10, 25, 50, 75, 100]. Create a table of valid combinations for each of the weightings values and criteria. There are ten criteria in the OEC. Using the six recommended weighting values results in over 60 million (6^10 = 60,466,176) combinations of importance weightings that could be used in the analysis. However, these values represent relative importance and their sum must always equal 100. For these recommended ranges, the number of valid cases is reduced to 2,425. Table 4.5 provides a sample listing of valid and invalid combinations of importance weighting values. Once the table of valid combinations of importance weighting values is generated, recalculate the utility value for each training alternative by repeating steps 3.C and 3.D. Each combination will change the overall utility value calculated for the training alternatives in step 3.C. The CUR will increase or decrease in comparison to the results from the new training program found in step 3.D.

Table 4.5: Example of Importance Weighting Values for OEC Sensitivity Analysis

w_1	w_2	w_3	w_4	w_5	w_6	w_7	w_8	w_9	w_{10}	**Sum**	**Valid/Invalid**
0	0	0	0	0	0	0	0	0	0	0	Invalid
0.1	0.1	0.1	0.1	0.5	0.1	0	0	0	0	1.0	Valid
0	0.1	0.1	0.25	0	0.1	0.1	0.1	0.25	0	1.0	Valid
0.1	0.25	0.1	0	0	0.5	0.75	1.0	0	0	2.7	Invalid
1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	10.0	Invalid

After completing steps 3.C and 3.D for each of the valid combinations of importance weighting values, a possible new training program will have been created for each combination. Using the recommend range values, this results in 2,425 training programs. If there are training lessons that have the same values for each of the ten criteria, then it is possible to reduce the number of cases that need to be repeated in steps 3.C and 3.D. The purpose of this sensitivity study is not to create multiple training programs, but to determine the impact importance weightings have on the attributes. This will provide insight into the necessity of each attribute in predicting training effectiveness.

In section 2.9 multiple methods were presented for performing DIFE analyses. There is not an industry or military standard for DIFE analyses, but similarities existed among the techniques. A semblance is that each method evaluated difficulty, importance, frequency, and consequence of error independently. These variables are tested for correlation after ratings are assigned. Although DIFE is grouped under the category of criticality ratings, it is assumed that independently assessing DIFE variables has a significant impact on training effectiveness. The same assumption applies to the other two attribute groupings of learning objectives and instructional strategies. The difference between the DIFE values under criticality ratings and the learning objectives and instructional strategies is that each of the methods for performing DIFE analysis in the *DoD Handbook Instructional Systems Development/Systems Approach to Training and Education (Part 2 of 5 Parts) and The Theory and Practice of Training* by Buckley consisted of hierarchy lists [103, 182]. The methods suggest that the analyst specifically evaluate one variable in relationship to another, which can lead to inherent correlation. The reader can check for dependency in the DIFE ratings by determining the statistical significance of these attributes. This process is broken into four steps. 1) First, conduct a correlation test between the values assigned for DIFE. If all four variables are not correlated, have a correlation (r) of less than 0.5, skip the rest of step 1, step 2, and step 3. Go directly to step 4. For any attributes that are correlated, have a correlation (r) of greater than 0.5, create a set of independent values for this particular variable. The most accurate means to create this new set of ratings is with the assistance of an SME. Only give the SME the training lesson title and/or description and request that they rate only that attribute. Do not show them the other ratings. For example, if the importance weighting is correlated with the consequence of error or difficulty or frequency, only show the SME the lesson title or description and ask them to assess the importance. If SME input is not available, the reader can randomly assign values for the correlated attribute. After the attributes have been independently assessed by an SME, or randomly generated, verify that the DIFE ratings are now uncorrelated, $r < 0.5$. This process may need to be repeated until each variable in uncorrelated. 2) Second, using the baseline importance weightings for all ten attributes repeat step 3.C and 3.D with the independent values for DIFE. 3) Third, use the matched pairs t-test and test the statistical difference between the alternative training programs generated using the original values for DIFE versus this newly created alternative training program. If the two are statistically different, then this information should be documented and reported to the instructional design team and the decision-maker. The design team may choose to re-evaluate the original ratings or proceed with the original values. At this step it is not a matter of which is right or wrong as long as the original design did follow a standard process for DIFE analysis. If the two are statistically the same continue to the fourth and final step. 4) The results from the sensitivity study conducted at the beginning of step 4.C identified ranges of weight factors where DIFE attributes are significant. Choose settings for the importance weightings where the DIFE variables are significant. Repeat steps 3.C and 3.D and create an alternative training program that will be driven in design by the DIFE attributes. Compare the alternative training program generated using the original values for DIFE at the baseline importance weight factors to the newly created alternative training program.

If the two are statistically different then assessing DIFE independently has a significant impact on predicting training effectiveness. Independent versus dependent DIFE ratings result in a difference between the recommended alternative training solutions. If the reader had to perform steps 1-4 within this paragraph, then revisit step 3. The information reported to the instructional design team and decision-maker is now of greater essentiality. Be clear in reporting to the decision-maker that the predicted effectiveness is limited by the use of dependent DIFE ratings. If SME input was used for the generation of the uncorrelated DIFE attributes then suggest replacing the original with the new values after following a proper verification process. Otherwise, recommend that the original training program be re-assessed with independent values for DIFE.

If the two are not statistically different then regardless of the methods used in the original design of the training program for DIFE analysis and any inherent correlation, it has no impact on predicting the effectiveness of the training program. With this result the importance weighting factors assigned to the DIFE attributes are meaningless. The author recommends removing these attributes from the multi-attribute utility function and dispersing the importance weight factors among the remaining six attributes. Then repeat steps 3.B – 4.C to increase the accuracy of the predicted cost-effectiveness of the training program.

The DIFE analysis techniques discussed in the literature review of Chapter 2 varied in terms of the complexity and level of details used. The advantage of a more comprehensive DIFE analysis in predicting training effectiveness is the ability to see how many lessons fall into each level of training. Plot the number of training hours and training methods versus DIFE ratings. Using these plots report to the decision-maker and design team how much training time and cost is being spent on average for low and high priority tasks.

Determine the sensitivity of the training method knowledge recall distribution ranges (4.D)

It is unlikely that funding will be provided to conduct an experiment to determine the knowledge recall distribution of the instructional methods in step 2.C. If the learning pyramid and/or SME input was used to derive the effectiveness of each training method, then a sensitivity analysis should be performed to test the robustness of the training program against the variance in knowledge recalled by each student for a given training method. Without performing an experiment on a sample population to accurately describe the type of distribution and key parameters, an assumption is inherent regarding how to best represent the variance of this probabilistic variable. If the reader followed the recommendation of the author in step 2.C the distribution for each training method is assumed triangular within the bounds of lower and upper values defined with a mode (most likely outcome). It a good design and evaluation practice to assess the robustness of any uncertainty variable. To complete this sensitivity study a new mode and limits need to be identified for each training method. The new upper and lower bounds can theoretically range from 0 to 100% effectiveness. However, the new bounds and mode should be realistic. Repeat steps 3.C and 3.D using the baseline importance weighting values and modified knowledge recall. Compare the alternative training program derived using the modified limits to the results from the baseline in step 3.D to draw conclusions whether the recommended training program is sensitive to changes in knowledge recall. The author recommends incrementally decreasing and increasing the bounds by a tenth of the percentage bounds until the point is reached where a statistically significant difference occurs. This process may require several iterations. This step identifies how much the knowledge recall can actually vary from the sample or assumed mean for each instructional method without having a significant impact in predicting the cost-effectiveness of the training program. This information should be reported to the decision-maker as a limitation on the applicability of the results.

4.2.5 Step 5: Decision Support

The evaluation results from step 4 should be presented first to the instructional design team and then to the decision-maker. The instructional design team may choose to make modifications to the original training program and have it re-evaluated before it is presented to the decision-maker. Present the percentage of the training program that uses lowest cost-utility instructional methods. Present the alternative training program that uses only the most cost-effective training alternatives. Present the results from the matched pairs t-test. Explain through the results of the sensitivity analysis the importance of each criteria and how the importance weightings effect the predicted cost-effectiveness of the original and recommended alternative training programs. Iterate the boundary limitations of the predicted results and their applicability. Reiterate that MPEET uses only a fraction of the information that is generated from any comprehensive training system design analysis. As discussed in the assumptions, there are numerous learning variables MPEET does include. These variables have an impact on training effectiveness and are included as part of the instructional design tasks/needs and learner analyses. With all the information presented and limitations explained the decision-maker has the option to use: the baseline training program, generated alternative training program, or to incorporate changes to the baseline from the alternative.

4.3 MPEET Summary

MPEET, as summarized in Figure 4.2, allows a designer or analyst the ability to map together learning objectives with feasible instructional strategies to create a portfolio of instructional strategy alternatives. Decision-maker preferences are requested for each

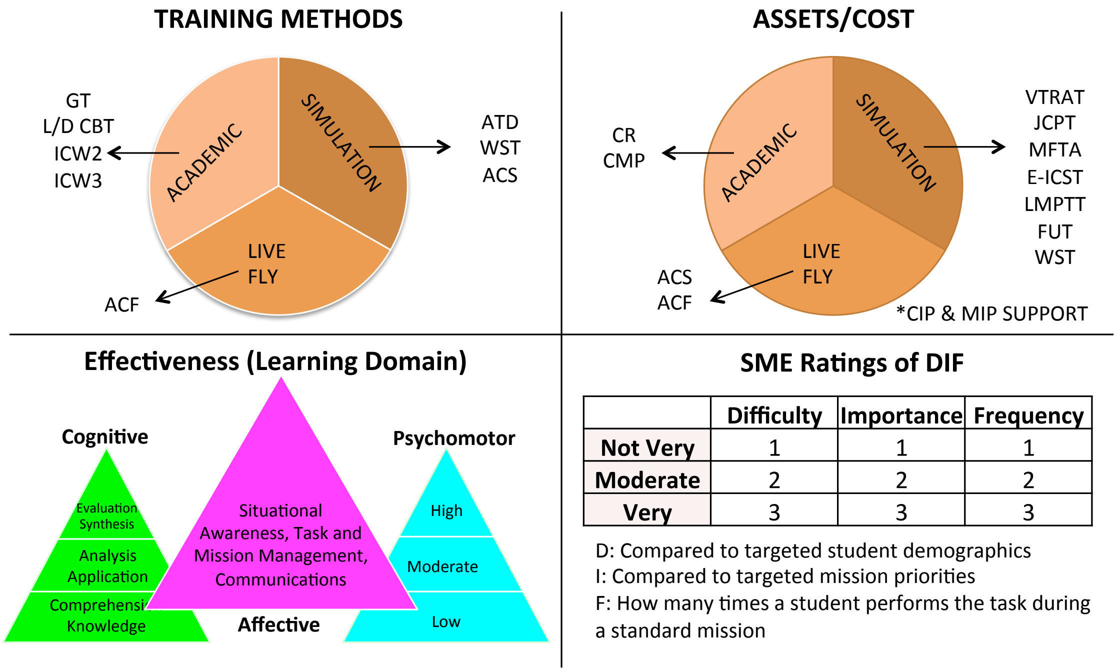
OEC criteria and utility values for each training objective is calculated. Then MPEET checks that the most effective methods are used for each learning objective. If there are alternatives with a lower CUR, compared to the baseline, a feasible alternative is generated. After each objective is evaluated MPEET creates a new alternative training program compiled of training lessons with the lowest value CUR based on the importance weightings assigned to the OEC criteria. MPEET verifies that instructional design best practices are utilized in the sequencing of tasks within the training program, and reports any findings. The sensitivity of attributes and probabilistic variables used in predicting the cost-effectiveness of the training program under evaluation and recommended alternative are identified and reported. Finally, the decision-maker is presented with the probabilistic effectiveness of the original training program as wells as the alternative. He or she may decide to select the original training program, incorporate specific changes from the alternative program, or implement the recommended alternative. MPEET provides a framework for evaluating the cost and effectiveness of training and meets the five necessary criteria for a training effectiveness model: 1) connecting training results to goals, 2) is based primarily on objective data (can be supported by subjective data), 3) accounts for variation of skill competency levels, 4) includes uncertainty analysis, and most importantly, 5) can be used to predict, rather than simply evaluate, performance results after training is complete.

CHAPTER 5

CASE STUDY of C-130J PILOT QTP USING MPEET

To demonstrate the feasibility of MPEET, a C-130J pilot qualification training program (QTP) will be used. The C-130J is relied upon by the USAF for clandestine or low visibility air refueling of helicopters and tilt-rotor aircraft, resupply of special operation forces by airdrop or airland in politically sensitive or hostile territories, search and rescue missions, humanitarian assistance operations, disaster response, and airdrops of leaflets [210, 211]. These missions require intense training to prepare the aircrew for planned and unexpected situations that require demonstrated knowledge and skills in multiple learning domains. Ensuring pilots receive the most effective training is critical to military and civilian safety both in the US and abroad. With an understanding of the defense budget constraints relative to the US economy, the affordability of C-130J pilot training must consist of a balance between effectiveness and cost. Determining the cost-effectiveness of the C-130J QTP is a prime candidate for an MPEET case study. An experienced instructional design team has designed the C-130J training program. The team included cognitive scientists, psychologists, certified adult education instructors, previous USAF C-130 flight instructors and evaluators, human performance engineers, human factors engineers, and training architects. The instructional design team presented the QTP in three training phases called initial qualification, tactical qualification, and special mission qualification. Completion of all three training phases is required; therefore, the training program is being evaluated based on the combined phases. Every training lesson in the QTP is administered using a specific training method and asset, is assigned a learning objective competency level in the cognitive, affective and psychomotor learning domains, and is rated for it's difficulty, importance and frequency. The information provided by the instructional design team is depicted below.

C-130J Aircrew Qualification Optimized Training Data Summary



	Difficulty	Importance	Frequency
Not Very	1	1	1
Moderate	2	2	2
Very	3	3	3

D: Compared to targeted student demographics
I: Compared to targeted mission priorities
F: How many times a student performs the task during a standard mission

5.1 Case Study Assumptions

There are assumptions that have been made prior to the proposed methodology being applied to the C-130J QTP. The effectiveness and cost-efficiency of any training program will partly depend on how well the learning objectives (LOs) are prioritized, clustered, and sequenced [103]. With the budget constraints in today's military environment, reduction in training time is necessary. This means that not every single task can be covered during a specific training course. LOs may need to be prioritized in order to provide the training that is most needed by the users. It is assumed that the training data provided has already been prioritized and determined that the lessons in this QTP are all required. There is not an analysis or experimentation being run to vary the length of the QTP. LOs are clustered and sequenced to present logical and meaningful portions of instruction throughout any training program [103]. Clustering can be done based on common prerequisites needed for other LOs, combining LOs that relate to the same system or require similar actions, teaching tasks that require common knowledge and skills jointly, or clustering LOs by the type of instructional strategy or method being used. This case study assumes that the clustering of LOs was designed for maximum effectiveness and optimal efficiency. Sequencing of LOs is extremely important to promote effective learning and to minimize risks, especially in terms of safety [103]. For example, a lesson that requires the use of dangerous complex equipment at night should not immediately follow multiple six hour per day exercises without allowing personnel the opportunity to sleep and adjust from day to night-time training. There are multiple sequencing orders to choose from in training design: job performance, chronological, cause and effect, criticality, simple-to-complex, and known-to-unknown order [103]. It is assumed that the lessons in the C-130J QTP data have been optimized for all sequencing orders just listed. If the reader is interested in learning more about prioritizing, clustering, and sequencing LOs, refer to the *DoD Handbook Instructional Systems Development/Systems Approach to Training and Education (Part 2 of 5 Parts)* [103]. In terms of sequencing, MPEET evaluates the use of instructional methods to verify passive and active techniques are used appropriately.

Training resources are the supplies and support necessary to design, develop, implement, support, operate, and maintain the instructional system. These resources are categorized into five major areas: equipment, facilities, funding, human resources, and time. It is assumed that the only types of equipment available for instructional methods are those currently used in the QTP. The facilities and time are set constraints. The training instructor is the only human resource that will be included in the trade space. Some instructional methods may require a teacher such as lecture and discussions and others will not, such as computer-based training (CBT). The equipment available is limited to that which is already being used in the QTP. The hourly costs of each instructional method type accounts for the development of a particular form of equipment.

5.2 MPEET Implementation

5.2.1 Training Requirements Defined

Use learning taxonomies to classify the stages of learning and competency levels (1.A)

The C-130J QTP uses cognitive, affective, and psychomotor classification systems to identify the expected LO and student capability for each training task and is summarized in Table 5.1. The cognitive classification system consists of four levels using a modified Bloom's taxonomy [114]. A training task can have no cognitive LO, a simplistic knowledge and comprehension (KC) LO, a moderate level of thinking capability for application and analysis (AA), or a higher order process involving synthesis and evaluation (SE). The affective classification system describes phases of situational awareness, task and mission management, and communications. Training tasks are designed to include the entirety of the uniquely developed categories or nothing at all involving affectivity. In this case, affectivity has two levels to evaluate, yes or no. The psychomotor classification system consists of four levels. The LO for each task is designated as no motor skill ability required, low, medium, or high psychomotor ability. This simplistic categorization system is strictly assessing the physical skills required for each training task. Within this training program all three taxonomies chosen to describe the cognitive, affective, and psychomotor competency levels are necessary because they are all specific to a learning domain.

Table 5.1: C-130J QTP Learning Objective Competency Levels

Level **Taxonomy**	**0**	**1**	**2**	**3**
Cognitive	None	KC	AA	SE
Affective	No	Yes		
Psychomotor	None	Low	Med	High

4 Cognitive x 2 Affective x 4 Psychomotor = 32 Combinations

Total Combinations of Training Task LO Classification System

5.2.2 Training Instructional Strategy Alternatives Defined

Map learning stages to corresponding instructional methods (2.A)

The C-130J training system design uses cognitive, affective, and psychomotor LOs. The psychomotor and affective taxonomies used by the instructional design team, shown in Table 5.1, are different than those recommended by the author. Therefore, the compatibility matrix developed in Table 4.3 cannot be applied directly in this evaluation. SME input was solicited and provided to create the following compatibility matrix.

Table 5.2: Compatibility Matrix: Competency Levels and Training Methods – Pilot QTP

	Cognition			Affective		Psychomotor				Training Method								
	KC	AA	SE	No	Yes	No	Low	Mod	High	GT	LD	CBT	ICW2	ICW3	ATD	ACS	SIM	ACF
Level	1	2	3	0	1	0	1	2	3	1	2	3	4	5	6	7	8	9
Cog-1		0	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Cog-2			0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Cog-3				1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Aff-0					0	1	1	1	1	1	1	1	1	1	1	1	1	1
Aff-1						1	1	1	1	1	1	1	1	1	1	1	1	1
Psy-0							0	0	0	1	1	1	1	1	1	1	1	1
Psy-1								0	0	1	1	1	1	1	1	1	1	1
Psy-2									0	1	1	1	1	1	1	1	1	1
Psy-3										1	1	1	1	1	1	1	1	1
TM-1											0	0	0	0	0	0	0	0
TM-2												0	0	0	0	0	0	0
TM-3													0	0	0	0	0	0
TM-4														0	0	0	0	0
TM-5															0	0	0	0
TM-6																0	0	0
TM-7																	0	0
TM-8																		0
TM-9																		

The compatibility matrix between the LO competency levels and training methods, Table 5.2, was completed with SME input from the instructional design team. There are 32 possible combinations of psychomotor, cognitive, and affective competency levels. Of the 32, only 23 (32 – 1 – 8 = 23) are compatible. The combination of no psychomotor, no cognitive, and no affective learning, from Table 5.1, was immediately deemed incompatible because no training lesson should be implemented if no learning is taking place. The training program was evaluated and it was verified that no lessons fall into this combination. As part of discussions with the design team, it was determined that every training lesson must involve some level of cognitive ability. This means that any combination that has zero cognition is unrealistic, resulting in eight incompatibilities. Because each training lesson contains some level of cognitive ability, the compatibility matrix excluded Level 0 for cognition. Only one competency level within a domain can be achieved during an individual training lesson. For example, a training lesson would not have a medium and high psychomotor competency level. The level would be medium or high. Each training methods is assumed capable of imparting all levels and types of learning. They are not equally as effective and that will be addressed in step 2.C where the effectiveness of each instructional method is determined. Per MPEET assumption #3 in section 4.1, only one training method can be assigned to each training lesson and the methods are incompatible with each other. Lessons that require multiple methods are separated into individual lessons based on the training methods used and time required for each subtask.

Match instructional methods with an appropriate method type(s) (2.B)

Eight media devices for the C-130J pilot training program are available for training. They are shown in Table 5.3 as training assets (TA). Each is identified by their hourly cost. These hourly costs are not exact for any specific C-130J training program. They are estimates listed in US Air Force Instruction (AFI) 65-503, *US Air Force Cost and Planning Factors* [212]. AFI 65-503 is the instruction that contains USAF cost and planning factors that AF activities use to estimate resource requirements and cost associated with AF force structures, missions, and activities. These costs are used herein for analysis purposes only. Similar to the assumption with instructional methods, it is assumed that each media type can be used to obtain any LO competency level. This does not consider differences in effectiveness, which is a part of step 2.C.

The media types or TAs are not compatible with each training method. Ground Training (GT), TM-1, can be given using no assets (TA-0) or in the classroom (TA-1). TM-1 is incompatible with all other devices (TA-2 through TA-8). Lectures and discussions (LD), TM-2, are given in the classroom (TA-1). Computer based lessons (TM-3) are given using in-class computers (TA-1) or individual laptops provided to students during training (TA-2). Instruction administered via interactive courseware (ICW) requires specific software designs; therefore, ICW-2 (TM-4) and ICW-3 (TM-5) are only compatible with TA-3 and TA-4, respectively. Aircraft training devices (ATDs) are low fidelity part task trainers (PTTs) that allow students to practice their skills on aircraft equipment. ATDs can consists of multiple devices. For analysis purposes, they are grouped into one category and the median hourly cost is used. ATD (TM-6) is only compatible with TA-5. There are two training methods that involve the aircraft TM-7 and TM-9. Training involving the aircraft can take place while the plane is grounded or flying. The difference between these two methods is that training on a standing aircraft (ACS), TM-7, cost less per hour than when the aircraft is flying (ACF), TM-9, because no fuel is consumed during aircraft training. TM-7 is compatible with TA-6 and TM-9 is compatible with TA-8. These are the representative costs for using the aircraft during training. Simulated (SIM) training, TM-8, requires specially designed software and a virtual environment for training. TM-8 is only compatible with TA-7.

All the training methods can involve an instructor or resource. Certain training methods and devices must have a resource to ensure safety of students and hardware. Instructors facilitate lectures and discussions. TM-2 will always utilize a resource; therefore, when instruction takes place in a classroom, TA-1, a resource is also required.

During training, pilots must be supervised by an instructor when practicing on the aircraft and engaging in simulated exercises. TM-7, TM-8, and TM-9 require a resource along with their corresponding training devices, TA-6, TA-7, and TA-8.

Table 5.3: Compatibility Matrix: Instructional Methods and Media Types – Pilot QTP

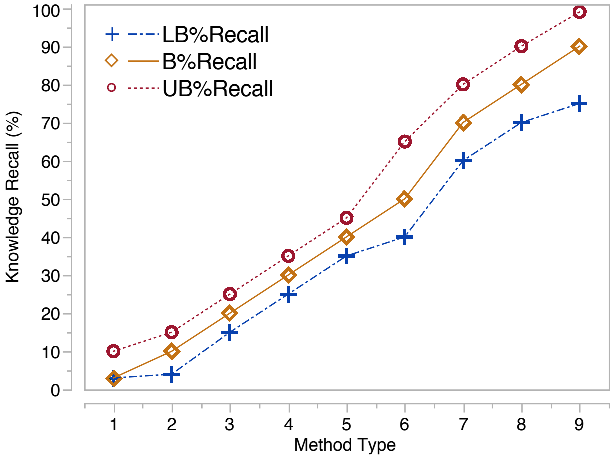
	Cognition			Affective		Psychomotor				Training Method									Training Media/Asset (Hourly Costs)									Resource	
	KC	AA	SE	No	Yes	No	Low	Mod	High	GT	LD	CBT	ICW2	ICW3	ATD	ACS	SIM	ACF	$0	$5	$10	$10	$75	$150	$2,500	$1,000	$7,500	No	Yes
Level	1	2	3	0	1	0	1	2	3	1	2	3	4	5	6	7	8	9	0	1	2	3	4	5	6	7	8	0	1
Cog-1		0	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Cog-2			0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Cog-3				1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Aff-0					0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Aff-1						1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Psy-0							0	0	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Psy-1								0	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Psy-2									0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Psy-3										1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
TM-1											0	0	0	0	0	0	0	0	1	1	0	0	0	0	0	0	0	1	1
TM-2												0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	1
TM-3													0	0	0	0	0	0	0	1	1	0	0	0	0	0	0	1	1
TM-4														0	0	0	0	0	0	0	0	1	0	0	0	0	0	1	1
TM-5															0	0	0	0	0	0	0	0	1	0	0	0	0	1	1
TM-6																0	0	0	0	0	0	0	0	1	0	0	0	1	1
TM-7																	0	0	0	0	0	0	0	0	1	0	0	0	1
TM-8																		0	0	0	0	0	0	0	0	1	0	0	1
TM-9																			0	0	0	0	0	0	0	0	1	0	1
TA-0																				0	0	0	0	0	0	0	0	1	1
TA-1																					0	0	0	0	0	0	0	0	1
TA-2																						0	0	0	0	0	0	1	1
TA-3																							0	0	0	0	0	1	1
TA-4																									0	0	0	1	1
TA-5																									0	0	0	1	1
TA-6																										0	0	0	1
TA-7																											0	0	1
TA-8																												0	1
R-0																													0
R-1																													

Determine the effectiveness (knowledge recall) of each instructional method (2.C)

The effectiveness of each training method in this case study was determined from a conglomerate of sources. To generate baseline knowledge recall values and bounds for each training method, the author initially used the NTL learning pyramid. Input was then solicited from SME, Dr. Steven Tourville. Dr. Tourville has over 25 years of experience in the research, development, and engineering of training systems, including ten plus years of operational experience with the USAF as an instructor and evaluator on multiple

C-130 variants. He participated in an experiment that evaluated 225 West Point students for effectiveness of different learning methods and media. This experiment tested cognitive, affective, and psychomotor LOs. TM-1 through TM-9 were included in the trials. Marginal effectiveness resulted from the methods only involving cognitive ability. When psychomotor and cognitive interaction were combined the effectiveness increased and the best performance resulted when the students were engaged in all three learning domains. The experiment discovered that the level of immersion in an activity is almost directly proportional to the level of effectiveness. As the media type changed there was a change in the effectiveness, but it was not exactly linear. Comparing information collected from the West Point student experiment with the NTL learning pyramid, the boundaries and baseline values of knowledge recall for each training method was established as shown in Figure 5.1. Some training designers would argue that flying the aircraft, TM-9, has an upper bound of 100% effectiveness. However, there are factors that confound the effectiveness of an actual flying mission in terms of individual task performance. It was decided to set the upper bound of TM-9 to 99% knowledge recall to highlight that uncertainty exists in training and even though a person is trained using live equipment and mission scenarios, there is not a guarantee that training always translates into perfect execution. Knowledge recall is stochastic. Each pilot will not remember exactly the same amount of information, but with the data from the learning pyramid and West Point study a range of expected performance is estimated. To capture this variation in performance, the training methods are represented as a triangular distribution and random variable sampling is used in the analysis process. The lower bound, upper bound, and most likely value for each training method is shown in Figure 5.1.

The compatibility matrix generated in Table 5.2 and 5.3 assumes that each training method and corresponding media device(s) can be used to teach LOs at any competency level. Given an unlimited amount of time and budget, this assumption is theoretically correct. Practically, the design team limits the interchangeability of the training methods based on their effectiveness and level of student engagement. This also prevents violation of MPEET assumptions #1 and #2. Assumption #1 assumes that the time provided for the baseline training lessons is correct based on the expertise of the instructional designer. MPEET does not assess the time allocated for training. Assumption #2 assumes that the time spent using one training method is equivalent to the time required to learn a lesson with any other training method. MPEET accounts for the difference in effectiveness between training methods, but does not include an algorithm to assess increases or decreases in time spent on a training lesson. Table 5.4 lists the training method options that can be used when trading training methods to maximize effectiveness.



TM	LB (*a*)	B ***(c)***	UB (*b*)	Student Engagement Level
1	3%	3%	10%	Passive
2	4%	10%	15%	Passive
3	15%	20%	25%	Passive
4	25%	30%	35%	Min. Active
5	35%	40%	45%	Min. Active
6	40%	50%	65%	Min. Active
7	60%	70%	80%	Min. Active
8	70%	80%	90%	Active
9	75%	90%	99%	Active

Figure 5.1: Effectiveness of Training Methods – Pilot QTP

Table 5.4: Training Method Alternative Limitations – Pilot QTP

Training Methods	Options
TM-1 (GT)	TM-1, TM-2, TM-3
TM-2 (LD)	TM-2, TM-3
TM-3 (CBT)	TM-3, TM-4
TM-4 (ICW2)	TM-4, TM-5
TM-5 (ICW 3)	TM-5, TM-6, TM-7
TM-6 (ATD)	TM-6, TM-7, TM-8
TM-7 (ACS)	TM-7, TM-8, TM-9
TM-8 (SIM)	TM-8, TM-9
TM-9 (ACF)	TM-9

5.2.3 Generation of Feasible Alternative Training System Design

The attributes that may influence training effectiveness included in the C-130J Pilot QTP are: learning objectives (LOs) in the cognitive and psychomotor domain, use of an instructor as a resource, and difficulty, importance, and frequency (DIF) ratings. LOs are not broken down by competency level for the affective learning domain, but each training lesson does identify whether affective training is involved. Affectivity in this QTP involves value judgment, communication, and situational awareness. The DIF ratings contained in the QTP are dependent and primarily based on the rating for difficulty. The difficulty rating was determined by SME input. The importance and frequency ratings were set equal to difficulty. This process is inconsistent with all of the DIFE analysis methods found in literature, discussed in section 2.9. The variable sensitivity studies in Step 4 will determine the significance of evaluating DIF independently, and discuss the impact of using this crude method for performing DIFE analysis. The C-130J Pilot QTP did not include a consequence of error rating. Because the rating scheme used for difficulty, importance, and frequency was to set importance and frequency equal to difficulty and the design team used an analysis method without considering consequence of error this attribute was dropped from the overall utility criteria. Crafting inaccurate data for this attribute was not of value in this case study, because independent values for importance and frequency had to be randomly generated later in step 4.C. Adding another variable that the instructional design team did not even consider confounds the results and is more of a detriment than benefit. This case study is includes nine of the ten criteria for predicting training effectiveness.

Convert each attribute into a common utility scale (3.A)

Single attribute utility functions for each attribute in the OEC are defined below using the proportional scoring formula [188]. The goal is to maximize the effectiveness of training and minimize training costs. In terms of training effectiveness, the following criteria are most effective when their values are higher: training methods, competency levels, difficulty, and importance ratings. These attributes have a positive correlation with their utility values. Based on the desire to use the most effective training method for a given cost, the utility values increase from TM-1 to TM-9 based on the increase in knowledge recall. The utility values for each competency level increase as the complexity of the skill increases. It is expected that each student who completes the C-130J pilot QTP will reach the maximum competency levels for each learning objective. This aligns with MPEET assumption #5, which assumes that the desire for the training program being assessed is to reach maximum competency levels for all learning objectives across all learning domains. From the research on DIFE analysis and Figure 2.14, training tasks that are performed frequently on the job and are not difficult are of low priority in formal training courses [182]. If the training task is performed frequently, is difficult, and not important, then training should be given to reach a basic performance standard. Only when a task is difficult, moderately or very important, and infrequent does it become a high priority training activity. In general, the more frequent a task is performed the less formal training is required. Therefore, a higher value of frequency corresponds to a lower utility value resulting in a negative correlation. The two cost attributes, resource and asset costs, are negatively correlated to utility. An increase in cost is undesirable and reflects a decrease in utility. Figures 5.2a and 5.2b contain plots of each of the criteria utility functions, which include tables of the original attribute values and corresponding utility values. Below are the utility functions for each of the OEC variables:

$$Training\ Method\ Utility\ Value = U_1(x_1) = \frac{TM-1}{9-1}$$

$$Resource\ Cost\ Utility\ Value = U_2(x_2) = 1 - \frac{Res-0}{75-0}$$

$$Asset\ Cost\ Utilty\ Value = U_3(x_3) = 1 - \frac{TA-0}{7500-0}$$

$$Affective\ Competency\ Level\ Utility\ Value = U_4(x_4) = \frac{Aff-0}{3-0}$$

$$Cognitive\ Competency\ Level\ Utility\ Value = U_5(x_5) = \frac{Cog-0}{3-0}$$

$$Psychomotor\ Competency\ Level\ Utility\ Value = U_6(x_6) = \frac{Psy-0}{3-0}$$

$$Difficulty\ Level\ Utility\ Value = U_7(x_7) = \frac{Diff-0}{3-0}$$

$$Importance\ Level\ Utilty\ Value = U_8(x_8) = \frac{Imp-0}{3-0}$$

$$Frequency\ Level\ Utility\ Value = U_9(x_9) = 1 - \frac{Frq-0}{3-0}$$

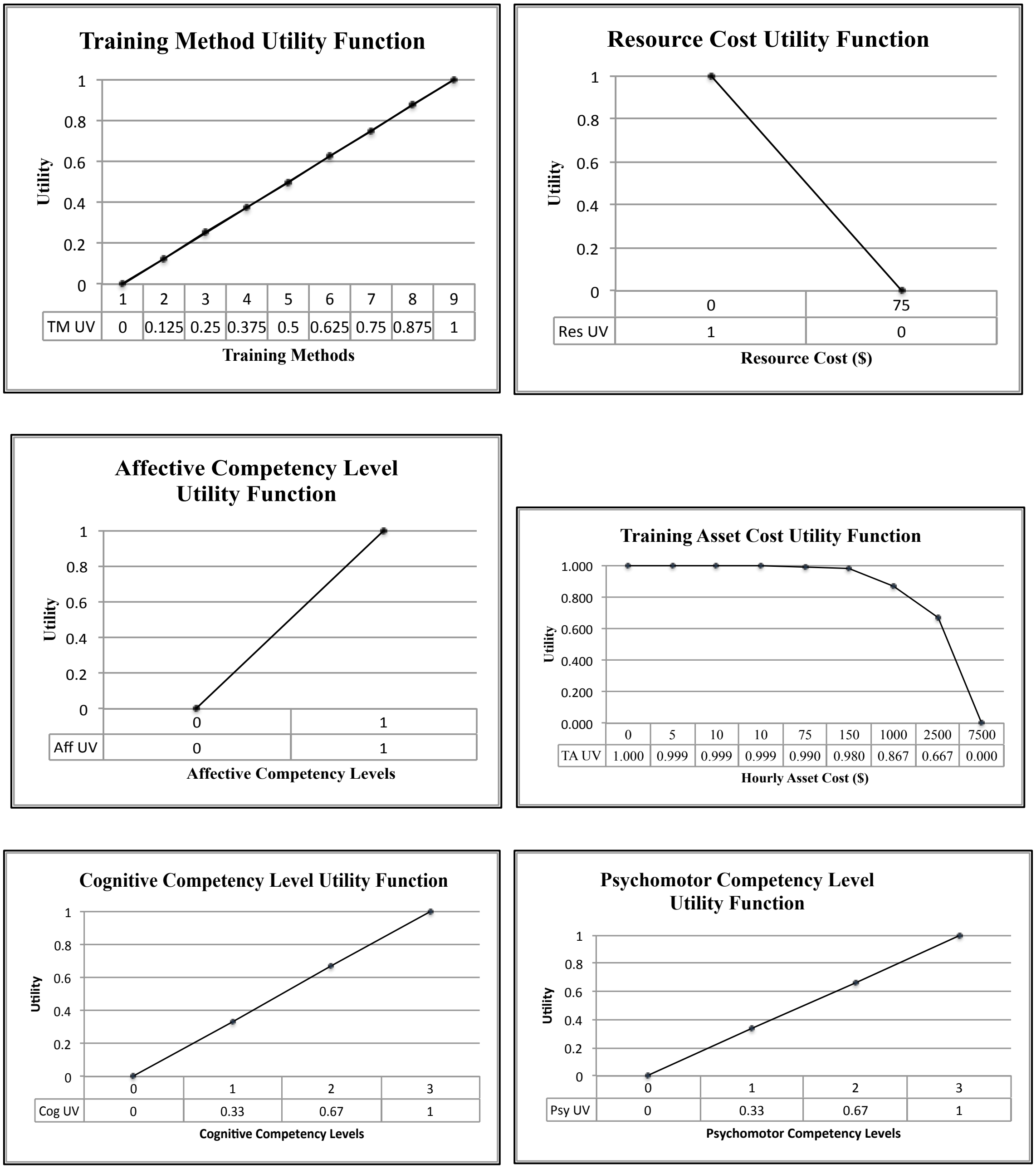


Figure 5.2a: Individual Criteria Utility Functions Part 1 of 2

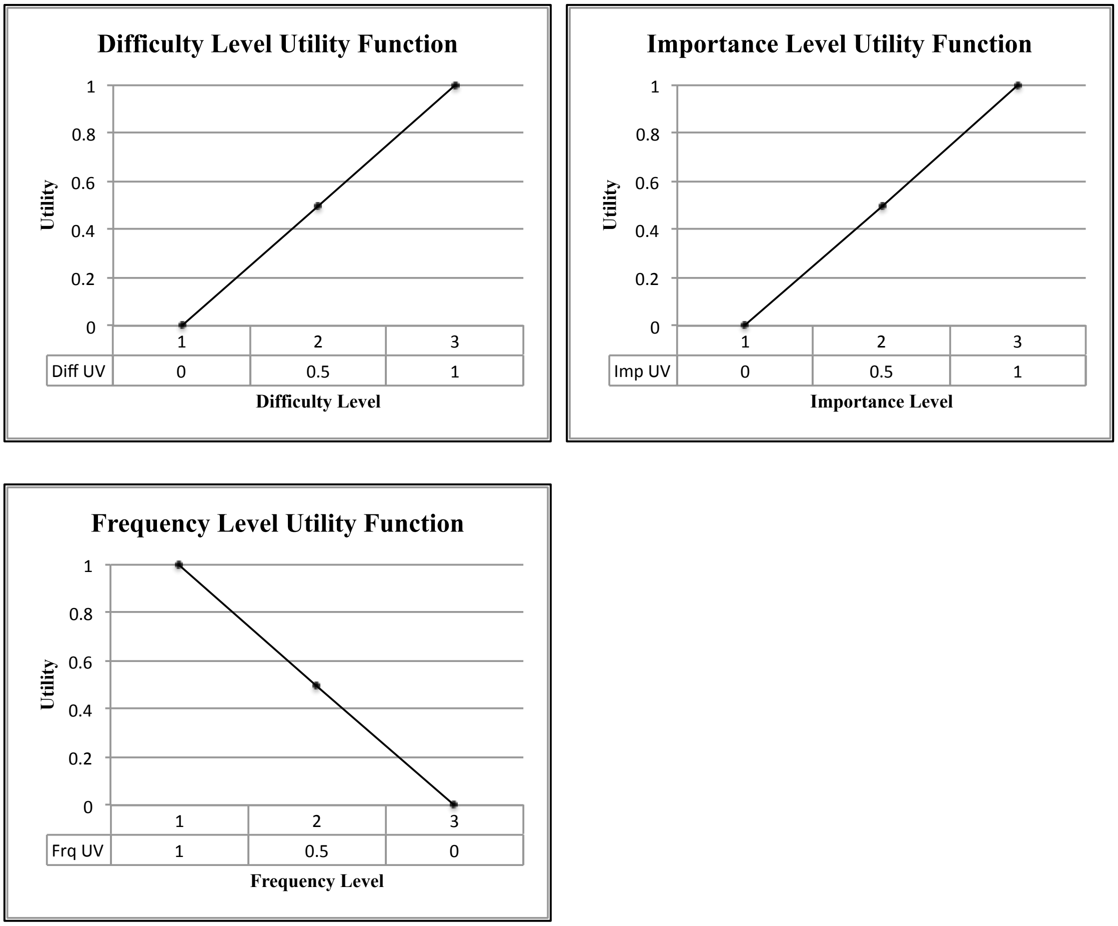


Figure 5.2b: Individual Criteria Utility Functions Part 2 of 2

Determine the importance weightings for each of the attributes (3.B)

The direct method was used to obtain the importance weightings for the nine attributes [188]. The instructional design team was asked to assign weightings based on the importance or influence each attribute has when determining which training method should be used to administer a training lesson. Each attribute could have a weighting between 0 and 100, but the sum of all weightings must equal 100. The weightings are shown in Table 5.5.

Table 5.5: OEC Attribute Importance Weightings – Pilot QTP

Attribute	Weighting Designator	Weighting
Training Method	w_1	10%
Resource Cost	w_2	5%
Asset Cost	w_3	10%
Affective Competency Level	w_4	20%
Cognitive Competency Level	w_5	20%
Psychomotor Competency Level	w_6	20%
Difficulty Level	w_7	5%
Importance Level	w_8	5%
Frequency Level	w_9	5%
		Sum = 100%

Determine the utility of each training alternative using the multi-attribute utility function (3.C)

The single attribute utility functions from step 3.A and importance weightings in step 3.B are combined into a multi-attribute utility function, referred to as the overall evaluation criterion (OEC):

$$OEC = U(x_1, \dots, x_9) = \sum_{i=1}^{9} w_i U_i(x_i)$$

An OEC value was calculated for each training lesson in the baseline C-130J pilot QTP. To determine if the baseline training method for each lesson is the most effective alternative based on the decision-maker preferences, all other possible training methods were evaluated. The compatibility matrices in Table 5.2 and 5.3 provide the relationship mappings between all the training attributes. Table 5.4 lists the training methods that can be used when trading training alternatives to maximize effectiveness. Using the information from Tables 5.2 – 5.4 an OEC value was calculated for each of the alternate training methods. For example, assume a training lesson has the baseline values for each attribute as listed in Table 5.6. Based on the training method alternatives for TM-6 in

Table 5.4, TM-7 and TM-8 are also options for administering this training lesson. Using Table 5.2, all three methods are compatible with the baseline values for the cognitive, affective, and psychomotor competency levels. Using Table 5.3, the asset cost for TM-7 is $2,500 and TM-8 is $1,000. TM-6 is compatible with or without a resource. The baseline training lesson in this example does not use a resource. However, according to the compatibility matrix in Table 5.3, TM-7 and TM-8 are only compatible with a resource. Therefore a resource is required for both TM-7 and TM-8. The utility values for training method (U1), resource cost (U2), and asset cost (U3), are updated based on the compatibility matrices. The OEC is calculated for the alternate methods, TM-7 and TM-8 in Tables 5.7 and 5.8 respectively. The OEC for TM-6 in this example baseline training lesson is 0.4855. The OEC for the alternate methods is 0.4167 for TM-7 and 0.4492 for TM-8. TM-7 and TM-8 have lower overall effectiveness. The training methods themselves are more effective, see Figure 5.4; that is represented by the increased training method utility values, but they also have a higher cost. Based on the importance weightings for the attributes TM-6 would be the best alternative if only considering overall effectiveness (not cost-effectiveness) for administering this example-training lesson because it has the highest OEC value. This same process described for the example training lesson was applied to each lesson in the C-130J pilot QTP. An OEC value was calculated for each training activity in the baseline program, then the applicable training method alternatives were analyzed to calculate their OEC values. The OEC values are used as the overall utility values when calculating the cost-utility ratio.

Table 5.6: Example Baseline Training Lesson Attribute Values and OEC – Pilot QTP

Attribute	**Baseline Value**	**Utility Value**	**Weighting**
Training Method	6	$U_1 = 0.625$	$w_1 = 0.10$
Resource Cost	\$0	$U_2 = 1$	$w_2 = 0.05$
Asset Cost	\$150	$U_3 = 0.98$	$w_3 = 0.10$
Affective Competency Level	0	$U_4 = 0$	$w_4 = 0.20$
Cognitive Competency Level	2	$U_5 = 0.67$	$w_5 = 0.20$
Psychomotor Competency Level	1	$U_6 = 0.33$	$w_6 = 0.20$
Difficulty Level	2	$U_7 = 0.5$	$w_7 = 0.05$
Importance Level	2	$U_8 = 0.5$	$w_8 = 0.05$
Frequency Level	2	$U_9 = 0.5$	$w_9 = 0.05$
OEC = 0.4855			

Table 5.7: TM-7 Alternate Training Lesson Attribute Values and OEC – Pilot QTP

Attribute	**Baseline Value**	**Utility Value**	**Weighting**
Training Method	7	$U_1 = 0.75$	$w_1 = 0.10$
Resource Cost	\$75	$U_2 = 0$	$w_1 = 0.05$
Asset Cost	\$2,500	$U_3 = 0.667$	$w_1 = 0.10$
Affective Competency Level	0	$U_4 = 0$	$w_1 = 0.20$
Cognitive Competency Level	2	$U_5 = 0.67$	$w_1 = 0.20$
Psychomotor Competency Level	1	$U_6 = 0.33$	$w_1 = 0.20$
Difficulty Level	2	$U_7 = 0.5$	$w_1 = 0.05$
Importance Level	2	$U_8 = 0.5$	$w_1 = 0.05$
Frequency Level	2	$U_9 = 0.5$	$w_1 = 0.05$
OEC = 0.4167			

Table 5.8: TM-8 Alternate Training Lesson Attribute Values and OEC – Pilot QTP

Attribute	**Baseline Value**	**Utility Value**	**Weighting**
Training Method	8	$U_1 = 0.875$	$w_1 = 0.10$
Resource Cost	\$75	$U_2 = 0$	$w_1 = 0.05$
Asset Cost	\$1,000	$U_3 = 0.867$	$w_1 = 0.10$
Affective Competency Level	0	$U_4 = 0$	$w_1 = 0.20$
Cognitive Competency Level	2	$U_5 = 0.67$	$w_1 = 0.20$
Psychomotor Competency Level	1	$U_6 = 0.33$	$w_1 = 0.20$
Difficulty Level	2	$U_7 = 0.5$	$w_1 = 0.05$
Importance Level	2	$U_8 = 0.5$	$w_1 = 0.05$
Frequency Level	2	$U_9 = 0.5$	$w_1 = 0.05$
OEC = 0.4492			

Determine the cost-utility ratio for each training alternative and select the most cost-effective training solution (3.D)

The training lesson cost for each activity in the C-130J pilot QTP was calculated using the following equation:

$$Training\ Lesson\ Cost = (Time * Asset\ Cost) + (Time * Resource\ Cost)$$

The costs of the alternative training methods were calculated based on the length of the original lesson, in agreement with assumption #2. If the alternative methods required resources, as in the example from step 3.C, that cost was included. For cases where the original method used a resource but there is an alternative that can be administered without a resource, such as TM-3 (CBT) being an alternative for TM-2 (lecture), the resource cost was properly excluded. The developmental costs for a new resource or asset are included in the hourly cost value, in accordance with assumption #6. Once the baseline lesson and alternative lesson costs were calculated, the corresponding cost-utility ratios (CURs) were computed using the following equation:

$$CU\ Ratio = \frac{Training\ Lesson\ Cost}{OEC}$$

The training methods with the smallest CUR, between the baseline and alternatives, were selected as the most cost-effective solution. The alternative with the smallest CUR is the lowest cost option per OEC value.

The example baseline training lesson from step 3.C will be used to demonstrate this analysis. The OEC for TM-6 was 0.4855. The OEC for the alternate methods was 0.4167 for TM-7 and 0.4492 for TM-8. Assuming the baseline-training lesson is administered for two hours (hr.), the training lesson cost for TM-6 is \$300 ((2 hr. x \$150) + (2 hr. x \$0)). Dividing the training cost by the OEC value results in a CUR of approximately 146 ($300 / 0.4855). Table 5.9 summarizes the results for TM-6, TM-7, and TM-8. TM-6 has the smallest CUR and would be chosen as the most cost-effective solution. In this example TM-6 was the most effective alternative (largest OEC value) and the most cost-effective solution (smallest CUR value). However, the most effective alternative is not always the most cost-effective solution; that is why it is important to consider cost and effectiveness together. For the C-130J training program, 21% of the most cost-effective alternatives were different from the most effective alternative.

Table 5.9: CUR of Example Training Alternatives – Pilot QTP

Training Option Alternatives	OEC	Hours	Asset Cost/hr.	Resource Cost/hr.	Training Lesson Cost	CUR
TM-6	0.4855	2	$150	$0	$300	618
TM-7	0.4167	2	$2,500	$75	$5,150	12,359
TM-8	0.4492	2	$1,000	$75	$2,150	4,786

5.2.4 Alternative Training System Design Evaluation

46.4% of the C-130J original pilot training program is administered using the most cost-effective instructional methods using a 95% confidence level. Training alternatives with a lower CUR were determined for the remaining 53.6% of the training lessons. Figure 5.3b plots the distribution of the training methods for the original QTP versus the new alternate QTP. It reveals that the vast majority of these differing training lessons occur during the use of passive and minimally active training methods. The alternative training program suggests the use of TM-3 in place of TM-1 and TM-2. TM-3 (CBT) has a higher effectiveness and is twice the asset cost of TM-2 (LD), but does not require a resource. In this case the expense of the resource is driving the training lesson costs up and thus the CUR. Investigating the cause of TM-3 (CBT) selection over TM-1

(GT), the higher effectiveness using CBT is the primary reason TM-3 is recommended instead of the original TM-1. The cost for TM-1 is less, but the knowledge recall is so low that the lower OEC causes a higher CUR in comparison to TM-3. The different training alternatives all use a training method that has a higher knowledge recall than the baseline training lesson. The total cost for the training program using the alternative methods is 4% less than the original QTP. This decrease in cost is also driven by a reduced number of resource hours resulting from the alternative using TM-3 (CBT) instead of TM-2 (LD). For 4% less in total training investment cost, over half the training lessons can be administered using more effective andragogical methods. These results are based on the attribute importance weightings given by the instructional design team in Table 5.5, AFI 65-503 cost factors, and the multi-attribute utility (OEC) function created by the author. MPEET assumption #7 should be considered at this point. Learner specific variables are not included in this analysis method. However, the design of the C-130J pilot QTP includes variables that characterize this trainee population, including their previous knowledge, ages, expected learning styles, and military training. Depending on the abilities of the trainees, foregoing traditional lectures in a classroom and just providing all CBT or ICW training may not prove as effective as predicted. One way to implement the alternative training program is to allow all the trainees to complete the initial training lessons via CBT and ICW. Before the students report in for training that requires on-site activity, a test can be given to verify the student has learned the knowledge and comprehension skills required to continue in training. For students who pass they continue with training as planned. Students that demonstrate a lack of understanding can be sent to traditional classroom training. To avoid scheduling issues the traditional classroom training could be held once a month or on a set schedule that is less regular than the current process. This would provide cost savings by reducing the number of students attending traditional classroom training, which reduces the number of required instructor resource hours.

Selecting the instructional methods with the lowest CUR for each training lesson created the C-130J alternate QTP, whether that is the original method used or a generated alternative. To determine if the new training program is statistically different than the original C-130J pilot QTP a matched pairs t-test was used to compare the difference between the CURs for each training lesson. At a 95% confidence level, the paired t-test results showed that on average the CUR of the new training program was less than the original QTP by 211.76. The small p-value (Prob > |t|) for the mean difference shown in Figure 5.3a indicates that this difference is statistically significant. The following summary statistics are shown for both the original and alternative QTPs in Table 5.10. The CUR expected value for the alternative training program is smaller than the original QTP, 2041.6 versus 2253.3. The standard deviations are very close, 7390.7 and 7413.3. From the results of the matched pairs t-test in Figure 5.3a, the mean standard error deviation was statistically insignificant. In Table 5.10, the minimum and maximum values of the CUR are equivalent. The upper and lower 95% mean values are the 95% confidence limits about the mean. They define an interval that is likely to contain the true population mean. The alternative training program is better than the original QTP based on the CUR summary statistics. There is 95% confidence that 46.4% of the original training program is cost-effective. The cost-effectiveness of 53.6% of the original training program can be improved. Therefore, 46.4% of the training lessons from the original QTP are contained in the alternative QTP. The other 53.6% of the alternative QTP is comprised of the most cost-effective instructional methods generated from the feasible alternatives.

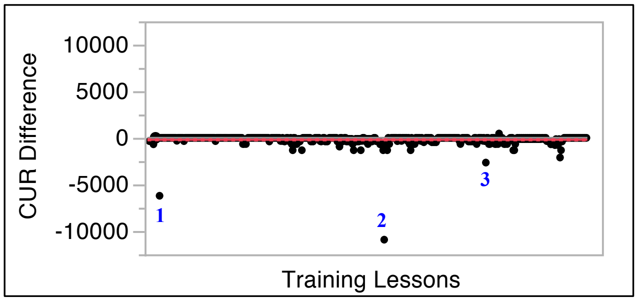
Table 5.10: CUR Summary Statistics – Pilot QTP

CUR Summary Statistics	Original QTP	Alternative QTP
Expected Value	2253.3	2041.6
Standard Deviation	7413.3	7390.7
Minimum Value	0	0
Maximum Value	47531.3	47531.3
Upper 95% Mean	2955.2	2741.3
Lower 95% Mean	1551.5	1341.9

Figure 5.3a contains a Tukey mean-difference plot of the CUR difference for each training lesson. The CUR difference between the new QTP and original QTP is plotted along the y-axis. The x-axis represents the training lessons in chronological order. There are three training lessons that are extreme outliers in comparison to the rest of the training program. Further investigation into these lessons determined that the values are valid. The left most outlier, labeled 1, is an aircraft familiarization task that uses the standing aircraft (ACS) in the original QTP. Based on the variable importance weightings, knowledge recall, and hourly cost difference for ACS and SIM the alternative QTP recommends using the simulator. The cost difference is the primary reason for the large difference in the CUR. The ACS hourly cost is $2,500/Hr. and the simulator is $1,000/Hr. The longer the ACS is used the greater the cost difference between these two instructional methods. This is reflected in the CUR difference plot. Pt. 1 where the training task is two hours long and pt. 2 is a four-hour training task. The training task represented at pt. 2 also used the ACS in the original QTP and the new alternative

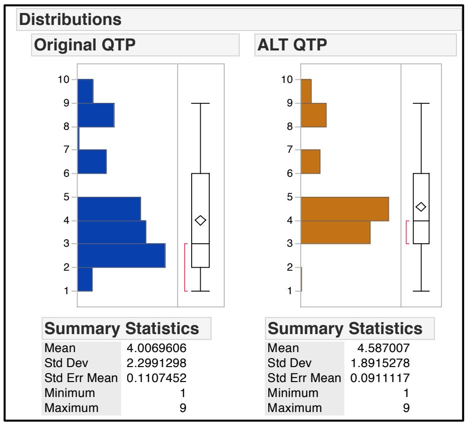
recommended the use of the simulator. The third point, labeled pt. 3 is an eight-hour night vision goggles training lesson originally taught via lecture and discussion in the classroom. The alternative QTP uses computer-based training for this activity. Lecture and discussion requires an instructor, which costs $75/Hr. plus the $5/Hr. classroom asset costs. In an eight-hour training session the cost for that lesson is $640 ($75/Hr. x 8Hr. + $5/Hr. x 8Hr.). In comparison computer-based training eliminates the instructor and has a $10/Hr. computer asset costs resulting in $40 total for the same lesson. This $600 cost difference is what is contributing the CUR difference at pt. 3 and making it appear to be an outlier. As an additional check, these three lessons that appeared as outliers were removed from the training program for statistical analysis purposes, the paired t-test results still concluded that the new alternative QTP is significantly different than original. The 95% confidence region does not include 0, which is another indicator that the two training programs are statistically different, as depicted in Figure 5.3a.

a)



New QTP CUR Mean	2041.56	t-Ratio	-6.52152
Original QTP CUR Mean	2253.32		
Mean Difference	-211.76	Prob > |t|	<.0001
Std Error	32.4711	Prob > t	1.0000
Upper 95%	-147.94	Prob < t	<.0001
Lower 95%	-275.58		

b)



Summary Statistics

	Original QTP
Mean	4.0069606
Std Dev	2.2991298
Std Err Mean	0.1107452
Minimum	1
Maximum	9

Summary Statistics

	ALT QTP
Mean	4.587007
Std Dev	1.8915278
Std Err Mean	0.0911117
Minimum	1
Maximum	9

Figure 5.3: New vs. C-130J Original Pilot QTP Paired t-test & Distribution Results

Determine how training tasks are distributed between lower and higher level learning objectives (4.A)

MPEET recommends two ways to review how many training tasks fall into lower and higher level training objectives. The initial development of MPEET only suggested the set of plots in Figure 5.4, which looks at the total number or training tasks and reviews the amount or percentage of training tasks that fall within each level. This method does not give any consideration to the length or hours spent in training. These results for the pilot QTP are shown in Figure 5.4. From these graphs the training tasks appear to be dominated by lower order LOs, but do include some LOs of higher order. Over half, 61%, of the training tasks in the cognitive learning domain require knowledge and comprehension skills. 34% of activities involve application and analysis. Only 5% percent of the LOs focus on synthesis and evaluation. Within the affective leaning domain, 12% of the training tasks address emotional learning states. 62% of the training tasks do not involve any motor skill ability. The remaining 38% of training tasks are split among low, medium, and high psychomotor skills. As the psychomotor levels increase the percentage of training tasks at each level decreases by nearly half, from 21% to 12% to 5%.

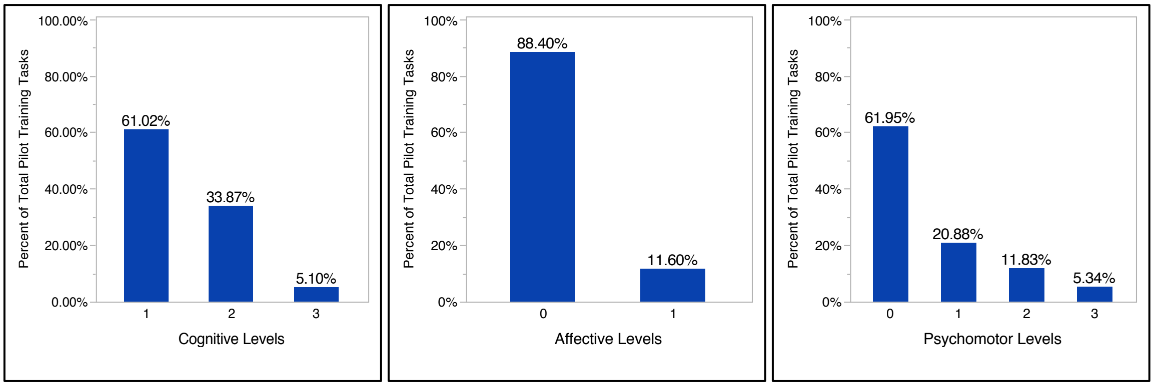


Figure 5.4: Percent of Training Tasks in each LO Competency Level – Pilot QTP

After seeing where the training tasks fit into the learning domains and competency levels independently, it was desirable to see the relationships between the cognitive, affective, and psychomotor domains for each task. If each training tasks fits within a cognitive, affective, and psychomotor competency level, then from Table 5.1 there are 32 possible combinations that a training task can have. Of these 32 possibilities how many are actually used in the training system design? Figure 5.5 shows the relationship between the LOs and competency levels for the training program. From the development of the compatibility matrix in step 2.A, the instructional design team stated that any combination with zero cognition was unrealistic. This resulted in the elimination of the zero or no competency level for cognition, and eliminated eight combinations. It was also determined that no training lesson should be implemented if there was no learning occurring. This meant that the combination of zero cognition, no affectivity, and no psychomotor learning was not compatible. This left 23 remaining possible combinations.

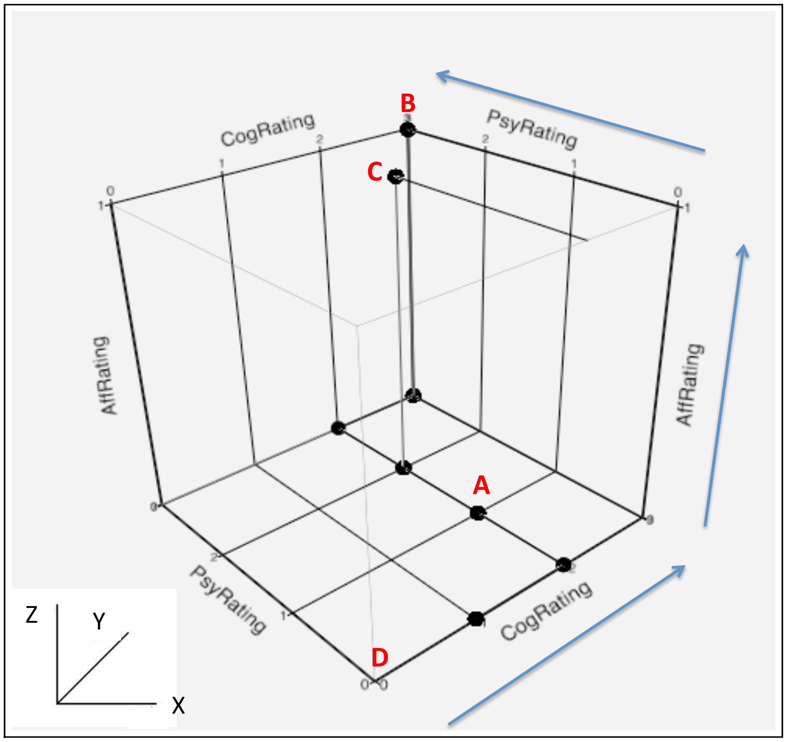


Figure 5.5: Relationships Between LOs and Competency Levels – Pilot QTP in a 3-D plot.

In Figure 5.5, the four cognitive competency levels from the QTP are plotted along the x-axis from front to back (left to right). The four psychomotor competency levels are plotted along the y-axis from front to back (right to left). The two affective levels are plotted along the z-axis from bottom to top. In this training program design eight of the 32 theoretical combinations are present. Although, the realistic number of combinations was determined to be 23, all 32 possibilities should and are plotted for evaluation. The black dots represent these eight combinations. For example, Point A represents training lessons with a cognitive level of 2 (application and analysis), psychomotor level of 1 (low), and 0 (no) affectivity. Point B shows there are training tasks involving the highest levels of all three learning domains, cognitive level of 3 (synthesis and evaluation), psychomotor level of 3 (high motor skill involvement), and affective level of 1 (emotional, situational awareness, and communication skills required). Point C represents training lessons with a cognitive level of 2, psychomotor level of 2 and includes affectivity skills training. Although only eight combinations are used, all 32 combinations are theoretically valid. However, Point D should raise immediate questions in terms of effectiveness. Point D characterizes training tasks with zero cognitive, psychomotor, or affectivity LOs. If there were a training activity at this point, an immediate question should be raised about the purpose this task served, especially if resources were spent on such an activity. Upon first review of this pilot QTP, a training task was listed that had this combination of essentially no learning involved. After discussion with the design team, it was discovered that this task was actually a day of rest provided to trainees as they switched from day to night flying exercises. No resources or assets are expended, but the actual time must be accounted for within the training design curriculum.

This discovery resulted in the second method to review how many training tasks fall into lower and higher level training objectives. Instead of simply looking at the number of, or percentage of, training tasks in each LO competency level, the lessons were charted based upon the number of hours spent in each category, Figure 5.6. When the training program was viewed in terms of training hours versus number of tasks, the distribution within all three learning domains changed. Approximately half, 52%, of the time spent in training focuses on application and analysis (level 2) cognitive ability. 12% of the time trainees are developing synthesis and evaluation skills (level 3) in comparison to what appeared to be less than half that when looking at the percent of cognitive training tasks in Figure 5.4. When comparing the affective and psychomotor competency levels based on hours of training versus training tasks, the same trend is observed. The percentage of *time* spent in training for higher competency levels is twice that of the percentage of *tasks*. This means that there are fewer individual training lessons teaching higher order LOs, but the instructional hours spent during these lessons is longer than the lessons requiring lower level skills.

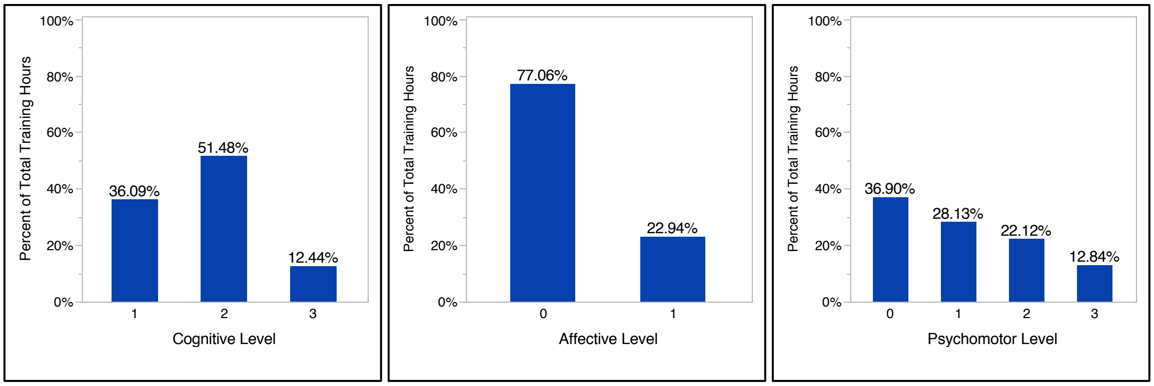


Figure 5.6: Percent of Training Hours in each LO Competency Level – Pilot QTP

Summarizing the training program based on the total tasks and total hours spent developing each learning domain competency level provides insight into the design of the training program. Both methods should be used in determining how training tasks are dispersed throughout each learning domain. These results provide an understanding of the training program in terms of the desired instructional content. The taxonomies used to classify each learning domain and competency level enable the training program to be quantified in terms of learning effectiveness. No effectiveness calculations or conclusions can be made at this step. The observation is that the training tasks vary between all competency levels for all three learning domains with a minimum of 5%. Based on the observations gathered from Figure 5.4 and 5.6 fail to reject hypothesis 1. The success criteria of at least one percent of all training activity requiring the highest level of competency for cognitive, affective, and psychomotor learning has been met. The LOs of this training system design are distributed across lower and higher competency levels. Analysis using MPEET may continue for this case study. If the training tasks had not varied among all learning objective competency levels (< 1%) this indicates that the training requirements are not being met. One of the five criteria identified as necessary to effectively evaluate training is to link training results to mission goals. If the instructional design of the training program does not include all competency levels desired, then this training program will not be effective in preparing students to perform their respective duties. This applies to any training situation and is not limited to this particular case study. Follow the recommendation made in MPEET and notify the design team that the training requirements are not met with the current design and request the training program be revised.

Determine the proper sequencing of instructional methods to reach each competency level (4.B)

The C-130J pilot QTP LO competency levels for all three learning domains are plotted in chronological order in Figures 5.7 – 5.9. Figure 5.7 plots the three cognitive competency levels. For the fist 20% of the training program level 1, KC, is the primary LO. There are a few AA tasks and no level 3, SE, requirements. SE, the highest cognitive competency level, is not introduced until 40% of training has taken place. There is a positive general trend that the training program begins with a lower cognitive competency level, and competency levels increase throughout the training program.

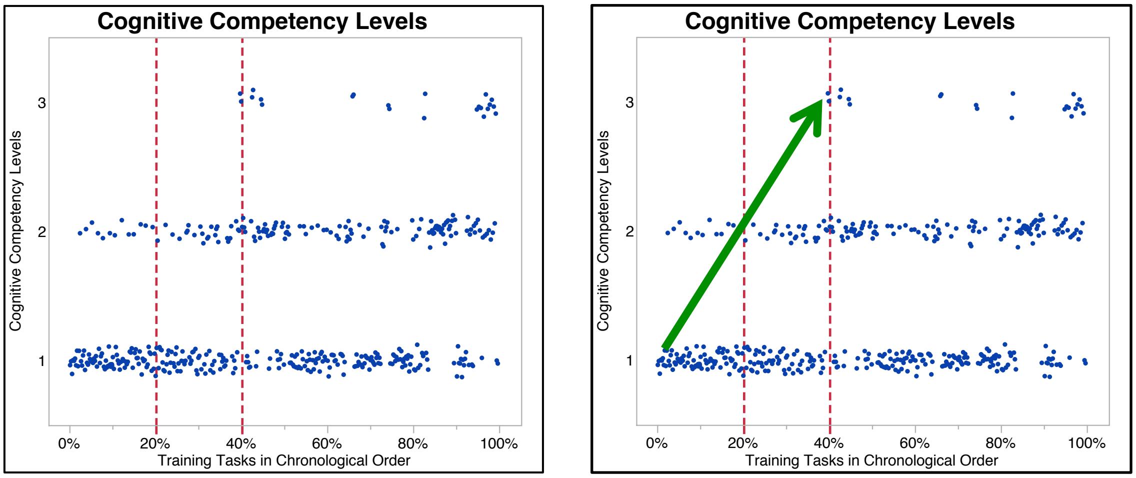


Figure 5.7: Cognitive LO Competency Levels versus Training Task Order

Figure 5.8 depicts the two affective competency levels. Affective LOs are introduced approximately 20% into the training program. They are required from 20-60% of the training program and then again for the last 20% of the course.

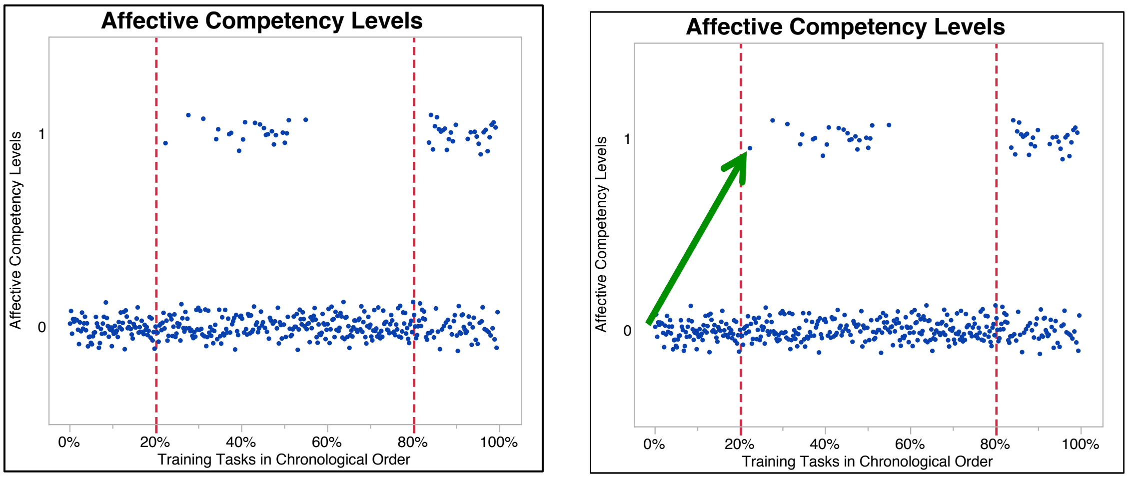


Figure 5.8: Affective LO Competency Levels versus Training Task Order

Figure 5.9 displays the psychomotor competency levels. The first 20% of training hardly involves any motor skills. During 20-40% of training, a low and medium level of psychomotor skills is required. Level 3, high motor skills, is not expected until students have completed 40% of the training program. For training tasks that involve psychomotor learning, there is a clear trend that shows skill development beginning low and building towards higher competency levels over time.

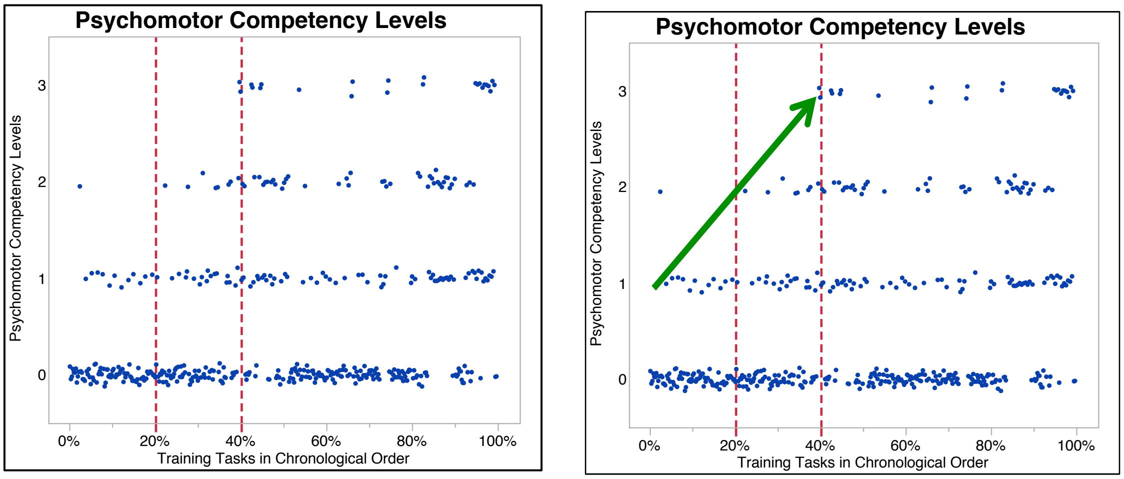


Figure 5.9: Psychomotor LO Competency Levels versus Training Task Order

A positive trend of increasing competency levels throughout training for the cognitive, affective, and psychomotor LOs is observed. The LOs competency levels all begin at a low level. Higher levels are not introduced until at least 20% of the training curriculum is completed.

The training methods (TMs) used in the C-130J pilot QTP are plotted in chronological order in Figure 5.10. Nine different instructional methods are used in this QTP, and are denoted as numbers 1 through 9 plotted along the y-axis in Figure 5.10a. Analogous with the determined effectiveness of each TM, shown in Figure 5.1, instructional method type 1 is a passive method that is the least effective of all nine. As the method numbers increase, their effectiveness increases and they involve more active student engagement. Method type 9 is most effective and is taught using active instructional techniques. The x-axis of Figure 5.10a depicts which method types are used from the beginning (0%) to end (100%) of the training program. This is the chronological order of the training tasks. This figure does not reflect the amount of time spent on each lesson. Figure 5.10b shows the aggregate view of the passive (TM-1, TM-2 and TM-3), minimally active (TM-4, TM-5. TM-6, and TM-7), and active (TM-8 and TM-9) training methods. The original pilot instructional design does not use TM-5 (Interactive Courseware 3) training method, although it is an option. Figure 5.10 shows that the first 20% of training lessons are taught using passive and minimally active methods. The middle of the training program (20-80%) uses a combination of passive, minimally active, and active methods. The use of minimally active instructional methods drops nearly exponentially when active methods are incorporated, as seen in Figure 5.10b. The final 20% of the program uses active and passive methods. At this point in the QTP there is alternation between instruction using passive instructional methods, the student practicing in a highly effective environment, and then the student receives passive feedback and performance evaluations. Overall, the QTP appears to progress from the use of passive towards active training methods. The initial use of passive instructional methods as observed from Figure 5.10 is a good instructional design sequence scheme because it allows students to be introduced to the information and procedures as bystander. They can learn basic concepts, safety precautions, and proper tool handling before they become actively engaged and risk injury to themselves, others, or equipment.

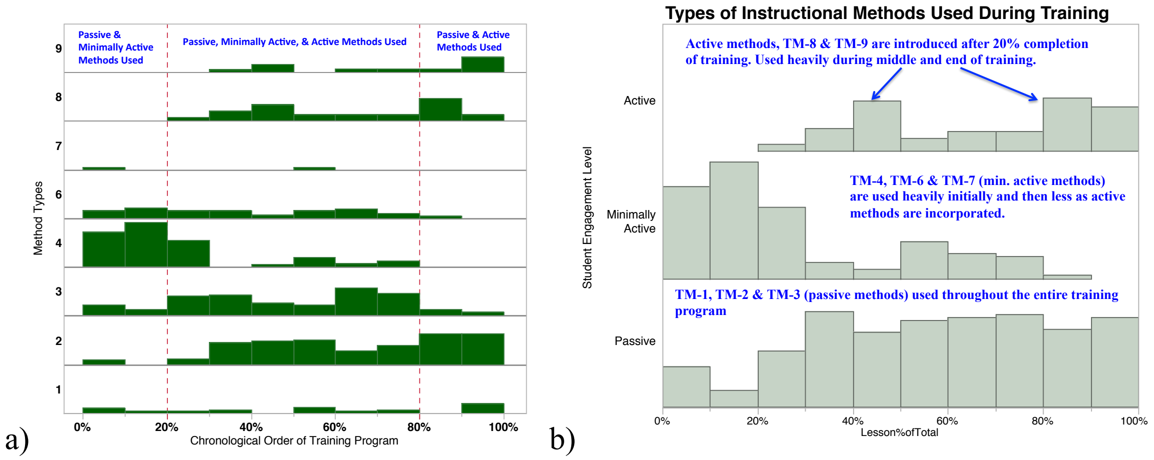
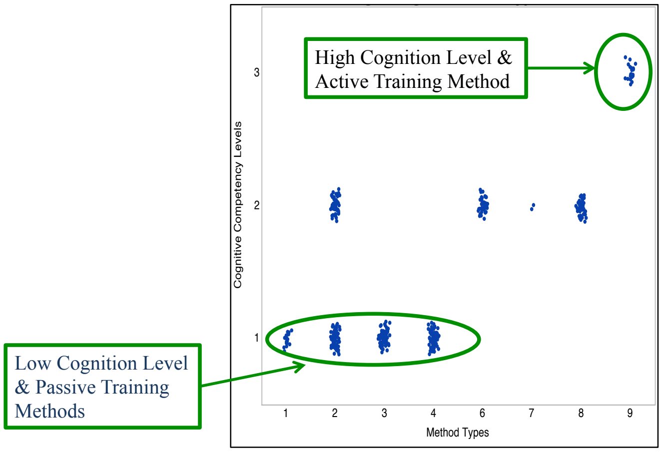


Figure 5.10: Training Methods Used Chronologically – Pilot QTP

Figures 5.11 – 5.16 plot the instructional methods utilized for each LO competency level. These plots test hypothesis 2. The lower competency levels must be administered using passive teaching methods, with a knowledge recall of 49% or less. The high competency levels must use active training methods with a knowledge recall of 50% or greater.



TM	**LB** (*a*)	**B** ***(c)***	**UB** (*b*)	**Student Engagement Level**
1 (GT)	3%	3%	10%	Passive
2 (LD)	4%	10%	15%	Passive
3 (CBT)	15%	20%	25%	Passive
4 (ICW2)	25%	30%	35%	Min. Active
5 (ICW3)	35%	40%	45%	Min. Active
6 (ATD)	40%	50%	65%	Min. Active
7 (ACS)	60%	70%	80%	Min. Active
8 (SIM)	70%	80%	90%	Active
9 (ACF)	75%	90%	99%	Active

Figure 5.11: Cognitive Competency Levels vs. Training Methods – Pilot QTP

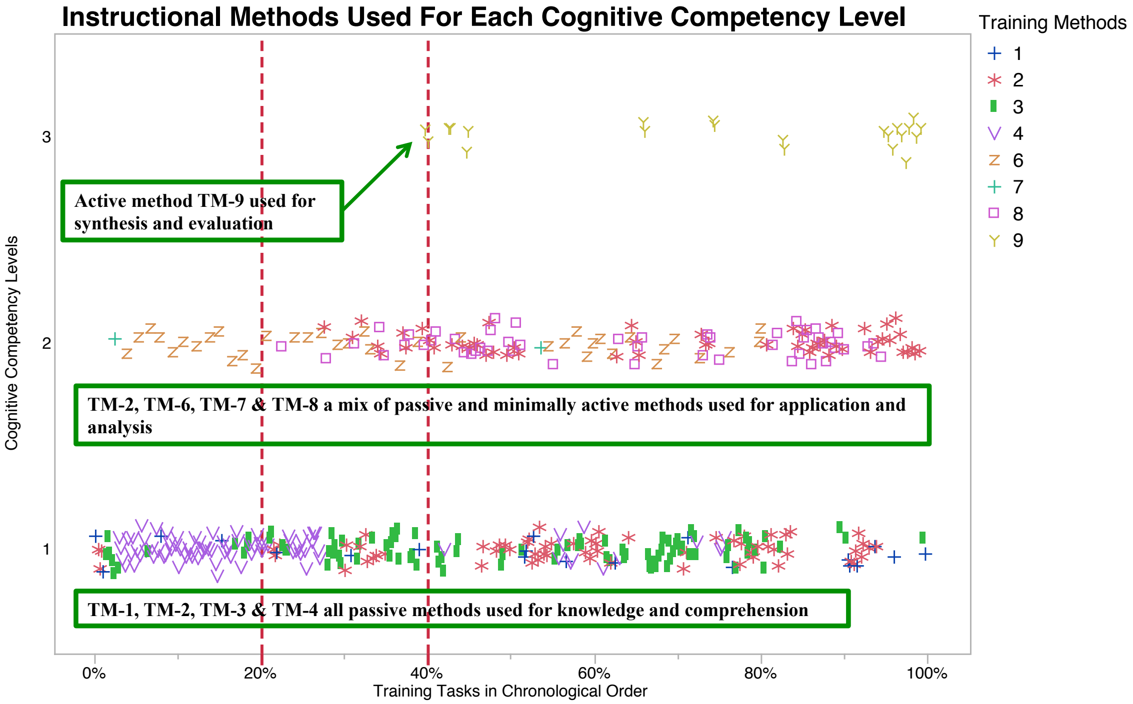
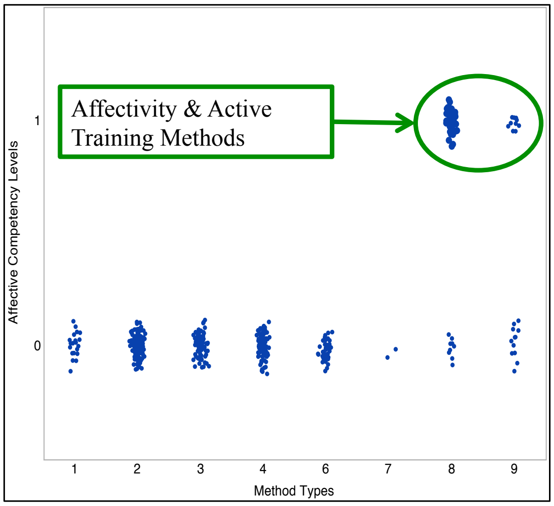


Figure 5.12: Training Methods Used For Cognitive LOs – Pilot QTP

In Figure 5.11, cognitive competency level 1 uses TM-1 through TM-4, all passive training methods (3-35% knowledge recall) for teaching knowledge and comprehension cognitive ability. Level 2, application and analysis, uses minimally active and active training methods. The most active training method, TM-9 (75-99% knowledge recall), is used for cognitive level 3, synthesis and evaluation. Figure 5.12 shows that the most complex or higher order cognitive ability is not included until almost half way through training (40%). This is acceptable because learners will not be capable of synthesizing and evaluating tasks until they have a solid foundation of knowledge and comprehension, and have the opportunity to apply and analyze the learned concepts. Training tasks involving knowledge and comprehension, cognitive competency level 1, and application and analysis, cognitive competency level 2, are spread over the entire training program.

The first 20% of training tasks heavily use passive training methods, in particular TM-4 (interactive courseware 2). This method allows for student independent learning. It uses a computer and does not require an instructor. This is a common method of instruction for adult learners as was discussed in section 2.8.5. TM-3 (computer-based training) is used throughout the training program as well. It is slightly less effective than ICW2 and also does not require the resource cost of an instructor. For application and analysis training tasks during the first 20% of training TM-6 (aircraft training devices) are used as the instructional method. From 20-80% of training task a combination of TM-6 and TM-8 (high fidelity simulation) is used to teach application and analysis cognitive ability. The simulator is more effective than aircraft training devices (ATDs), however ATDs cost a fraction (0.15) of the simulator cost on an hourly basis. Deciding between the use of an ATD or simulator can be heavily based on the weight placed on effectiveness versus cost. This is a good example where the objectivity in the MPEET process can determine the most cost-effective method considering the importance weightings of the decision-maker. From the data in the original QTP it is unclear how that decision was made.



TM	LB (*a*)	B (*c*)	UB (*b*)	Student Engagement Level
1 (GT)	3%	3%	10%	Passive
2 (LD)	4%	10%	15%	Passive
3 (CBT)	15%	20%	25%	Passive
4 (ICW2)	25%	30%	35%	Min. Active
5 (ICW3)	35%	40%	45%	Min. Active
6 (ATD)	40%	50%	65%	Min. Active
7 (ACS)	60%	70%	80%	Min. Active
8 (SIM)	70%	80%	90%	Active
9 (ACF)	75%	90%	99%	Active

Figure 5.13: Affective Competency Levels vs. Training Methods – Pilot QTP

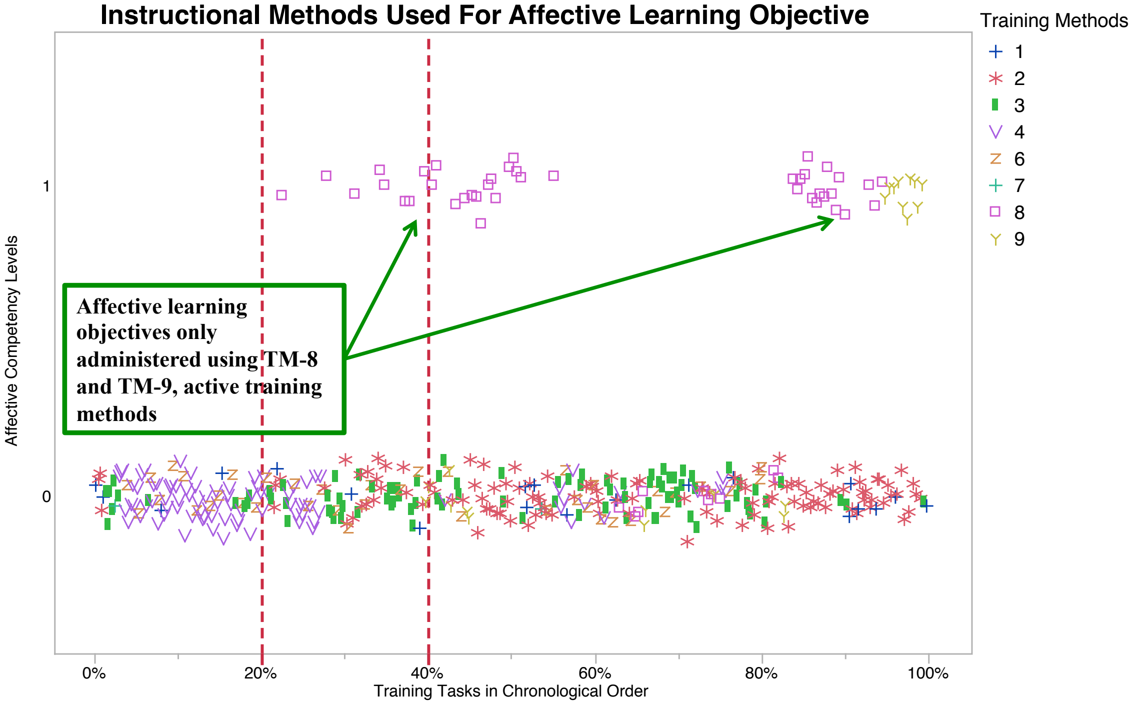


Figure 5.14: Training Methods Used For Affective Learning Objectives – Pilot QTP

Figure 5.13 shows that only active training methods are used for teaching lessons involving affective skills (70-99% knowledge recall). Figure 5.14 shows that affective learning objectives are administered using high fidelity simulation (TM-8) 20% into the training program. From approximately 20-56% of the training program, and towards the end 85-95% of the training program high fidelity simulation is used. The most effective instructional method TM-9, which involves the pilot actually flying the aircraft, is not used for affective learning objectives until the last 10% of the QTP. In comparison to the psychomotor and cognitive higher order skills training in this C-130J training program, lessons involving affective learning objectives are grouped during specific times. In Figures 5.12 and 5.16, the use of TM-8 and TM-9 for higher competency levels starts

40% into the training program, but then is more evenly dispersed throughout the entire QTP. There is a break in the affective skills training during the middle of the QTP, 56-85%. During this time there are training lessons that focus on building complex psychomotor and cognitive skills but not affectivity. Figure 5.14 implies that affective learning objectives can only be administered using active instructional methods because TM-8 and TM-9 are the only methods used in teaching affectivity. However the affective learning domain in the C-130J QTP is not broken down into subcategories. There was not a classification difference in complexity or competency level for situational awareness, task and mission management, or communications. Either the training lesson required them all or nothing. Because no distinction was made in competency levels, the author observed and noted that the only methods used to teach affectivity were active training methods, but does not conclude that these are the only training methods that can be used.

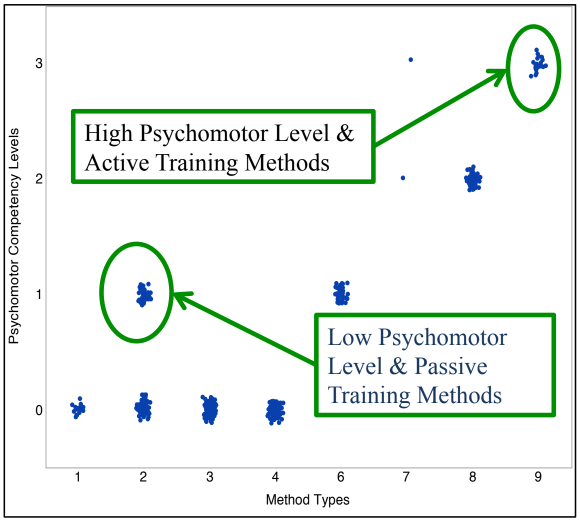
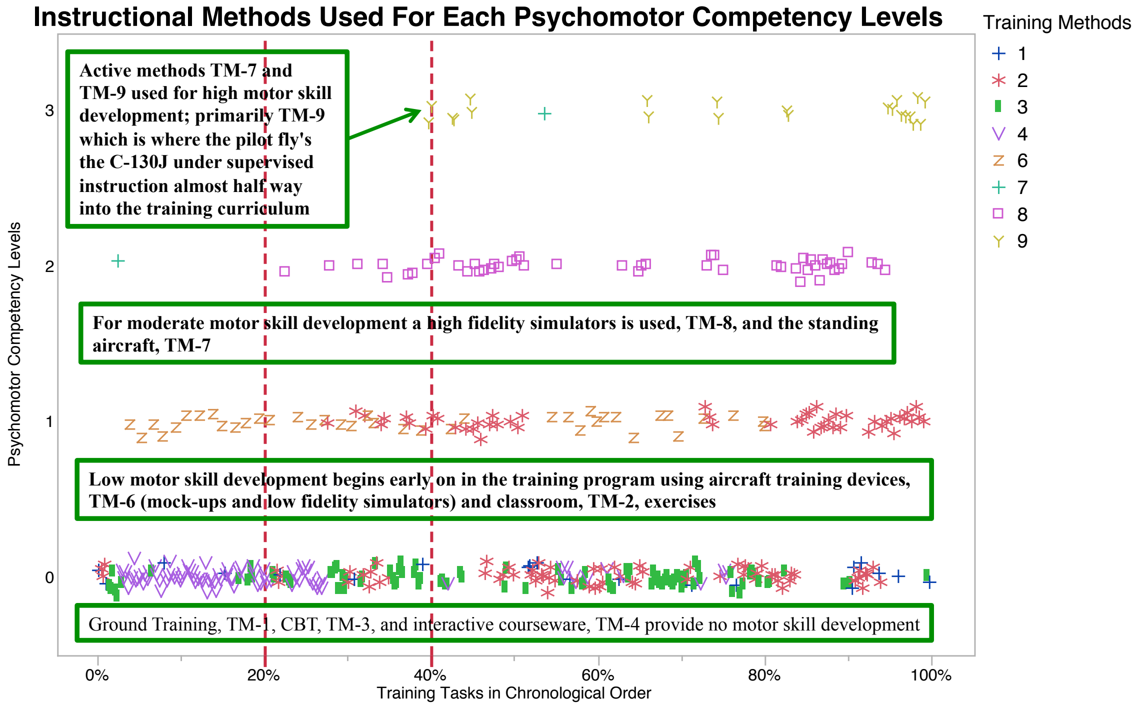


Figure 5.15: Psychomotor Competency Levels vs. Training Methods – Pilot QTP



Active methods TM-7 and TM-9 used for high motor skill development; primarily TM-9 which is where the pilot fly's the C-130J under supervised instruction almost half way into the training curriculum

For moderate motor skill development a high fidelity simulators is used, TM-8, and the standing aircraft, TM-7

Low motor skill development begins early on in the training program using aircraft training devices, TM-6 (mock-ups and low fidelity simulators) and classroom, TM-2, exercises

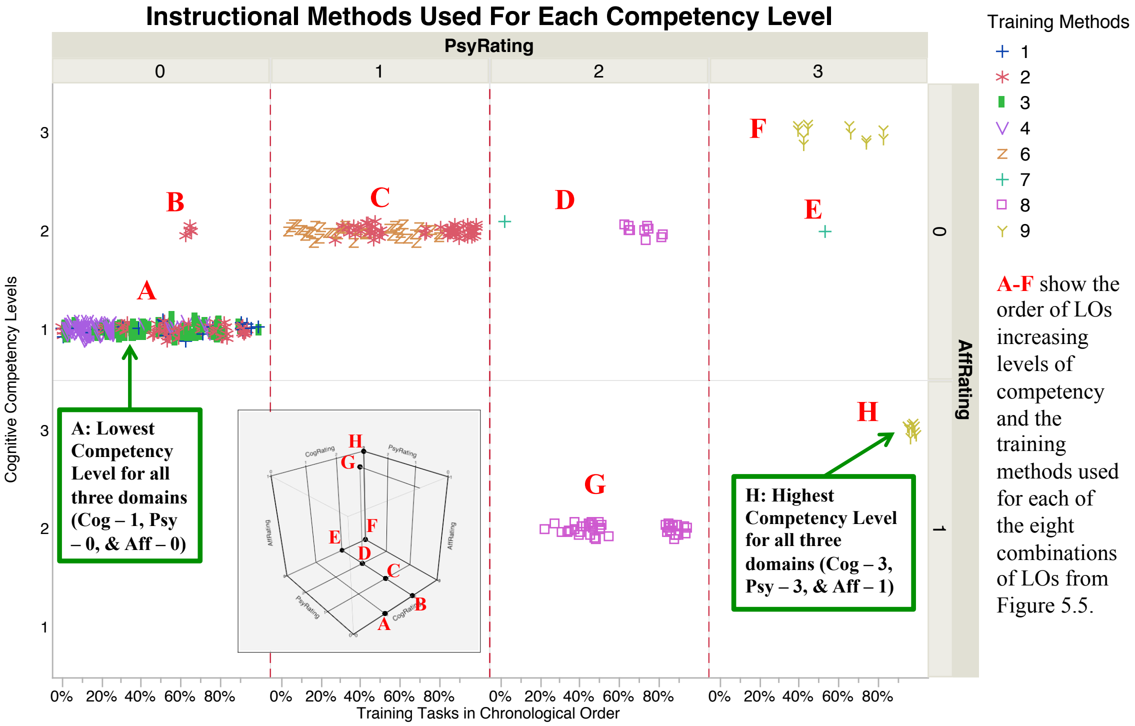
Ground Training, TM-1, CBT, TM-3, and interactive courseware, TM-4 provide no motor skill development

Figure 5.16: Training Methods Used For Psychomotor Learning Objectives – Pilot QTP

The psychomotor competency levels are plotted versus the training methods in Figure 5.15. All training activity that does not involve motor ability is taught using passive training methods (3-35% knowledge recall). As the motor competency level increases, active training methods are used. The highest psychomotor competency levels use training methods that have the most effectiveness, TM-8 and TM-9 (70-99% knowledge recall). One lesson that requires a level 3 psychomotor competency level uses TM-7, which falls into the minimally active range. The knowledge recall for TM-7 is 60-80% and is above the 50% knowledge recall success criteria.

To teach and learn the highest skill level of psychomotor competency the actual on-the-job environment must be used, according to Figure 5.16. Unlike the similar observation made for affective learning objectives, it is logical that active training methods are necessary to instruct high competency level motor skills. Certainly, the trainees should be tested in the environment before there is an expectation of high performance during a mission. From Figure 5.16, the highest psychomotor skill level for a pilot can only be taught on the aircraft. One lesson is taught using the standing aircraft, TM-7, (not actually flying) with the expectation of high motor skill experience and productivity. All other training lessons that involve high motor skills use the flying aircraft, TM-9. For a moderate level of psychomotor involvement the simulator, TM-8, is used, and again one time the standing aircraft is used. For low motor skill development aircraft training devices such as low-fidelity simulators or aircraft mock-ups, TM-6, are used along with classroom training, TM-2, for mission planning, briefing, and debriefing exercises.

In the cognitive, affective, and psychomotor learning domains passive training methods were used to teach non-complex learning objectives and the complex or higher levels of performance are taught using active instructional methods. Figure 5.17 plots the instructional methods used for all three domains and their competency levels. Figure 5.5 is shown in the bottom left corner to remind the reader of the eight combinations of affective, cognitive, and psychomotor learning objectives identified earlier of this training program. Figure 5.17 is an aggregate of Figures 5.12, 5.14, and 5.16. The same observations made from those three previous figures can be seen in Figure 5.17. The difference in Figure 5.17 is that the three learning domains are plotted together and it can be seen that the active training methods are used for the combined highest competency levels (bottom right). The passive instructional methods are used for the lowest competency levels in all three domains (top left). This is true regardless of the learning domain. This information can be used for evaluation of future training programs.



A: Lowest Competency Level for all three domains (Cog – 1, Psy – 0, & Aff – 0)

H: Highest Competency Level for all three domains (Cog – 3, Psy – 3, & Aff – 1)

A-F show the order of LOs increasing levels of competency and the training methods used for each of the eight combinations of LOs from Figure 5.5.

Figure 5.17: Training Methods Used For All Three Learning Domains – Pilot

In the C-130J pilot QTP the lower level competency levels use passive instructional strategies of 35% or less effectiveness satisfying the success criteria of 49% or less in all three learning domains. Therefore, fail to reject hypothesis 4. If any training lesson had been observed where a passive instructional method was used when an active method should have been used resulting in a failure of hypothesis 2, identify a feasible alternative instructional method to use and provide these findings to the instructional design team for updates to the training program. If any training lesson had been observed where an active instructional method was used when a passive method should have been used, a feasible alternative will not automatically be produced. Remember the philosophy that drives the MPEET process is maximum training effectiveness. The active instructional methods have a higher effectiveness (knowledge recall) than the passive methods, so any alternative is going to also be an active method. If this scenario occurs, assume that the higher level of student engagement is required and notify the instructional design team. They can choose to add a training lesson that instructs the trainees using less complex instructional methods to allow the students to build up skills and be prepared for the active instructional method. Or they may move that training task to later in the schedule when the trainee will have developed all the skills required to participate and gain the most benefit from the lessons. The instructional design team may respond and state that the activities leading up to this task are sufficient and based on learner specific variables (not considered in MPEET) active methods can be used to administer lower level competency skills without risk or safety concerns to the trainee, instructor, or equipment. If this is the response, then the success criteria for hypothesis two may be too restrictive for the advanced training program being evaluated. Work with the design team to determine the appropriate percentage of knowledge recall that is expected for the instructional methods as they are used for administering lower and higher level learning objectives. Go back to step 2.C and compare these values to those in the equivalent of Figure 5.1. If there are discrepancies between the knowledge recall and the associated engagement level for the instructional methods used in step 2.C versus what is now determined the two sets of conflicting information need to be resolved. This involves both the evaluation analyst and the design team. If an experiment was not conducted in step 2.C to determine the knowledge recall info, it is recommended that one be performed. If that is not feasible, the reference information used needs to be re-examined and a search for other sources is warranted. After the correct knowledge recall percentage values and applicable hypothesis 2 success criteria are determined, repeat steps 3 and 4.

Determine the sensitivity of the OEC criteria (4.C)

A sensitivity analysis was conducted to determine how sensitive the training alternative predictions were to the importance weightings used in the multi-attribute utility function. The following recommended range of importance weightings were used: [0, 10, 25, 50, 75, 100]. Translating the compatibility matrix into possible training alternatives and adding in DIF ratings results in a full factorial of 1,296 training alternatives. (9 TM Options x 2 Resource Options x 3 Cognitive Levels x 2 Affective Levels x 4 Psychomotor Levels x 3 DIF Levels = 1,296). Instead of evaluating each training lesson at the 1,071 combinations of weight factors[7], an analysis of the C-130J pilot QTP revealed that all of the training lessons can be described with 15 different relationships. Previously, in step 3, the full factorial design was used for evaluations and predictions because the experimental design needed to analyze all possible alternatives to find the most cost-effective solution. The purpose of this sensitivity study is to specifically identify the impact importance weightings have on the attributes. The original QTP is used in this case to prevent confounding results. If the training alternatives were used, then between one and three other variables may have already changed before the weight factors are considered. Using the original QTP, one training

---

[7] C-130J pilot QTP contains nine criteria in the OEC. Using the six recommended importance weighting values, [0, 10, 25, 50, 75, 100], results in over ten million (6^9 = 10,077,696) combinations of importance weightings that could be used in the analysis. However, these values represent relative importance and their sum must always equal 100. Of the ten million combinations the number of valid cases is reduced to 1,071.

lesson that represents each of the 15 relationships must be analyzed. Prediction profilers were used to view the significance of the changes in weighting factor values. Sample prediction profilers are shown in Figure 5.18 for the example baseline-training lesson in step 3.C and 3.D. The slope of the line in each profiler indicates the impact that the weight factors have upon the attributes and OEC value. In the example, Figure 5.18, each weight factor (WF#OEC) shows significance except WF5OEC, which is the weight factor for the affectivity level attribute. In this example affectivity is zero; therefore, any change in the weight factor has no impact on the OEC value.

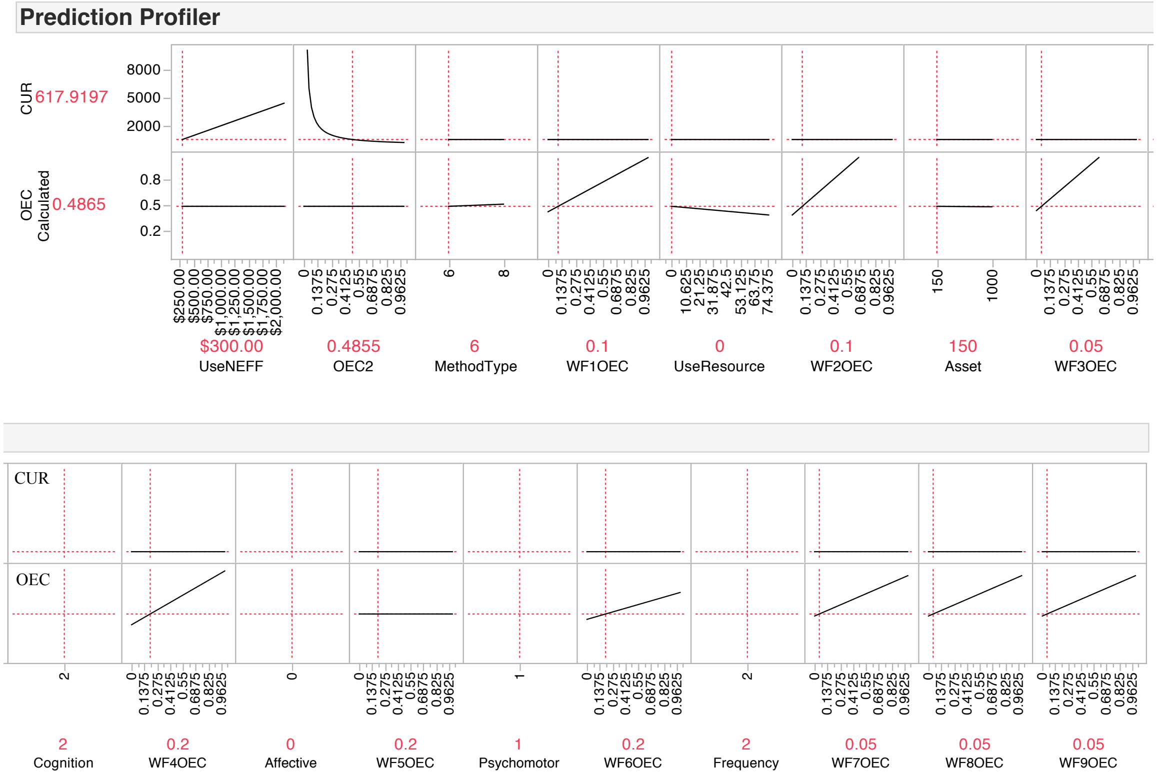


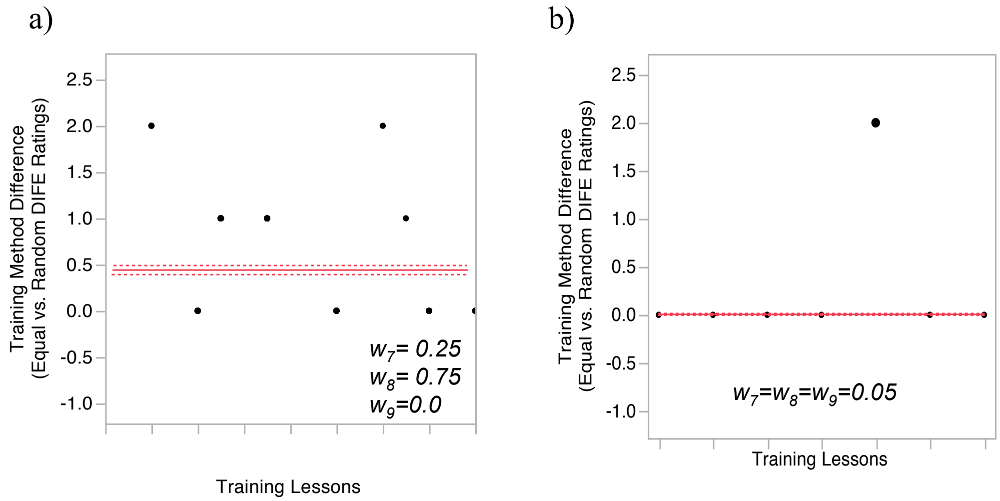
Figure 5.18: Example Prediction Profilers for TM-6 – Pilot QTP

For the C-130J pilot QTP changes in the importance weighting factors did significantly change the training system design, supporting hypothesis 3. The criteria included in the

OEC are all important in determining the cost-effectiveness of a training system design. The criteria are not equally sensitive to the range of importance values assessed. Initial results caused the author to believe that the difficulty, importance, and frequency attributes were all insignificant. However, across the range of importance values these attributes are significant, primarily when the assigned individual importance weight factors for difficulty, importance, or frequency is greater than 0.5.

Although the C-130J data set listed DIF separately, discussions with the instructional design team revealed that the importance and frequency ratings are based on the rating for difficulty. All three DIF had the same values for each training lesson ($r = 1.0$). Therefore, the OEC and CUR results in Step 3.C and 3.D are based on dependent DIF values. The reference documents used to present DIFE analysis and evaluation techniques in Chapter 2 consists of hierarchal lists that asks for a rating for difficulty or importance first and then branch off into frequency and consequence of error [103, 182]. The way the DIFE analysis methods are displayed could lead to inherent correlation. In the C-130J QTP the DIF ratings were deliberately rated dependently. Following MPEET steps, the QTP was checked to determine the impact the correlation between the DIF variables has on the predicted effectiveness and alternative training program results. Unfortunately, a data set with SME input for independent DIFE ratings was not available for the C-130J pilot QTP. The ratings for difficulty are the exact values based on SME input, and were used without any modification. To simulate independent ratings for importance and frequency, random values between low, moderate, and high, were assigned to each lesson. Randomly assigning values for importance and frequency will not provide accurate results for the C-130J training program, but it will provide indications concerning the importance of the technique used when conducting DIFE analyses and how it can impact training effectiveness predictions.

With the OEC criteria set to the baseline weighting values used in step 3.C and 3.D, less than 1% of the recommended training methods differed when the DIF ratings were independently simulated versus the dependent original data. Figure 5.19b shows that there was no statistical difference between the two alternative training programs as a result of assigning random values between DIF ratings when evaluating at the baseline attribute importance weightings. The baseline importance weightings for DIF were each 0.05.



	(a)	(b)		(a)	(b)
ALT QTP DIFE Same	4.80046	4.80046	t-Ratio	17.60228	1
ALT QTP DIFE Randomized	4.35963	4.79582			
Mean Difference	0.4408|	0.00464	Prob > |t|	<.0001	0.3179
Std Error	0.02504	0.00464	Prob > t	<.000	0.1589
Upper 95%	0.49006	0.01376	Prob < t	1.0000	0.8411
Lower 95%	0.39161	-0.0045			

Figure 5.19: Equivalent vs. Randomized DIFE Ratings ALT TM Programs – Pilot QTP

The results of the OEC sensitivity study above showed that the DIF attributes were significant when the importance weighting factor for difficulty, importance, or frequency was greater than 0.5. The model was also run at importance weight factors values that had a combination of 0.75 and 1.0 for w_7, w_8 and w_9, as shown in Figure 5.19a. Approximately 11% or more of the training program differed when the DIF ratings were independently simulated versus the dependent original data and the attribute weight values were increased above 0.5. Based on the results of the match paired t-test assessing DIF ratings independently has a significant impact on predicting training effectiveness when the decision-maker has a high preference for DIF. After discussions with the instructional design team and decision-makers it was determined that to assign an importance weighting of 0.5 or greater to either of the DIF variables was unreasonable. If difficulty or importance or frequency had an importance weight factor of 0.5, then there would only be 0.5 left to distribute over the remaining eight criteria, and the three learning domains will always be given the same or equal value for importance. If the remaining 0.5 is spread evenly amongst cognitive, affective, and psychomotor learning that gives an importance weighting of 0.17 each. The learning domains are always more important than DIF ratings and that preference is not possible if either of DIF is 0.5 or greater. The philosophy of MPEET is maximum effectiveness and minimal costs. Most likely, any variable given a 50% or higher importance weighting would be a cost or learning competency level variable. In the case of the C-130J pilot QTP decision-maker preferences, the DIF attributes are relatively unimportant, so using a method that determines the DIFE rating independently is not imperative for predicting the training effectiveness. For future training evaluations determining the maximum reasonable importance weighting for each criteria can prevent time and effort spent analyzing unrealistic results. However, if the decision-maker did have a preference for assessing DIFE with a high relative importance the training program under evaluation needs to have been designed with independent DIFE ratings for the most accurate cost-effectiveness prediction and alternative training program recommendations.

Figure 5.20 shows the percentage of training lessons at each difficulty level and the training methods used for the C-130J baseline pilot QTP. 36% of the time in training is spent on non-difficult activity. Over half of the training program is spent teaching moderately difficulty tasks, and 12% of the time students are engaged in very difficult lessons. The difficulty rating is directly correlated to the cognitive competency levels, shown in figure 5.6. Further investigation into the C-130J pilot QTP revealed that the cognitive levels assigned are the same as DIF. In terms of using DIF for OEC attributes, they are unnecessary variables for this particular training program. DIF is completely correlated to cognition (r=1.0). DIF provides no additional information and is therefore inutile in predicting training effectiveness for this particular case study.

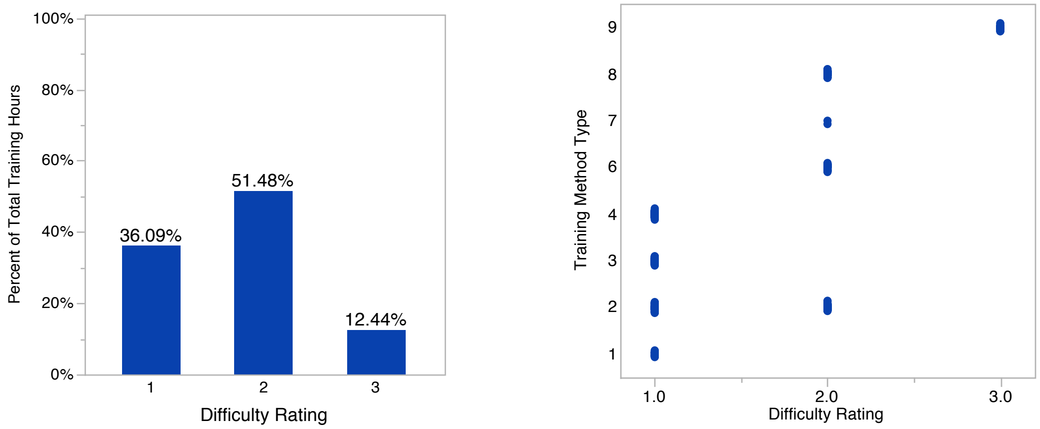


Figure 5.20: Difficulty Rating – Pilot QTP

MPEET is a predictive forecasting tool. In addition to using the results from the OEC sensitivity study to determine proper sequencing of instructional methods and the significance of the criteria and importance weightings, the data can be used to predict what instructional methods are best based on varying decision-maker input. This will make the results robust. In many problems, circumstances change or budgets are cut and the decision-maker input changes. Having an analysis tool that captures the uncertainty in the decision-making process is advantageous. The modeling environment can be used to predict how changes in inputs will affect the recommended training alternatives. In the C-130J pilot QTP example that has been used throughout this case study, the prediction profilers can be appended by adding probability distributions to any of the criteria or importance weighting values. In the example problem from step 3.D TM-6 was found to have a lower cost utility than the alternatives TM-7 and TM-8. To predict when variable changes would result in TM-8 being the most cost-effective alternative the importance weightings for each of the criteria were changed from fixed variables to a random distribution ranging from zero to one as shown in Figure 5.21.

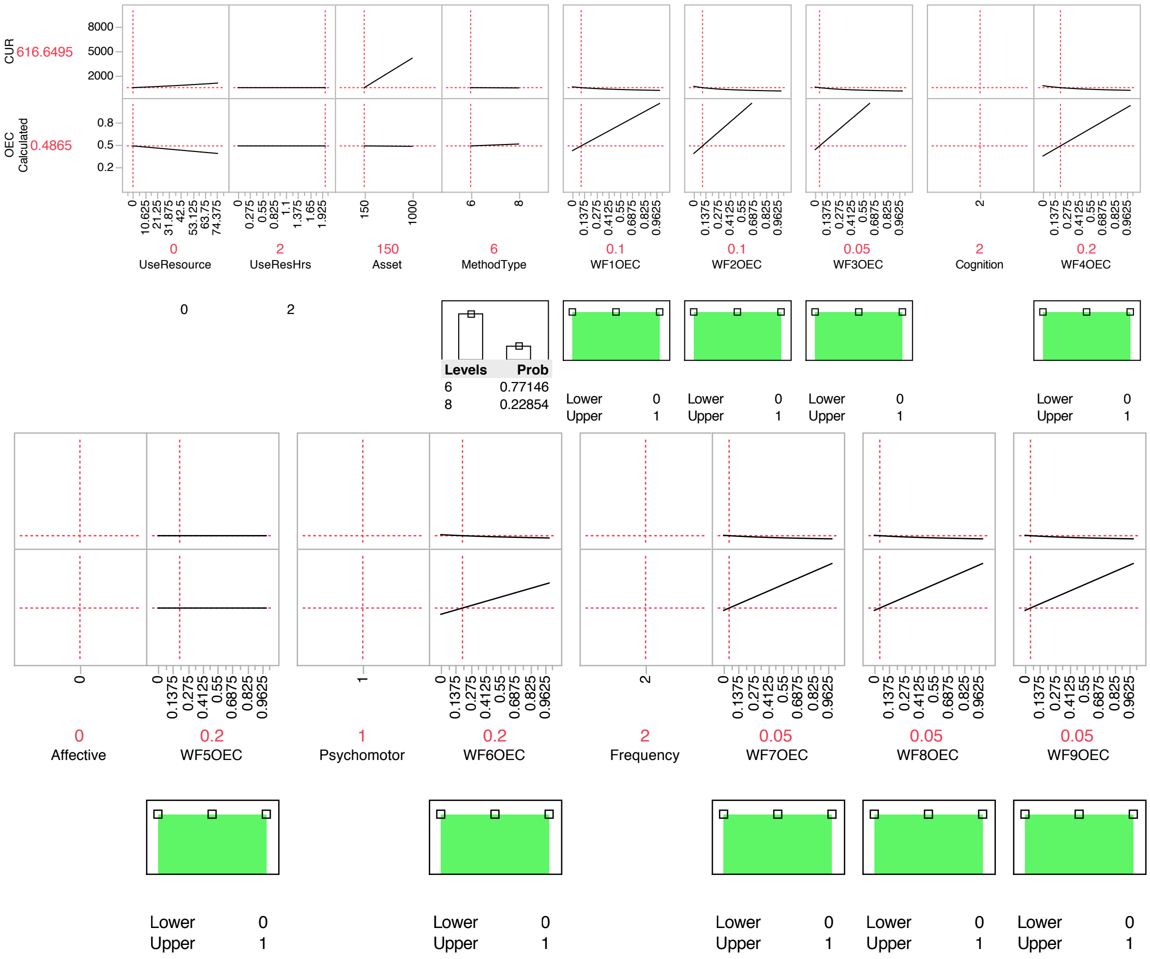
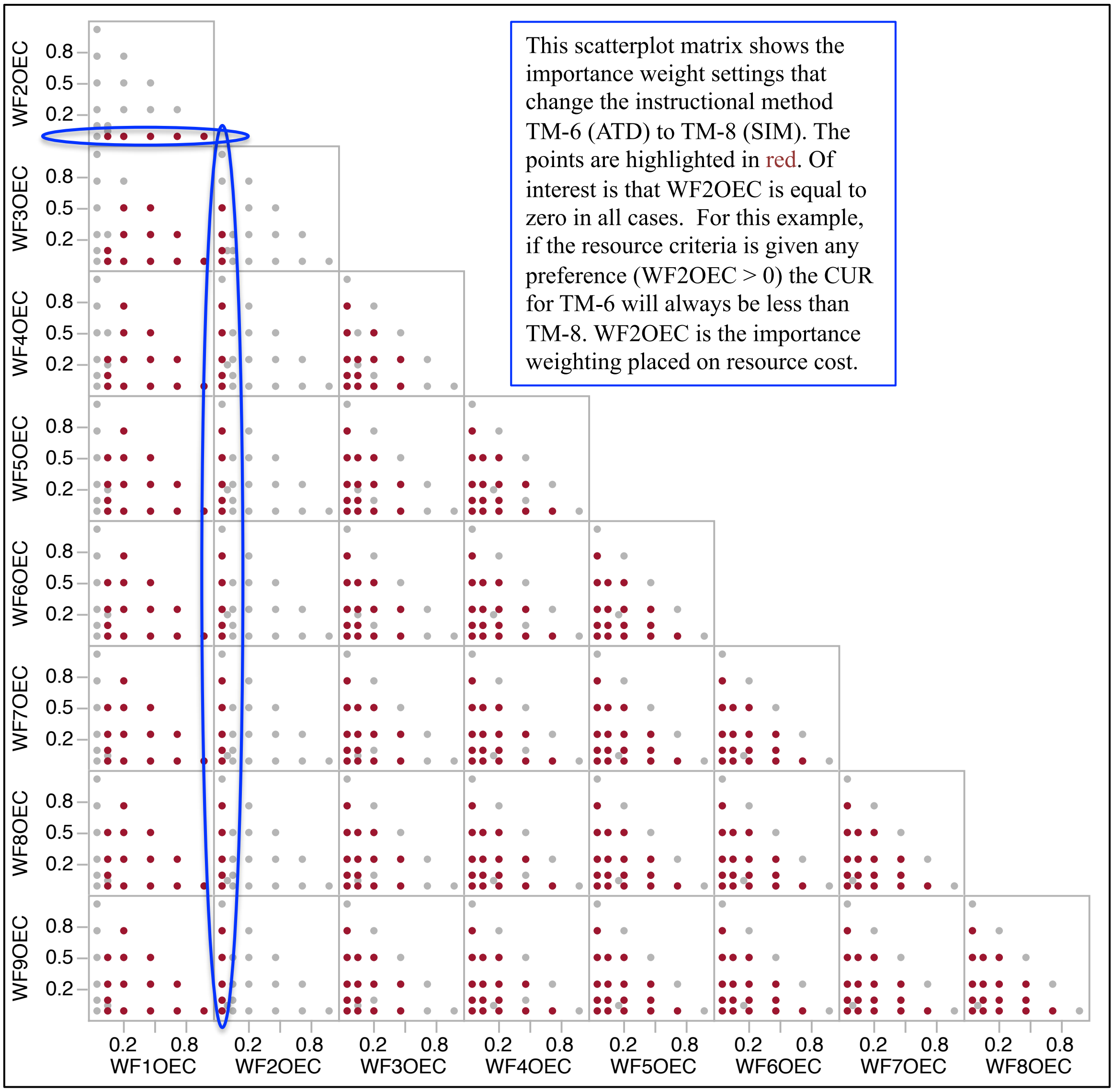


Figure 5.21: Example Monte Carlo Simulation Variable Input – Pilot QTP

A 10,000 case Monte Carlo was run where the simulator randomly selected values for each OEC weight factor, between zero and one, and the CUR was calculated for both TM-6 and TM-8. A scatterplot matrix was formed that consists of the importance weights for all nine criteria, shown in Figure 5.22. This matrix contains all 1,071 valid combinations of importance weightings. A data filter was added to highlight only the points that use TM-8 as the instructional method. At the baseline importance weightings TM-6 had the lowest CUR value. The red points in Figure 5.22 show the settings where the cost-utility ratio for TM-8 is actually less than TM-6. A key observation is that

WF2OEC is zero when TM-8 has a lower CUR than TM-6. WF2OEC is the importance weighting for the resource cost attribute. In this example, any weighting given for resource cost will automatically result in TM-6 being selected over TM-8. Making and documenting these observations during the initial MPEET analysis will save significant time if the decision-maker makes modifications to importance weightings after the initial cost-effectiveness predictions are presented.



This scatterplot matrix shows the importance weight settings that change the instructional method TM-6 (ATD) to TM-8 (SIM). The points are highlighted in red. Of interest is that WF2OEC is equal to zero in all cases. For this example, if the resource criteria is given any preference (WF2OEC > 0) the CUR for TM-6 will always be less than TM-8. WF2OEC is the importance weighting placed on resource cost.

Figure 5.22: Example Scatterplot Matrix for TM-8 Alternative Selection – Pilot QTP

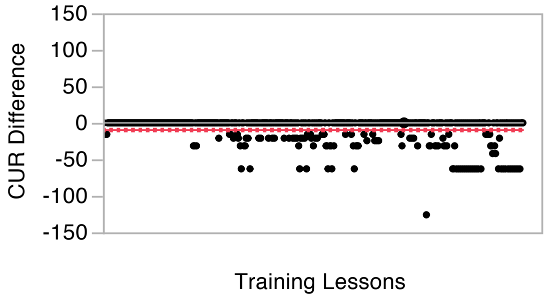
Determine the sensitivity of the training method knowledge recall distribution ranges (4.D)

During this research effort, funding was not available to conduct an experiment to test a sample population of C-130J pilots. As discussed in step 2.C, the knowledge recall values for the most likely value and upper and lower limits were derived from the learning pyramid and SME input [148]. The triangular distribution was decreased and increased by 10%, as applicable. The absolute minimum of 3%, and maximum of 99% are maintained. Training will not be greater than 100% effective, or less than 3%. The two new sets of distributions are shown in Table 5.11, with subscript "1" and "2" along with the original values. A third set of upper and lower bounds was created by combining the 10% decrease and increase of sets 1 and 2, which creates the widest distributed range. To maintain the triangular shape of the original distributions in step 2.C, the baseline value for the third set is proportionally increased. Steps 3.C and 3.D were repeated using the baseline importance weighting values in Table 5.5.

Table 5.11: Expanded Distribution Ranges for TM Sensitivity Analysis – Pilot QTP

	Original (Step 1.e)			**10% Decrease (1)**			**10% Increase (2)**			**10% Decrease & Increase (3)**		
TM	**LB** (*a*)	**B** (*c*)	**UB** (*b*)	**LB** (a_1)	**B** (c_1)	**UB** (b_1)	**LB** (a_2)	**B** (c_2)	**UB** (b_2)	**LB** (a_3)	**B** (c_3)	**UB** (b_3)
1	3%	3%	10%	3%	3%	3%	13%	13%	20%	3%	3%	20%
2	4%	10%	15%	3%	3%	5%	14%	20%	25%	3%	15%	25%
3	15%	20%	25%	5%	10%	15%	25%	30%	35%	5%	20%	35%
4	25%	30%	35%	15%	20%	25%	35%	40%	45%	15%	30%	45%
5	35%	40%	45%	25%	30%	35%	45%	50%	55%	25%	40%	55%
6	40%	50%	65%	30%	40%	55%	50%	60%	75%	30%	48%	75%
7	60%	70%	80%	50%	60%	70%	70%	80%	90%	50%	70%	90%
8	70%	80%	90%	60%	70%	80%	80%	90%	99%	60%	79.5%	99%
9	75%	90%	99%	65%	80%	89%	85%	99%	99%	65%	86.25%	99%

A 10% decrease or increase in the distribution ranges for each training method resulted in a statistically insignificant change in the recommended alternative training program. The OEC is not sensitive to a 10% positive or negative shift in the distribution ranges. When a combined 10% decrease and increase was added to the lower and upper bounds, the resultant alternative training program was statistically different, as shown in Figure 5.23. Considerations that the OEC is sensitive to wide spread variations in knowledge recall should be acknowledged and reported.



ALT QTP Original Step 2.f CUR Mean	2041.56	t-Ratio	-10.6786
ALT QTP 10% D&I CUR Mean	2051.44		
Mean Difference	-9.8795	Prob > |t|	<.0001
Std Error	0.92516	Prob > t	1.0000
Upper 95%	-8.0611	Prob < t	<.0001
Lower 95%	-11.698		

Figure 5.23: New vs. C-130J Original Pilot QTP Paired t-test Results

5.2.5 Decision Support

MPEET predicts with 95% confidence that 46.4% of the C-130J pilot QTP uses the most cost-effective instructional methods. Training alternatives with a lower CUR were determined for the remaining 53.6% of the training lessons. The vast majority of the difference between the original QTP and the recommended alternative training lessons occur during the use of passive and minimally active training methods. The alternative training program suggests the use of TM-3 in place of TM-1 and TM-2. TM-3 (CBT) has a higher effectiveness and is twice the asset cost of TM-2 (LD), but does not require a resource. In this case the expense of the resource is driving the training lesson costs up and thus the CUR. Investigating the cause of TM-3 (CBT) selection over TM-1 (GT), the higher effectiveness using CBT is the primary reason TM-3 is recommended instead of the original TM-1. The cost for TM-1 is less, but the knowledge recall is so low that the lower OEC causes a higher CUR in comparison to TM-3. The different training alternatives all use a training method that has a higher knowledge recall than the original training lesson. The total cost for the training program using the alternative methods is 4% less than the original QTP. This decrease in cost is also driven by a reduced number of resource hours resulting from the alternative using TM-3 (CBT) instead of TM-2 (LD). For 4% less in total training investment cost, over half the training lessons can be administered using more effective andragogical methods. These results are based on the attribute importance weightings given by the instructional design team in Table 5.5, AFI 65-503 cost factors, and the multi-attribute utility (OEC) function created by the author. Learner specific variables are not included in this analysis method. However, the design of the C-130J pilot QTP includes variables that characterize this trainee population. This can also be a factor in the difference between the original and alternative training programs. Based on the results of this case study, the author recommends that the instructional design team and decision-maker consider a reduction in the amount of training lessons taught via lecture and discussion and ground training during the first 20% of the training tasks. If these training tasks can be administered using CBT and/or ICW2, not only would there be a 4% savings in terms of administering the QTP, but with the technology capabilities of web-based instruction it could reduce the time a pilot has to spend training at a particular site location. This could result in additional savings in travel costs and facility fees.

5.3 Observations and Lessons Learned from Case Study

The author used the C-130J Pilot QTP as a benchmark to determine if MPEET could indeed evaluate a training program. Conclusions were drawn after the evaluation that can aid the instructional design process, and the decision-maker responsible for the training program implementation. Findings during the C-130J pilot QTP evaluation using MPEET required the author to make updates and modifications that were incorporated into the process presented in Chapter 4.

As the C-130J QTP was analyzed the importance of viewing the training lesson learning objectives in a multi-dimensional format was highlighted. In step 4.A, when the lessons were initially viewed in table form and 2D plots everything appeared reasonable. Looking at the same set of data in 3D showed a training lesson that had no desired learning objective in any learning domain. A training lesson with no learning objective is questionable because time and money spent that results in no training performance is not a good use of resources. After discussions with the design team it was discovered that this lesson represented a rest day for trainees as they switched between day and night time practice exercises. Another finding that occurred in step 4.A was the necessity to view the how many training lessons fell into each learning competency level in terms of percent of total training tasks and percent of total time in training. Percent of time in training is more useful information that the percent of tasks during training because a training task or lesson may take one hour or eight hours. Determining the number of tasks helps to identify if learning is occurring at all competency levels, which is required to reach maximum training effectiveness, but it does not give a clear picture of the training program. There were no standards discovered in the literature review of Chapter 2 that stated exactly how long a task should take for the student to comprehend or be able to perform at the required competency level. Determining this requires knowledge about the learners themselves, but plots of hours spent in training versus the learning objective competency levels show the dispersion of time spent training the students at each level. This aids in verifying that the training program design meets the training requirements and instructional intent.

MPEET can assess the overall training program by verifying that the design has a general trend of using passive then active instructional methods and that the learning objective competency levels increase over time. In the C-130J pilot QTP a positive tend of increasing competency levels over time was observed. Complex skill levels were not introduced until at least 20% of the training curriculum is completed. In effective training system design passive instructional methods are used during the introductory phase of training to allow students to become familiar with the concepts and tools being used. There was a general progression from the use of passive to active training methods in the C-130J training program. The passive instructional methods are used for the lowest competency levels in all three domains. Active student engagement instructional methods are used for the highest competency levels regardless of the learning domain. Observing when active instructional methods are introduced to trainees, and what methods are used for teaching specific skill levels in future training evaluations adds empirical data to the information found in literature regarding the use of passive and active training methods. The results from the C-130J case study correspond to the data collected in Tables 4.1 and

4.2 These tables mapped learning objective competency levels to training methods based on the results of the Sitzmann et al. study and Morrison et al. instructional design handbook [63, 153]. In the C-130J pilot QTP only active instructional methods were used to teach affective learning objectives. In the future a classification system that breaks down affective complexity levels should be used. It is not a good idea to assume that active training methods always have to be used to teach affective skills. The affective skills involved for the C-130J pilot training are very complex (situational awareness, task and mission management, and communications). Communications is something that could be demonstrated and rehearsed in a classroom before actually putting the pilot in a simulated environment. Task management is a skill that could be broken down and rehearsed using low fidelity equipment. If the affective domain had been classified in a taxonomy that indicated increasing complexity there would be more variation among the instructional methods used for affective learning objectives.

To ensure students are fully capable of performing high competency level psychomotor skills, instructional methods with the most trainee engagement must be used and rehearsed in an environment that represents the actual environment. For the C-130J pilots that was the aircraft itself. To properly prepare students for on-the-job expectations it may require supervised performance in the actual situation and environment. The more complex and higher safety risk fields such as piloting, nursing, surgery, etc. will require practice with feedback in a simulated environment before on-the-job supervised instruction. This pattern is observed in the C-130J QTP and is recommended in instructional design.

The results of MPEET, recommended using CBT and ICW for the entire C-130J pilot training introductory phase. Learner specific variables are not included in this analysis method. However, the design of the C-130J pilot QTP includes variables that characterize this trainee population, including their previous knowledge, ages, expected learning styles, and military training. Depending on the abilities of the trainees, foregoing traditional lectures in a classroom and just providing all CBT or ICW training may not prove as effective as predicted. One way to implement the alternative training program is to allow all the trainees to complete the initial training lessons via CBT and ICW. Before the students continue training a test can be administered to verify that each student has learned the knowledge and comprehension skills required to continue in training. For students who pass they should continue with training as planned. Students that demonstrate a lack of understanding can be sent to traditional classroom training. To avoid scheduling issues the traditional classroom training could be held routinely on a schedule less regular than the current process. This would provide cost savings by reducing the number of students attending traditional classroom training, which reduces the number of required instructor resource hours.

MPEET predicts that 46.4% of the C-130J pilot QTP is administered using the most cost-effective methods. An alternative training program was generated that used only the most cost-effective methods based on the importance weightings provided by decision-maker input. The significance of the training effectiveness criteria was tested using a sensitivity study. The results of the C-130J OEC sensitivity study showed all ten MPEET training effectiveness criteria to be significant. The DIF attributes were significant when the importance weighting factor for difficulty importance, or frequency was greater than 0.5. In general, the decision-maker will not rate any DIFE attribute with an importance value of 0.5 or higher because the cost and learning effectiveness variables are more important. After verifying that the decision-maker does not have a high preference for DIFE, the user may choose to eliminate these four attributes and allow the importance values that would have been assigned to these variables to be added to one or more of the other six OEC criteria.

In addition to sensitivity studies where the limitations of using probabilistic variables such as knowledge recall was determined, the predictive capability of MPEET was demonstrated using multivariate analysis. MPEET can be used to predict the importance weight settings that will result in a change in the instructional method alternative recommendations. This capability adds robustness against the uncertainty in decision-maker preference selection and can be used for other uncertainty variables such as cost in future training program evaluations.

In general, it was discovered that deciding between the uses of various instructional methods could be heavily centered around the weight placed on effectiveness versus cost variables. The objectivity in the MPEET process can systematically determine the most cost-effective method considering the importance weightings of the decision-maker. Because the effectiveness of the C-130J pilot QTP had to be determined before the training had occurred, MPEET used learning classification systems to identify if and to what extent learning objectives were met during training. The taxonomies used to classify each learning domain and competency level enabled the training program to be quantified in terms of learning effectiveness without the need for post-training evaluation results.

CHAPTER 6

SUMMARY & CONCLUSIONS

This research effort was motivated by observing the lack of available information describing how training can be assessed and compared to the other materiel and non-materiel alternatives during the early phase of the defense acquisition process. A gap in the amount of information and detail was discovered in the government reference documents and guidelines for completing CBAs. Information on how to perform a CBA is provided in the following documents: *Appendix A to Enclosure B of the Manual for the Operation of the JCIDS, Capabilities-Based Assessment (CBA) User's Guide* by JCS J-8, and *Capabilities-Based Assessment (CBA) Guide, Version 3.1* by The U.S. Army Training and Doctrine Command (TRADOC) [15-17]. All three of these documents emphasize the requirement to determine if a non-materiel approach can partially or entirely mitigate any identified capability gaps by recommending changes to existing capabilities in one or more of the DOTMLPF-P areas. Definitions of each alternative are given, but details and recommendations of what to include or where to find information to properly assess each area is only provided for certain solution alternatives. Minimal information is provided to a CBA analyst on how to include "training" as an independent variable amongst DOTLMPF-P. The only document that lists examples of what an analyst should consider when evaluating the training alternative during a CBA analysis is the *CBA Guide, Version 3.1* authored by TRADOC. They suggest that the analyst consider several questions such as: "Is existing training being delivered effectively? Are training results being monitored and analyzed for effectiveness? Is training properly staffed and/or funded? Are there training devices, simulators, or simulations that, if developed and fielded, would close or mitigate the gap?" [17]. These questions provide an analyst, who may or may not have a background in training, a place to begin doing research and seek SME input. However, TRADOC does not provide any references or information on how to answer these questions. To date documents that provide assistance for conducting CBAs have yet to address how to include training in the DOTMLPF-P analysis process. Further investigation into the methods used for evaluating training effectiveness within the DoD, revealed that the Government standards and guidebooks all suggested post-training evaluation techniques. If the recommended techniques are all implemented after training has occurred, determining the effectiveness of training as part of a CBA analysis is impossible. The first part of solving this problem was to find an appropriate method for measuring training effectiveness early in the defense acquisition process. This required a thorough literature review of existing training effectiveness models. A set of criteria was proposed that a training effectiveness model must contain to provide results that allow training benefits to be independently considered as a non-materiel solution. The model must 1) connect training results to mission specific goals, 2) be primarily based on objective data (can be supported by subjective data), 3) account for variation of skill levels, 4) include uncertainty analysis, and most importantly 5) has the ability to predict training effectiveness. The method of Bahlis and Tourville fully met four of the five criteria [61]. It lacked an uncertainty analysis for competency level assessment and cost data. With the addition of uncertainty analysis the method of Bahlis and Tourville provides a means for evaluating training effectiveness because it does have a predictive capability. The method of Dietchman fully met three, and partially met the other two criteria [56]. The case study used by Dietcchman provides a demonstrated example that training effectiveness can be predicted, quantified in monetary terms, and compared to the benefits of added force structure or new equipment. There is a difference in the approaches used by Bahlis and Tourville versus Deitchman. The type of information used to measure effectiveness in the case of Deitchman's study was very task specific. He used a traditional method of applying improvement and degradation factors to mission success objectives based on the amount of training time invested [56]. In comparison, Bahlis and Tourville use a qualitative method that prioritizes training tasks and eliminates those that have the least impact within the allocated budget [61]. Although both methods can be used to predict the effectiveness of training, they highlight the fact that different strategies exist. This led to a second primary observation: a framework is needed that can assess training systems, irrespective of the type of training situation. With the development of MPEET, the benefits of training can be expressed in terms of cost and overall effectiveness. This allows an analyst assigned to complete a CBA the ability to use a cost-benefit analysis to compare training as an independent alternative.

The first research question, "What is an appropriate method of measuring training effectiveness during early phase defense acquisition to aid decision-makers in DOTMLPF-P alternative selection," was answered through a literature review of existing methods for evaluating training effectiveness, determining the criteria that a training effectiveness method must meet to be beneficial in a CBA analysis, and the inclusion of these criteria in a new methodology development named MPEET. The MPEET process consists of five major steps similar to the systems engineering decision-making process. MPEET predicts the cost-effectiveness of a training program by determining how well the instructional strategies used in the curriculum design meet the required learning objectives. By using these two primary elements of instructional design, instructional strategies and learning objectives, MPEET eliminates the need for post-training evaluation data, which was one of the capability gaps in most existing training effectiveness methods. The instructional design elements of learning objectives and methods form the basic components for ensuring and increasing human performance during training. MPEET provides a means for understanding and showing the training alternative space. It evaluates the effectiveness and cost of a baseline training system design and creates more cost-effective alternatives for individual training tasks based on decision-maker preferences for the variables used to measure training effectiveness. MPEET is most useful when post-training evaluation information is not available, the cost to collect actual data is infeasible, or there is a desire to mitigate the risk of implementing an ineffective training program. These are common situations when conducting CBAs during the early phases of the defense process. However, MPEET is not limited in application to the DoD acquisition process. MPEET can be used in any case where the assumptions are met, and there is a need to enhance the instructional design process by adding a verification step to objectively evaluate and predict the cost-effectiveness of a training program. MPEET was tested on the design of a C-130J pilot QTP.

The second research question contained two parts regarding how to actually measure training effectiveness. RQ2.1 sought to quantify the benefits of soldiers training in terms of effectiveness, and RQ2.2 questioned how resources should be allocated to gain maximum training effectiveness. These questions were answered through literature review, and experimentation using MPEET. To quantify training in terms of effectiveness, literature was reviewed from training, education, and psychology fields. The best taxonomies for describing the stages of learning and competency were reviewed. Instructional strategies were examined to determine which training methods and media resources resulted in maximum trainee knowledge retention and ability. This literature review also provided answers to research question three, how to quantify increased knowledge, skills, and attitudes in training system design. Based upon the literature review, an overall evaluation criterion was created that enveloped the various attributes of training system design and importance weightings were assigned from the decision-maker preferences. It was hypothesized that the following ten attributes were necessary to predict the effectiveness of a training program: learning objectives in the cognitive, affective, and psychomotor domains, instructional methods, instructional media, use of an instructor as a resource, and difficulty, importance, frequency, and consequence of error ratings. Because there was no standard criterion to use, a sensitivity experiment was conducted on all the criteria to ensure that the selected criteria were indeed relevant in predicting training effectiveness. This experiment was not part of the initial planned work, but observations were made that the difficulty, importance, frequency, and consequence of error (DIFE) weightings appeared to have no effect and was not the cause of any changes in the new training programs. Another sensitivity analysis was conducted to determine how sensitive the training alternative predictions were to the importance weightings used in the multi-attribute utility function. Variance in the assigned importance weightings for each attribute resulted in significant changes in the recommended training alternative program. The decision-maker preferences will have a direct impact on the training methods used and effectiveness evaluation of the training system design. A final sensitivity experiment was performed on the distribution ranges of the knowledge recall expected from each training method. There is disagreement amongst the training and education community regarding the percentage of information a person remembers based upon a certain training method. Therefore, a sensitivity analysis was run to prove that knowledge recall could vary without having a significant impact on the effectiveness of a training system. A 10% increase or decrease in the distribution limits had an insignificant effect in predicting the overall effectiveness of the training system design. A combined 10% increase and decrease resulted in a statistically significant difference in the overall effectiveness of the training system design. If combined variations up to 10% on the upper and lower bounds of the training method distributions are feasible, then the predicted overall effectiveness may be confounded by the training method attribute.

The metrics used in the C-130J case study were those provided as part of the data set. Ideally, one would want to use classification systems that best align with the learning objectives that the student will accomplish in training because these are the best indicators of training benefits. However, using valid data is better than guessing or forcing something to fit a model. In the case of the C-130J data set, the method used for performing the DIFE analysis was rudimentary in nature. Every training task was given the same ratings for difficulty, importance, and frequency. Every task was either assigned a rating of: 1) not difficult, not important, not frequent 2) moderately difficult, moderately important, moderate frequency or 3) very difficult, very important, very frequent. From the literature review, scholars that use DIFE analysis would find this assignment method inadequate [136, 182, 183]. Training tasks are going to vary in difficulty, importance, and frequency. Assuming everything that is difficult is also important and frequent is not representative of all tasks, as explained in section 2.9. As a test to prove if performing DIFE analysis with one of the recommended techniques found from literature has an effect on predicting training effectiveness, an experiment was run that enabled a comparison between the training effectiveness using the baseline data provided versus randomly assigning the DIFE ratings so that each training task had a mixture of not, moderate, and very DIFE. Assessing DIFE ratings independently has a significant impact on predicting training effectiveness when the decision-maker has a high preference for DIFE. If these attributes are relatively unimportant as in the case of the C-130J pilot QTP then using a method that determines the DIFE rating independently is not imperative.

To answer the second part of research question two, "For a given set of monetary resources how should one allocate resources to gain maximum training effectiveness?" a cost utility analysis was performed. A multi-attribute utility function was developed from the baseline importance weightings and combined utility functions for each criteria. The cost of each training alternative was divided by the calculated overall evaluation criteria (OEC) and the alternative with the smallest cost utility ratio (CUR) was selected. The new training program created using this approach provides the most utility at the lowest cost, considering the relative importance of ten OEC criteria. Using the OEC created as part of MPEET the most cost-effective training method alternatives can be determined based on the importance weightings of each criteria (LOs, cost, resource (instructor), method type, DIFE rating). A comparison is made between the original training program and the program developed by MPEET. MPEET predicts the effectiveness of the original training system design based on the number of original training lessons that use the most cost-effective training methods. For the C-130J case study MPEET predicts with 95% confidence that 46.4% of the C-130J pilot QTP is administered using the most cost-effective methods. An alternative training program was generated that used only the most cost-effective methods based on the importance weightings provided by decision-maker input. The difference between the C-130J pilot QTP and the alternative QTP generated during MPEET primarily occurred during the first 20% of the training program. Based on the decision-maker importance weightings for each OEC attribute, MPEET recommended the use of varying levels of engagement via computer-based training (CBT) for the initial 20% of training. The original QTP uses traditional instructor led lectures and discussions in addition to CBT during this same time frame. Using only CBT reduced the total training program cost by 4%, and increased the overall effectiveness because knowledge retention using CBT is higher than classroom lecture although both are typically passive instructional methods. A contributing difference between the original QTP and alternative could be due to the fact that MPEET does not consider learner specific variables in the analysis process. The original C-130J QTP is designed based on historical data of C-130 pilots personal characteristics and demonstrated capabilities. The military pilot selection process is very well understood by the instructional design team and common background information such as learning styles, instructional strategy preferences, previous knowledge etc. is considered. MPEET does not include these learner specific variables and therefore the results and applicability of MPEET are limited.

This initial development of MPEET does not include an evaluation of every variable that impacts training. Constraints in length or time of training and the inclusion of variables that are unique to individual students are not considered. Even without these variables MPEET can enhance the current defense acquisition process by providing an assessment of the effectiveness of the training DOTMLPF-P solution alternative during CBAs. Because the effectiveness of training programs must be determined before the training has occurred, MPEET uses learning classification systems to identify if and to what extent learning objectives will be met during training, and probability analysis to quantify the variable uncertainty. The taxonomies used to classify each learning domain and competency level enable the training program to be quantified in terms of learning effectiveness without the need for post-training evaluation results.

6.1 Summary of Contributions

This work has resulted in several contributions to the fields of instructional design and development, human performance engineering, and CBA. First, a cross-domain literature search has combined information from psychology, education, training, systems and aerospace engineering, mathematics, business, and economics to create a comprehensive list of criteria, which should be used to predict the effectiveness of training systems design. A modeling and simulation approach was used in the development of a methodology, MPEET, for predicting training effectiveness during the early phase of the defense acquisition process using the list of criteria. MPEET leverages concepts and tools from systems engineering and uses them to communicate relationships between instructional design elements, predict characteristics, and support decision-making. Tools such as the compatibility matrices and morphological matrices were used to map instructional methods to learning objectives and create cost-effective training alternatives. The MPEET process aligns with the generic IPPD decision-making process making the steps easy to follow and results in not only an evaluation but also a recommendation of a more cost-effective training program. MPEET enhances the instructional design process with the objective verification of the training system design using modeling and simulation. It predicts with probabilistic uncertainty the cost-effectiveness of the training program. MPEET was applied to a C-130J pilot training program. The criteria used in MPEET stems from primary instructional design elements and business and economic cost analysis methods that enable the prediction of training effectiveness. The ability to predict training effectiveness and quantify the variable uncertainty contributes to the JCIDS CBA process. Adding MPEET as a reference for evaluating the training alternative of DOTMLPF-P fills an informational gap in CBA process guidance.

6.2 Recommendations for Future Research Areas

Cost-utility analysis was used in MPEET to determine the best training alternatives. This technique works well when comparing and analyzing the most effective training method. In order to compare training as an alternative to other DOTMLPF-P alternatives a cost-benefit analysis is necessary. The benefits gained from training will differ from those obtained from a doctrinal, organization, or facilities change. Each of the DOTMLPF-P alternatives will need to be assessed in terms of their overall worth. To accomplish this task, one must determine the appropriate benefits of training (i.e. reduced loss of life, increased mission success probability) that should be translated into monetary values, and how to translate those benefits into outcome measures that are easily expressed in units of currency. MPEET could then be updated to use cost-benefit analysis in addition to cost-utility. Once the benefits of training are converted to monetary values, the next step is to compare training as an alternative to other non-materiel solutions. This research effort discusses but does not attempt an experiment or demonstration of a CBA analysis to compare training to the other DOTMLPF-P alternatives.

The aim of this research was to provide a proof of concept for the overall MPEET process. In the future, to adequately assess all aspects of training effectiveness, the addition of unique student or trainee characteristics and learning style preferences that are not considered in this effort will enhance MPEET and provide a more robust measure of all aspects of training effectiveness. The only learner component of instructional design included in MPEET is the impact of age on the effectiveness of training methods. As previously discussed in section 2.7.1, there are many other variables that influence training effectiveness that require pre- and post-training comparisons. Although these variables are not included during this initial development of MPEET, their existence is acknowledged and their inclusion may result in more accurate predictability.

REFERENCES

1. Secretary of Defense, *Quadrennial Defense Report*, 2010: Washington, DC.
2. The Secretary of Defense, *Sustaining U.S. Global Leadership Priorities for 21st Century Defense*, 2012: Washington, DC.
3. Bumliller, E. and T. Shanker, *Obama Puts His Stamp on Strategy for a Leaner Military*, in *The New York Times*2012.
4. Dempsey, M.E., *Capstone Concept for Joint Operations: Joint Force 2020*, 10 September 2012.
5. Alexander, D. and A. Shalal *Budget cuts to slash U.S. Army to smallest since before World War Two*. 2014.
6. Williams, B., *Story 3*, in *NBC Nightly News*2014: New York.
7. Chairman of the Joint Chiefs of Staff Instruction, *Joint Capabilities Integration and Develpment System 3170.01H*, 10 January 2012.
8. Joint Chiefs of Staff, J-8, *White Paper on Conducting a Capabilities-Based Assessment (CBA) Under the Joint Capabilities Integration and Development System (JCIDS)*, January 2006.
9. Schwartz, M., *Defense Acquisitions: How DOD Acquires Weapon Systems and Recent Efforts to Reform the Process*, 2013, Congressional Research Service.
10. Chairman of the Joint Chiefs of Staff, *Manual for the Operation of the Joint Capabilities Integration and Development System*, 2012.
11. Joint Chiefs of Staff, *Joint Staff Organizational Chart*, U.S. Department of Defense, Editor 2012.
12. Brown, B., *Introduction to Defense Acquisition Management*. 10th ed. 2010, Fort Belvoir, VA: Defense Acquisition University.
13. Griendling, K.A., *ARCHITECT: The architecture-based technology evaluation and capability tradeoff method*, 2011, Georgia Institute of Technology: United States -- Georgia. p. 403.
14. Department of Defense, *Interim DoD Instruction 5000.02, "Operation of the Defense Acquisition System"*, D.S.o. Defense, Editor 2013: Washington, D.C.
15. Chairman of the Joint Chiefs of Staff, *Capabilities Based Assessment Guide.* 2012. **Appendix B to Enclosure A**.

16. Joint Chiefs of Staff, J.-. *Capabilities-Based Assessment (CBA) User's Guide Version 3*, March 2009.

17. TRADOC, *Capabilities-Based Assessment (CBA) Guide, Version 3.1*, 2010.

18. Simpson, H., *Cost Effectiveness Analysis of Training in the Department of Defense*, 1995, Defense Manpower Data Center: Seaside, CA.

19. Department of Defense, *Military Handbook Interactive Courseware (ICW) for Military Training, Manager's Guide for Development, Acquisition, and Management of (Part 1 of 3 Parts)*, Department of the Navy, Editor 1992: Washington, DC.

20. Department of Defense, *Department of Defense Handbook Guidance for Acquisition of Training Data Products and Services*, Naval Air Systems Command, Editor 1996: Lakehurst, NJ.

21. United States Department of the Air Force, *Air Force Handbook 36-2235. Information for Designers of Instructional Systems Application to Aircrew Training*, 1993, Headquaters US Air Force,: Washington, D.C.

22. United States Department of Defense, *Department of Defense Handbook Guidance for Acquisition of Training Data Produces and Services (Part 1 of 4 Parts)*, Naval Air Warfare Center Aircraft Division, Editor 1999: Lakehurst, NJ.

23. United States Government Accountability Office, *Military Training: Navy and Air Force Need to More Fully Apply Best Practices to Enhance Development and Management of Combat Skills Training*, 2009: Washington, DC.

24. United States Goverment Accountability Office, *Ballistic Missile Defense Actions Needed to Improve Training Integration and Increase Transparancy of Training Resources*, 2011.

25. Oxford University Press, *Oxford Dictionaries Online*, 2013.

26. Random House, Inc, *Dictionary.com Unabridged*, Dictionary.com website.

27. Joint Chiefs of Staff, J-8, *Joint Capability Areas Definitions and Framework*, J7/Joint Concepts to Capabilities Division (JCCD), Editor 8 April 2011.

28. Defense Acquisition University, *Glossary of Defense Acquisition Acronyms and Terms*, July 2011.

29. Alvarez, K., E. Salas, and C. Garofano, *An Integrated Model of Training Evaluation and Effectiveness.* Human Resource Development Review, 2004. **3**(4): p. 385-416.

30. Kirkpatrick, D.L. and J.D. Kirkpatrick, *Evaluating Training Programs : The Four Levels*. 3rd ed. 2006, San Francisco, CA: Berrett-Koehler.

31. Hewlett-Packard, *Summary Process for measuring ROI of hp*, 2004, Hewlett-Packard Development Company, L.P.

32. Jung, C. *Evaluation in Instructional Design - Kirkpatrick's 4 Level Model*. 2009 1 March 2009 12 February 2013]; Available from: http://etec.ctlt.ubc.ca/510wiki/Evaluation_in_Instructional_Design_-_Kirkpatrick%27s_4_Level_Model.

33. Ascend Corporate Learning. *Kirkpatrick's Four Levels of Learning*. 2010 [cited 2013 16 April]; Available from: http://www.acend.com/learning-methods/kirkpatricks-four-levels-of-learning-/.

34. Sugrue, B. and R.J. Rivera, *State of the Industry: ASTD's Annual Review of Trends in Workplace Learning and Performance*, 2005, American Society of Training and Development: Alexandria, VA.

35. Aguinis, H. and K. Kraiger, *Benefits of training and development for individuals and teams, organizations, and society.* Annual Review Of Psychology, 2009. **60**: p. 451-474.

36. Twitchell, S., E.F. Holton III, and J.W. Trott Jr, *Technical Training Evaluation Practices in the United States.* Perform Improvement Quarterly, 2000. **13**(1): p. 84-109.

37. Tourville, S., *Lockheed Martin Training Solutions*, Lockheed Martin Global Training and Logistics: Orlando, FL.

38. American Management Association, *Effective Training: Implementing the Four Levels*, 2007.

39. Kirkpatrick Partners, LLC. *Training Evaluation That Delivers Measurable Business Results*. 2013 [cited 2013 17 April]; Available from: http://www.kirkpatrickpartners.com/Home/tabid/38/Default.aspx.

40. Kirkpatrick, D.L., *Evaluating Training Programs : The Four Levels*. 1998, San Francisco, Calif: Health Press.

41. Praslova, L., *Adaptation of Kirkpatrick's four level model of training criteria to assessment of learning outcomes and program evaluation in higher education.* Educational Assessment, Evaluation and Accountability, 2010. **22**(3): p. 215-225.

42. Patel, L., *ASTD State of the Industry Report 2010*, 2010, American Society for Training & Development: Alexandria, VA.

43. Clark, D.R. *Kirkpatrick's Four Level Evaluation Model*. 2012 21 February 2013; Available from: http://www.nwlink.com/~donclark/hrd/isd/kirkpatrick.html.

44. Kirkpatrick, J.D. and W.K. Kirkpatrick, *Three Steps to Effectiveness*, in *Chief Learning Officer Magazine* 2010, Mediatec Publishing Inc.

45. Clark, D.R. *Flipping Kirkpatrick*. 2008 17 December 2008 22 April 2013]; Available from: http://bdld.blogspot.com/2008/12/flipping-kirkpatrick.html - links.

46. DuBose, E. *Kirkpatrick does a 180: That's more like it*. 2011 8 February 2011 22 April 2013]; Available from: http://ipttoolkit.wordpress.com/2011/02/08/kirkpatrick-does-a-180-thats-more-like-it/.

47. Gram, T. *Evaluating Training and Learning Circa 2011*. 2011 17 February 2011 22 April 2013]; Available from: http://performancexdesign.wordpress.com/2011/02/17/evaluating-training-and-learning-circa-2011/.

48. Wick, C.W., R.V.H. Pollock, and A. Jefferson, *The Six Disciplines of Breakthrough Learning*. 2006, San Francisco, CA: Pfeiffer.

49. Rudman, R., *Performance Planning & Review*. 1998, Warriewood, Australia: Business & Professional Publishing.

50. Gilbert, T., *A Leisurely Look at Worthy Performance*, in *The 1998 ASTD Training and Performance Yearbook*, J.A. Woods and J.W. Cortada, Editors. 1998, McGraw-Hill: New York.

51. Alliger, G.M., et al., *A meta-analysis of the relations among training criteria.* Personnel Psychology, 1997. **50**(2): p. 341-358.

52. Kirkpatrick, D.L. *The four-level model*. 1960 22 April 2013]; Available from: http://idassessment.wikispaces.com/Kirkpatricks+four+levels+of+evaluation.

53. Deitchman, S.J., *Preliminary Exploration of the Use of a Warfare Simulation Model to Examine the Military Value of Training*, 1988, Institute for Defense Analyses: ALEXANDRIA VA.

54. Deitchman, S.J., *Further Explorations in Estimating the Military Value of Training*, 1990, Institute for Defense Analyses: Alexandria, VA.

55. Goman, G.P.F., (USA, Ret.), *The Military Value of Training*, 1990, Institute for Defense Analyses: Alexandria, VA.

56. Deitchman, S.J., *Quantifying the Military Value of Training for System and Force Acquisition Decisions: An Appreciation of the State of the Art*, 1993, Institute For Defense Analyses: Alexandria, VA.

57. Bell, H.H. and W.L. Waag, *Evaluating the effectiveness of flight simulators for training combat skills: A review.* International Journal of Aviation Psychology, 1998. **8**(3): p. 223-242.

58. Schreiber, B.T., M. Schroeder, and W. Bennett, *Distributed Mission Operations Within-Simulator Training Effectiveness.* International Journal of Aviation Psychology, 2011. **21**(3): p. 254-268.

59. Gehr, S.E., B. Schreiber, and W. Bennett, *Within-Simulator Training Effectiveness Evaluation*, in *The Interservice/Industry Training, Simulation & Education Conference* 2004, National Security Industrial Association: Orlando, FL. p. 1652-1661.

60. Schreiber, B.T., M. Schroeder, and W. Bennett, *Distributed Mission Operations Within-Simulator Training Effectiveness Baseline Study: Metric Development and Objectively Quantifying the Degree of Learning*, Air Force Research Laboratory, Warfighter Readiness Research Division,, Editor 2006.

61. Bahlis, J. and S. Tourville, *Where Training Resources Should be Allocated*, in *The Interservice/Industry Training, Simulation & Education Conference (I/ITSEC)*2005.

62. Phillips, J.J., *Return on Investment in Training and Performance Improvement Programs*. 2nd ed. 2003: Butterworth-Heinemann.

63. Morrison, G.R., et al., *Designing Effective Instruction*. 7th ed. 2013, Hoboken, NJ: John Wiley & Sons, Inc.

64. Cornell-d'Echert, B., *Beyond training: New ideas for military forces operating beyond war.* New Directions for Adult and Continuing Education, 2012. **2012**(136): p. 17-27.

65. Salas, E., et al., *The Science of Training and Development in Organizations: What Matters in Practice.* Psychological Science in the Public Interest, 2012. **13**(2): p. 74-101.

66. Campbell, J.P., *Personnel Training and Development.* Annual Review of Psychology, 1971. **22**: p. 565-602.

67. Kraiger, K., J.K. Ford, and E. Salas, *Application of Cognitive, Skill-Based, and Affective Theories of Learning Outcomes to New Methods of Training Evaluation.* Journal of Applied Psychology, 1993. **78**(2): p. 311-328.

68. Salas, E. and J.A. Cannon-Bowers, *The Science of Training: A Decade of Progress.* Annual Review of Psychology, 2001. **52**: p. 471-499.

69. Hilgard, E.R., *Theories of learning*. 4th ed. ed. The Century psychology series, ed. G.H. Bower. 1975, Prentice-Hall: Englewood Cliffs, N.J.

70. Mowrer, R.R. and S.B. Klein, *Handbook of Contemporary Learning Theories*. 2001, Mahwah, NJ: L. Erbaum Associates.

71. Ormrod, J.E., *Human Learning*. 5th ed. 2008, Upper Saddle River, NJ: Pearson/Merrill Prentice Hall.

72. Bloom, B.S., *Taxonomy of Educational Objectives: The Classification of Education Goals. Handbook I: Cognitive Domain*. 1956, New York: Longmans, Green and Company.

73. Eisner, E.W., *Benjamin Bloom, 1913-99.* Prospects: The Quarterly Review of Comparative Education, 2000. **XXX, no. 3**(115): p. 124.

74. Anderson, L.W. and L.A. Sosniak, *Bloom's taxonomy: A forty-year retrospective*. National Society for the Study of Education Yearbooks. 1994, Chicago, IL: University of Chicago Press.

75. Lewy, A. and Z. Bathory, *The taxonomy of educational objectives in Continental Europe, the Mediterranean, and the Middle East*, in *Bloom's taxonomy: A forty-year retrospective*. 1994, University of Chicago Press: Chicago, IL.

76. Anderson, L.W. and D.R. Krathwohl, *A taxonomy for learning, teaching, and assessing : a revision of Bloom's taxonomy of educational objectives: Complete edition*, ed. B.S. Bloom. 2001, Longman: New York.

77. Driscoll, M.P., *Psychology of Learning for Instruction*. 2nd ed. 2000, Needham Heights, MA: Allyn & Bacon.

78. Iskandar, Y.H.P., L. Gilbert, and G.B. Wills. *Pedagogy in Computer-Based Sport Training*. in *Advanced Learning Technologies (ICALT), 2011 11th IEEE International Conference*. 2011. Athens, GA.

79. Smith, P.L. and T.J. Ragan, *Instructional design*. 1999: Merrill Upper Saddle River, New Jersey.

80. Driscoll, M.P., *Psychology of Learning for Instruction*. 1994, Needham Heights, MA: Allyn & Bacon.

81. Hassan, O.A.B., *Learning theories and assessment methodologies – an engineering educational perspective.* European Journal of Engineering Education, 2011. **36**(4): p. 327-339.

82. Kyllonen, P.C. and V.J. Shute, *Taxonomy of Learning Skills*, 1988, Air Force Human Resource Laboratory: Brooks Air Force Base, TX.

83. Pope, C.C., *Once More with Feeling: Affect and Playing with the TGfU Model.* Physical Education and Sport Pedagogy, 2005. **10**(3): p. 271-286.

84. Shulman, L.S., *Making Differences: A Table of Learning.* Change, 2002. **34**(6): p. 36-44.

85. Ertmer, P.A. and T.J. Newby, *Behaviorism, cognitivism, constructivism: Comparing critical features from an instructional design perspective.* Performance Improvement Quarterly, 2013. **26**(2): p. 43-71.

86. Winn, W., *Some implications of cognitive theory for instructional design.* Instructional Science, 1990. **19**(1): p. 53-69.

87. Rush, L., *Archaeology, Cultural Property, and the Military*. Heritage Matters. 2010, Rochester, NY: Boydell Press.

88. Svinicki, M.D., *New Directions in Learning and Motivation.* New Directions for Teaching & Learning, 1999(80): p. 5.

89. Jackson, L.D., *Revisiting Adult Learning Theory through the Lens of an Adult Learner.* Adult Learning, 2009. **20**(3/4): p. 20-22.

90. Skinner, B.F., *Beyond Freedom and Dignity [by] B. F. Skinner*. 1971, New York: Knopf.

91. Gagne, R.M., *The Conditions of Learning*. 4 ed. 1985, New York: Holt, Rinehart & Winston.

92. Knirk, F.G. and K.L. Gustafson, *Instructional Technology: A Systematic Approach to Educaton*. 1986, New York.

93. Dave, R.H., *Psychomotor Levels*, in *Developing and Writing Behavioral Objectives*, R.J. Armstrong, Editor. 1970, Educational Innovators Press: Tucson, AZ.

94. Simpson, E.J., *The Classification of Educational Objectives in the Psychomotor Domain: The Psychomotor Domain*. 1972, Washington, DC: Gryphon House.

95. Harrow, A.J., *A Taxonomy of the Psychomotor Domain: A Guide for Developing Behavioral Objectives*. 1972, New York: David McKay Co.

96. Ferris, T.L.J., *Bloom's Taxonomy of Educational Objectives: A Psychomotor Skills Extension for Engineering and Science Education.* International Journal of Engineering Education, 2010. **26**(3): p. 699-707.

97. Bloom, B.S., *Learning for Mastery.* Evaluation Comment, 1968. **1**(2): p. 12.

98. Cregger, R. and M. Metzler, *PSI for a College Physical Education Basic Instructional Program.* Educational Technology, 1992. **32**(8): p. 51-56.

99. Draper, S. *Taxonomies of learnign aims and objectives: Bloom, neoBloom, and crticisms*. 2005 Feb 15, 2013 [cited 2014 Feb 2]; Available from: http://www.psy.gla.ac.uk/~steve/best/bloom.html.

100. Utley, R., *Theory and Research for Academic Nurse Educators: Application to Practice*. 2010, Burlington, MA: Jones and Bartlett Learning.

101. Merritt, R.D., *The Psychomotor Domain.* EBSCO Research Starters, 2008.

102. Clark, D.R. *Bloom's Taxonomy of Learning Domains*. 1999 1 May 2013 6 January 2014]; Available from: http://www.nwlink.com/~donclark/hrd/bloom.html.

103. United States Department of Defense, *Department of Defense Handbook Instructional Systems Development/Systems Approach to Training and Education (Part 2 of 5 Parts)*, Naval Air Warfare Center Aircraft Division, Editor 2001: Lakehurst, NJ.

104. Engineers Australia, *Engineers Australia National Generic Competency Standards—Stage 1 Competency Standard for Professional Engineers*, 2006: Canberra.

105. Engineers Australia, *Engineers Australia National Generic Competency Standards—Stage 1 Competency Standard for Engineering Technologist*, 2007, Engineers Australia: Canberra.

106. ABET, *Accreditation Policy and Procedure Manual Effective for Evlauations During 2009-2010 Accreditation Cycle*, 2008: Baltimore, MD.

107. Yilmaz, K., *The Cognitive Perspective on Learning: Its Theoretical Underpinnings and Implications for Classroom Practices.* Clearing House, 2011. **84**(5): p. 204-212.

108. Winn, W. and D. Snyder, *Cognitive Perspectives in Pshychologoy*, in *Handbook of research for educational communications and technology : a project of the Association for Educational Communications and Technology*, D.H. Jonassen, Editor. 1996, New York: Macmillan Library Reference USA.

109. Grippin, P. and S. Peters, *Learning Theory And Learning Outcomes: The Connection*. 1984, Lanham, MD: University Press of America.

110. Anderson, J.R., L.M. Reder, and H.A. Simon, *Rejoinder: Situative versus Cognitive Perspectives: Form versus Substance.* Educational Researcher, 1997. **26**(1): p. 18-21.

111. Greeno, J.G., A.M. Collins, and L. Resnick, *Cognition and Learning*, in *Handbook of Educational Psychology*, D.C. Berliner and R.C. Calfee, Editors. 1996, Macmillan: New York.

112. Collins, A., J.S. Brown, and S.E. Newman, *Cognitive Apprenticeship: Teaching the Craft of Reading, Writing, and Mathematics*, 1987, Bolt, Beranek and Newman, Illinois Univ., Urbana Center for the Study of Reading: Arlington, VA. p. 37.

113. Brown, A.L., *Knowing When, Where, and How to Remember: A Problem of Metacognition*, 1977, Illinois Univ., Urbana Center for the Study of Reading, Bolt, Beranek and Newman, Inc.: Washington, DC. p. 152.

114. Bloom, B.S., et al., *Taxonomy of Educational Objectives The Classification of Educational Goals Handbook I: Cognitive Domain*. 1979, London: Longman Group.

115. Athanassiou, N., J.M. McNett, and C. Harvey, *Critical Thinking in the Management Classroom: Bloom's Taxonomy as a Learning Tool.* Journal of Management Education, 2003. **27**(5): p. 533-555.

116. Putnam, R.T. and H. Borko, *What Do New Views of Knowledge and Thinking Have to Say about Research on Teacher Learning?* American Educational Research Association, 2000. **29**(1): p. 4-15.

117. Liu, M., et al., *Understanding the Connection Between Cognitive Tool Use and Cognitive Processes as Used by Sixth Graders in a Problem-Based Hypermedia Learning Environment.* Journal of Educational Computing Research, 2004. **31**(3): p. 309-334.

118. Wilson, L.O. *Beyond Bloom - A new Version of the Cognitive Taxonomy*. 2006 January 9, 2014]; Changes to Bloom's Figure]. Available from: http://www4.uwsp.edu/education/lwilson/curric/newtaxonomy.htm.

119. Shank, P. *eLearning Guild Research: Reconsidering Bloom's Taxonomy (Old AND New)*. Learning Solutions Magazine, 2013.

120. Feezel, J.D., *Toward a Confluent Taxonomy of Cognitive, Affective, and Psychomotor Abilities in Communication* Communication Education, 1985. **34**(1).

121. Krathwohl, D.R., *A Revision of Bloom's Taxonomy: An Overview.* Theory into Practice, 2002. **41**(4): p. 212-218.

122. Gagne, R., M. and M.P. Driscoll, *Essentials of Learning for Instruction*. 1988, Englewood Cliffs, NJ: Prentice-Hall.

123. Merrill, M.D., *Component Display Theory*, in *Instructional Design Theories and Models*, C.M. Reigeluth, Editor. 1983, Lawrence Erlbaum Associates: Hillsdale, NJ.

124. Merrill, M.D., *First Principles of Instruction.* Educational Technology Research and Development, 2002. **50**(3): p. 43-59.

125. Biggs, J.B. and K.F. Collins, *Evaluating the Quality of Learning - The SOLO Taxonomy*. 1982, New York: Academic Press.

126. Reigeluth, C.M. and J. Moore, *Cognitive Education and the Cognitive Domain*. Instructional-Design Theories and Models: A New Paradigm of Instructional Theory, ed. C.M. Reigeluth. Vol. II. 1999, Hillsdale, NJ: Lawrence Erlbaum Associates.

127. Krathwohl, D.R., B.S. Bloom, and B.B. Masia, *Taxonomy of Educational Objectives: The Classification of Educational Goals*. Handbook II: Affective Domain. Vol. 1. 1964, New York: David McKay Company, Inc.

128. Colwell, R., *MENC Handbook of Research Methodologies*. 2006, New York: Oxford University Press.

129. Hauenstein, D.A., *A Conceptual Framework for Educational Objectives: A Holistic Approach to Traditional Taxonomies*. 1998, Lanham, MD: University Press of America.

130. Skryabina, E. and N. Reid, *Attitude Towards Physics.* Research in Science & Technological Education, 2002a. **20**(1): p. 67-81.

131. Reid, N., *A Scientific Approach to the Teaching of Chemistry.* The Royal Society of Chemistry, 2008. **9**: p. 51-59.

132. Wickens, C.D. and J.G. Hollands, *Engineering Psychology and Human Performance*. null, ed. C.D. Wickens and J.G. Hollands. 2000, Upper Saddle River, NJ: Prentice Hall.

133. Spiker, A., et al., *Team Performance during Combat Mission Training: A Conceptual Model and Measurement Framework*, 1996, Armstrong Laboratory Human Resources Directorate Aircrew Training Research Division: Mesa, AZ.

134. Orasanu, J.M. and U. Fischer, *Finding Decisions in Hatural Environments: The View from the Cockpit*, in *2nd Naturalistic Decision Making Conference*, C.E. Zsambok and G.A. Klein, Editors. 1997, Lawrence Erlbaum Association: Dayton, OH. p. 343-357.

135. Predmore, S.C. *Microcoding of Cockpit Communications in Accident Analyses: Crew Coordination of the United 811 and United 232.* in *Sixth International*

Symposium on Aviation Psychology. 1991. Columbus, Ohio: The Ohio State University, Department of Aviation.

136. United States Department of Defense, *Department of Defesnse Handbook Instructional Systems Development/Systems Approach to Training and Education (Part 2 of 5 Parts)*, Naval Air Warfare Center Aircraft Division, Editor 2001.

137. Vosniadou, S. and W.F. Brewer, *Mental models of the earth: A study of conceptual change in childhood.* Cognitive Psychology, 1992. **24**(4): p. 535-585.

138. Brown, J.S., A. Collins, and P. Duguid, *Situated Cognition and the Culture of Learning.* Educational Researcher, 1989. **18**(1): p. 32-42.

139. Tomei, L.A., *Designing Instruction for the Traditional, Adult, and Distance Learner : A New Engine for Technology-based Teaching*. 2010, Hershey, PA: Information Science Reference.

140. Entwistle, N.J., *Styles of Learning and Teaching: Integrated Outline of Educational Psychology for Students, Teachers and Lecturers*. 1981, Chichester, England: Wiley.

141. Witkin, H.A. and D.R. Goodenough, *Cognitive Styles: Essence and Origins*. 1981, New York: International Universities Press.

142. Marzano, R.J., *A Different Kind of Classroom: Teaching with Dimensions of Learning*. 1992, Alexandria, VA: Association for Supervision and Curriculum Development. 213.

143. Schmeck, R.R., *Learning Strategies and Learning Styles*. Perspectives on Individual Differences. 1988, New York: Plenum Press.

144. Weinstein, C.E. and R.E. Mayer, *The Teaching of Learning Strategies.* Innovation Abstracts, 1983. **5**(32).

145. Baldwin, T.T. and J.K. Ford, *TRANSFER OF TRAINING: A REVIEW AND DIRECTIONS FOR FUTURE RESEARCH.* Personnel Psychology, 1988. **41**(1): p. 63-105.

146. Ford, J.K. and K. Kraiger, *The Application of Cognitive Constructs and Principles to the Instructional Systems Model of Training: Implications for Needs Assessment, Design, and Transfer.* INTERNATIONAL REVIEW OF INDUSTRIAL AND ORGANIZATIONAL PSYCHOLOGY, 1995. **10**: p. 1-48.

147. Clark, D.R. *Learning Strategies or Instructional Strategies*. 1999 July 5, 2010 5 February 2014]; Available from: http://www.nwlink.com/~donclark/hrd/strategy.html.

148. NTL Institute for Applied Behavioral Science, *The Learning Pyramid*: 300 N. Lee Street, Suite 300, Alexandria, VA 22314. 1-800-777-5227.

149. Sitzmann, T., et al., *A multilevel analysis of the effect of prompting self-regulation in technology-delivered instruction.* Personnel Psychology, 2009. **62**(4): p. 697-734.

150. Gagne, R.M. and M.D. Merrill, *Integrative Goals for Instructional Design.* Educational Technology and Research and Development, 1992. **38**: p. 23-30.

151. Tennyson, R.D. and K. Breuer, *Psychological foundations for instructional design*. Instructional design: International perspectives: Vol. 1. Theory, research, and models, ed. R.D. Tennyson, et al. 1997, Mahwah, NJ: Erlbaum. 113-134.

152. Newman, F. and J. Scurry, *Online technology pushes pedagogy to the forefront.* The Chronicle of Higher Education, 2001. **47**: p. B7-B10.

153. Sitzmann, T., et al., *The comparative effectiveness of web-based and classroom instruction: A meta-analysis.* Personnel Psychology, 2006. **59**(3): p. 623-664.

154. Clark, R.E., *Media will never influence learning.* Educational Technology Research and Development, 1994. **42**: p. 21-29.

155. Dumont, R.A., *Teaching and Learning in Cyberspace.* IEEE Transactions on Education, 1996. **39**: p. 192-204.

156. Hiltz, S.R. and B. Wellman, *Asynchronous learning networks in a virutal classroom.* Communications of the ACM, 1997. **40**: p. 44-49.

157. Sullivan, P., *Gender differences and the on-line classroom. Male and female college students evaluate their differences.* Communiy College Journal of Research and Practice, 2001. **25**: p. 805-818.

158. Goldstein, I.L. and J.K. Ford, *Training in organizations: Needs assessment, development and evaluation*. 4th ed. 2002, Belmont, CA: Wadsworth.

159. Welsh, L.T., et al., *E-learning: Emerging uses, best practices, and future directions.* International Journal of Training and Development, 2003. **7**: p. 245-258.

160. Zhao, Y., et al., *What makes the difference? A practical analysis of research on the effectiveness of distance education.* Teachers College Record, 2005. **107**: p. 1836-1884.

161. Fletcher, J.D., *Effectiveness of cost of interactive videodisc instruction in defense training and education.* Institute for Defense Analyses Paper, 1990. **P-2372**.

162. Kulik, C.-L.C. and J.A. Kulik, *Effectiveness of Computer-Based Instruction: An Updated Analysis.* Computers in Human Behavior, 1991. **7**: p. 75-94.

163. Kulik, J.A., *Meta-analytic studies of findings on computer-based instruction*, in *Technology Assessment in Education and Training*, E.L. Baker and H.F. O'Neil Jr., Editors. 1994, Erlbaum: Hillsdale, NJ.

164. Yaakub, M.N., *Meta-Analysis of the Effectiveness of Computer-Assisted Instruction in Technical Education and Training*, 1998, Virgina Polytechnic Insitute and State University: Blacksburg.

165. Liao, Y.-K.C., *Effects of Hypermedia on Students' Achievement: A Meta-Analysis.* Journal of Educational Multimedia and Hypermedia, 1999. **8**: p. 255-277.

166. Olson, T.M. and R.A. Wisher, *The Effectiveness of Web-Based Instruction: An Inital Inquiry.* International Review of Research in Open and Distance Learning, 2002. **3**(2).

167. Means, B., et al., *Evaluation of Evidence-Based Practices in Online Learning: A Meta-Analysis and Review of Online Learning Studies*, 2010, U.S. Department of Education Office of Planning, Evaluation, and Policy Development Policy: Washington, DC.

168. Knowles, M.S., *The Adult Learner: A Neglected Species*. 3rd ed. 1984, Houston, TX: Gulf Publishing Company.

169. White, R.D., *On Pedagogy and Andragogy: Balancing the Learning Needs of Pre-Service and In-Service MPA Students.* Journal of Public Affairs Education, 2000. **6**(2): p. 69-78.

170. Durant, R.F. and W.A. Taggart, *Mid-Career Students in MPA Programs: Implications for Pre-Service Student Education.* Public Administration Review, 1985: p. 301-308.

171. Knowles, M., E.F. Holton, and R.A. Swanson, *The Adult Learner: The Definitive Classic in Adult Education and Human Resource Development*. 7th ed. 2011, Burlington, MA: Taylor & Francis.

172. Graham, B.W., *Attitude, Motivation, and Anxiety in Traditional and Nontraditional Students at Institutes of Higher Education*, 1990, The University of New Mexico: Albuquerque.

173. Tallent-Runnels, M.K., et al., *Teaching Courses Online: A Review of the Research.* Review of Educational Research, 2006. **76**: p. 93-135.

174. Hylmo, A. and L. Speckman, *Knowledge Transfer and Mentoring the Next Generation of Aerospace Engineers: A Case Study for Strategic Intervention*, in

*AIAA SPACE 2010 Conference & Exposition*2010, American Institute of Aeronautics and Astronautics: Anaheim, CA.

175. McGlynn, A.P., *Teaching Millenials, Our Newest Cultural Cohort.* Education Digest, 2005. **71**(4): p. 12-16.

176. Townsend, J., A. Dillon, and S. Sequeira, *Student-Centered Learning and Metacognitive Strategies in Engineering Thermal-Fluids Courses*, in *46th AIAA Aerospace Sciences Meeting and Exhibit*2008, American Institute of Aeronautics and Astronautics: Reno, NV.

177. Barnes, K., R.C. Marateo, and P.S. Ferris, *Teaching and Learning with the Net Generation.* Innovate, 2007. **3**(4).

178. Jentsch, F. and J.A. Cannon-Bowers, *Evidence for the Validity of PC-Based Simulations in Studying Aircrew Coordination.* International Journal of Aviation Psychology, 1998. **8**: p. 243-260.

179. Magennis, S. and A. Farrell, *Teaching and Learning Activities: Expanding the Repertoire to Support Student Learning.* All Ireland Society for Higher Education, 2005. **1**.

180. Edgar, D., *A Truncated Section of the Cone of Experience.* Theory into Practice, 1970. **9**(2): p. 96-100.

181. Lalley, J.P. and R.H. Miller, *The Learning Pyramid: Does it Point Teachers in the Right Direction?* Education, 2007. **128**(1): p. 64-79.

182. Buckley, R. and J. Caple, *The Theory & Practice of Training*. 6th ed. Theory and Practice of Training. 2009, London: Kogan Page.

183. Cascio, W.F. and R.A. Ramos, *Development and Application of a New Method for Assessing Job-Performance in Behavioral Economic Terms.* Journal of Applied Psychology, 1986. **71**(1): p. 20-28.

184. Curren, R., *Connected Learning and the Foundations of Psychometrics: A Rejoinder.* Journal of Philosophy of Education, 2006. **40**(1): p. 17-29.

185. Ghiselli, E.E., J.P. Campbell, and S. Zedeck, *Measurement Theory for the Behavioral Sciences*. 1981, San Franciso: W. H. Freeman & Company.

186. Mattson, B.W., *Using the Critical Outcome Technique to Demonstrate Financial and Organizational Performance Results.* Advances in Developing Human Resources, 2005. **7**: p. 102-120.

187. Canada, J.R., W.G. Sullivan, and J.A. White, *Capital Investment Analysis for Engineering Management*. 2nd ed. 1996, Upper Saddle River, NJ: Prentice Hall.

188. Levin, H.M. and P.J. McEwan, *Cost-Effectiveness Analysis: Methods and Applications*. 2nd ed. 2001, Thousand Oaks, CA: Sage Publications. 308.

189. Domercant, J.C., *ARC-VM: An Architecture Real Options Complexity-Based Valuation Methodology for Military System-of-Systems Acquisitions*, in *Aerospace Engineering*2011, Georgia Institute of Technology: Atlanta, GA.

190. Ackerman, F. *Critque of Cost-Benefi Analysis, and Alternative Approaches to Decision-Making: A report to Friends of the Earth England, Wales and Northern Ireland*. 2008.

191. Defense Acquisition Guidebook, *Online Publication*, 2013.

192. Office of Aerospce Studies, A.F.M.C.A.O.A., *Office of Aersopace Studies: Analysis of Alternatives (AoA) Handbook*, 2013: Kirtland AFB, NM.

193. United States Goverment Accountability Office, *Cost Assessment Guide: Best Practices for Estimating and Managing Program Costs (Exposure Draft)*, 2007.

194. McCabe, T.J. and C.W. Butler, *Design Complexity Measurement and Testing.* Communications of the ACM, 1989. **32**: p. 1415-1425.

195. Mavris, D.N., *Introduction to Decision Making and Mult-attribute Decision Making (MADM)*, 2011, Georgia Institute of Technology, School of Aerospace Engineering, Advanced Systems Design Laboratory: Atlanta, GA.

196. Keeney, R.L. and H. Raiffa, *Decisions with Multiple Objectives: Preferences and Value Tradeoffs*. 1976, New York, NY: John Wiley & Sons, Inc.

197. Keeney, R.L. and H. Raiffa, *Decisions with Multiple Objectives: Preferences and Value Tradeoffs*. 1993, Cambridge, UK: Cambridge University Press.

198. Hazelrigg, G.A., *Systems Engineering: An Approach to Information-Based Design*. Prentice-Hall International Series in Industrial and Systems Engineering. 1996: Prentice Hall, Inc.

199. Clemen, R.T., *Making Hard Decisions: An Introduction to Decision Analysis*. 2nd ed. 1996, Pacific Grove, CA: Duxbury Press.

200. INCOSE, *Systems Engineering Handbook: A Guide for System Life Cycle Processes and Activities*, C. Haskins, Editor 2006, International Council of Systems Engineering.

201. Idaho National Laboratory. *System Modeling and Simulation*. [cited 2014 22 May]; Available from: https://inlportal.inl.gov/portal/server.pt/community/systems_engineering/334/system_modeling/1861.

202. Fallin, J.H.K., *Modeling and Simulation (M&S) Use in the Army Acquisition Process*, in *Program Management*1997, Defense Acquisition University. p. 78-80.

203. National Research Council, *Statistics, Testing, and Defense Acquisition: New Approaches and Methodological Improvements*, ed. M.L. Cohen, J.E. Rolph, and D.L. Steffey. 1998, Washington, DC: The National Academies Press.

204. Thomas, J.J. and K.A. Cook, *Illuminating the Path: The Research and Development Agenda for Visual Analytics*, 2005, National Visualization and Analytics Center.

205. Schrage, D.P. and D.N. Mavris, *Recomposistion: The Other Half of the MDO Equation*, in *ICASE/NASA Langley Workshop in Multi-disciplinary Design Optimization - State-of-the-Art*1995.

206. U.S.A. Department of Defense, *DoD Guide to Integrated Product and Process Development (Version 1.0)*, 1996, Office of the Under Secretary of Defense (Acquisition and Technology): Washington, DLC.

207. Schrage, D.P. and D. Mavris, *Technology for affordability - How to define, measure, evaluate, and implement it?*, in *50th National Forum of the American Helicopter Society*1994: Washington, D.C. p. 1039-1053.

208. Dieter, G.E., *Engineering Design: A Materials and Processing Approach*. 3rd ed. 2000, Boston: McGraw-Hill.

209. Yoon, S.-w., *In Search of Meaningful Online Learning Experiences.* New Directions for Adult and Continuing Education, 2003. **2003**(100): p. 19-30.

210. U.S. Air Force, *U.S. Air Force Fact Sheet MC-130J COMMANDO II*, U.S.A.F.S.O. Command, Editor 2011, U.S. Air Force Special Operations Command Public Affairs Office: Hurlburt Field, FL.

211. U.S. Air Force, *U.S. Air Force Fact Sheet HC-130J COMBAT KING II*, A.C. Command, Editor 2011, Air Combat Command Public Affairs Office: Langley Air Force Base, VA.

212. Department of the Air Force, *US Air Force Cost and Planning Factors, AFI 65-503*, 1994, Government Printing Office: Washington, D.C.