1.0 INTRODUCTION

The construct of task learning difficulty originally grew from the need of Air Force managers for a mechanism to assess job difficulty. This need stemmed from a desire of managers, as Mead (1970a) stated:

> "...(a) to develop reassignment systems that would insure that individuals are assigned jobs of increasing difficulty and responsibility as they progress in their career ladder; (b) to assist in establishing minimum aptitude requirements for positions and classes of positions; (c) to compare the difficulty level of work assigned to individuals at various aptitude levels; (d) to compare the difficulty level of work assigned to technical school graduates, individuals bypassing the technical school, and individuals receiving directed duty assignments; (e) to investigate the interaction between job difficulty, job satisfaction, felt utilization of talent, aptitude, and career decisions; (f) to determine the appropriate grade requirements for positions; and (g) to guide decisions about modifications in the classification structure (p.1)."

Over the years that followed, task learning difficulty has received ever increasing emphasis in the measurement of job difficulty levels within and across Air Force specialties. Additionally, task learning difficulty is now a major component in the assessment of minimum aptitude entry requirements and in the assignment of enlistees to Air Force jobs. Task learning difficulty indices have also been used to evaluate training mode utility, training outcomes, job satisfaction, and personnel utilization. Recently, its application has been further expanded to include use as a task screening variable in the development of Specialty Knowledge Test (SKT) weighted outlines (Brown and Weissmuller, 1987) and as an index to stratify task performance domains to insure adequate ranges of Walk Through Performance Test (WTPT) item difficulties (Alba and Dickinson, 1985).

1.1 Purpose

That validation of task learning difficulty has proceeded at a slower pace than the application of this index to real-world problems is a fair assessment. The purpose of this study is to extend the research pertaining to the validation of the construct of task learning difficulty. This purpose will be accomplished by first identifying potential task learning difficulty validation criteria and then analyzing their relationships with task learning difficulty in terms of magnitude and direction.

1.2 Review of Literature

In a series of three studies, Mead (1970a & b) and Mead and Christal (1970) demonstrated a method for evaluating the difficulty levels of Air Force enlisted jobs within the Medical Materiel, Vehicle Maintenance, and Accounting and Finance Career Ladders. In these studies, the job difficulty evaluation policy of supervisors was captured and their judgment decisions simulated with multiple regression equations consisting of three significant predictors: Number of tasks performed (NTP), NTP squared, and average task difficulty per unit time spent (ATDPUTS). The obtained correlations with the criterion for each of the three career ladders ranged from .93 to .95.

Of special significance in these studies was the use of a relative scale to assess task difficulty and the use of "time required to learn to perform tasks satisfactorily" as an operational definition of task difficulty. Both this operational definition and method of assessment have remained remarkably stable over time. Although these studies successfully demonstrated a method for evaluating job difficulty within career ladders, the computed indices could not be used to compare job difficulty across career ladders. Additionally, these studies did not directly address the relationship between job difficulty and aptitude.

Fugill (1972 & 1973) in an important set of studies examined the relationship between relative difficulty level and relative aptitude for groups of mechanical, electronic, and administrative tasks. He found that the

correlations between judgments of difficulty level and judgments of aptitude requirements ranged from .89 to .93. Of particular importance in these two studies is an implicit linking of aptitude requirements and difficulty as an underlying hypothesis. More specifically, given a "time to learn" definition of task difficulty and given that "ability to learn" is a function of aptitude, difficulty and aptitude were assumed to be functionally related. Based on this hypothesis, Fugill proceeded to explore the feasibility of constructing benchmark scales that would anchor judgments about task difficulty and task aptitude to a common frame of reference and allow meaningful comparisons of job difficulty across career ladders.

Christal (1974), Goody and Watson (1975), and Goody (1976) all contributed to the early development of prototype benchmark scales for the general, administrative, mechanical, and electronic aptitude groups. Benchmark scale evolution reached its present state with the work of Burtch, Lipscomb, and Wissman (1981); and Garcia, Ruck, and Weeks (1985). They developed, validated, and implemented three 25-point benchmark scales corresponding to general/administrative, mechanical, and electronics Air Force specialty groups. These scales were based on task learning difficulty defined as time required to learn to perform a task satisfactorily. Using these scales, relative task learning difficulty is mapped onto the benchmark scale by linear transformation. The resulting benchmark value is then cross multiplied by the corresponding percent time spent value to yield an average task difficulty per unit time spent (ATDPUTS) index. This index can then be used to compare task and job difficulties within as well as across specialties comprising the aptitude group of interest.

More recently, Weeks (1984) using average ATDPUTS values, computed indices of occupational learning difficulty for 200 enlisted specialties. He then used these indices to determine the order of Air Force specialty aptitude requirement minimums. Underlying this effort is the implicit relationship (as previously pointed out by Fugill) between aptitude and task difficulty as defined by "time to learn".

The development of methodology to assess task, job, and occupational level learning difficulty has been both thorough and systematic. However, efforts to empirically establish its construct validity have been less organized. Evidence that can be construed to address task learning difficulty validity has often been obtained indirectly during the study of broader questions not directly linked to the construct validity issue.

McCormick and Tombrink (1960) in a comparison of work activity statements provided evidence of test-retest reliability and interrater consistency of scale responses when work activities were reported at the task level. Mean test-retest coefficients reported for the four scales used (Frequency of Performance, Time Required, Mental Difficulty, and Physical Difficulty) were .80, .80, .63 and .64, respectively. Of the four scales, "Mental Difficulty," defined in terms of the degree of mental difficulty involved in task performance had the lowest test-retest reliability. In a later study of task and scale reliabilities and scale interrelationships, Cragun and McCormick (1967) reported a reliability coefficient of .35 for the "Difficulty" scale. Difficulty in this study was defined as a combination of mental difficulty, physical difficulty, and the difficulty of personal relationships involved in task performance. With the shift in the definition of difficulty away from performance and to "time required to learn to perform tasks satisfactorily," reliabilities increased dramatically to levels beyond .90 (Mead, 1970a), (Christal, 1974).

Lecznar (1972) in a study of airmen across eight career fields found positive relationships between TAFMS and each of three criterion variables: Job Difficulty Index (JDI) comprised of NTP, NTP squared, and ATDPUTS; ATDPUTS values; and NTP. He also detected negative relationships between TAFMS and both "utilization of talents and training" and "job interest" for all but the AFS 811XX career field. However, Gould (1972) in a study of job satisfaction of airmen across 97 career ladders reported that the percentage of subjects who felt poorly utilized at the 3-, 5-, and 7- skill levels was 24, 22, and 10 percent, respectively. He further noted that "...as skill levels increase, the tasks performed should become more demanding and hence better utilize talents and training (p.1)." This apparent conflict is most probably a function of the

samples used. Lecznar's sample was restricted to first-term airmen while Gould sampled airmen across 3-, 5-, 7-, and 9- skill levels. In a later study of job satisfaction, McFarland (1976) calculated ATDPUTS, "felt utilization", and job interest indices for 30 job types within the AFS 97XX (Nurse) and AFS 902X0 (Medical Service) career fields. The resulting correlations between job difficulty as measured by average ATDPUTS values and both "felt utilization" and "job interest" were .70 and .50, respectively. Both correlations were significant beyond the .01 alpha level for a sample that included personnel with from less than 1 to more than 20 years TAFMS. Finally, Finstuen and Edwards (1980) in a longitudinal study of job attitudes of Radio Operators (AFS 293X3) found an increasing relationship between ATDPUTS and felt utilization of training.

In an analysis of the difficulty of jobs performed by first-term airmen in 11 career ladders, Wiley (1972) used 40 predictors involving service time, aptitude, place, enlistment age, and technical school graduation in an attempt to account for the variance of three separately treated criterion variables: JDI, ATDPUTS, and NTP. The maximum amounts of ATDPUTS variance accounted for ranged from 22.3 to .07 percent. Time, consisting of one or more aspects of time on the job, time in the career ladder, or TAFMS, was the best predictor. In several career fields, they were the only factors of practical consequence. This appears to imply that the more experienced personnel perform the more difficult tasks. The relationship between aptitude and job difficulty was nonsignificant over half the time and always small. However, the relationship of aptitude, as represented by AFQT score and the applicable aptitude index, with the ATDPUTS criterion was much stronger than the relationships of aptitude with JDI or NTP. This finding was interpreted as a tendency to assign fewer but more difficult tasks to brighter personnel.

Burtch, Lipscomb, and Wissman (1982) made the first formal attempt to validate the learning difficulty index. They compared the relative ratings of senior technicians with corresponding benchmark ratings of occupational experts. The resulting correlations ranged from .54 to .96 for 100 different job specialties. Weeks, Mumford, and Harding (1985) broadened the scope of validation to the occupational level. Using a sample of 5970 students within

48 initial-skills courses, a wide variety of independent variables falling into three general categories (Student Input, Training Outcomes, and Course Content) were used to predict the learning difficulty (LD)[1] index obtained for the job specialty associated with each of the 48 initial-skills courses. The strongest relationships noted were with the Course Content variable set. Among these, course length and course diversity had correlations of .50 and .54, respectively, with the criterion. These relationships were in the expected direction since one would expect the more difficult to learn jobs to be associated with the longer and more diverse subject matter. All observed correlations between LD index and the student input variables were insignificant. This result was also expected since the student input variables were measures of personnel attributes which should not be related to a job property such as learning difficulty. Both results appear to support the construct validity of task learning difficulty.

Ruck, Thompson, and Stacy (1987) in a study of task training emphasis (TE) for first-termers, compared supervisors' ratings on the task factors of TE, Task Learning Difficulty, Probable Consequences of Inadequate Performance, Task Delay Tolerance, and Relative Time Spent. In general, the task learning difficulty factor had significant negative correlations with TE and Percent Members Performing-First Job and significant positive correlations with grade level for 12 of the 18 career fields sampled. The negative correlations found between task learning difficulty and TE are reasonable in that tasks recommended for training emphasis for first-termers should be those tasks that they routinely perform during their first job, not the more difficult tasks in the specialty. We would expect the more difficult tasks to be performed by more senior personnel which is supported by the positive relationship between task learning difficulty and grade level.

In a study of task testing importance for Specialty Knowledge Tests (SKTs), Brown and Weissmuller (1987) examined the policy used by senior NCOs in

---

[1]The LD index was computed by transforming relative ratings to benchmark scale ratings and then weighting the resulting values by the corresponding percent time spent. The resulting indices were then aggregated to produce an occupational-level index of learning difficulty.

assigning "testing importance" values[2] to between 150 to 200 technical tasks within each of six Air Force specialties. The relationships among rated testing importance, task learning difficulty, and percent members performing for these tasks are specified by Appendix A. Evaluation of the obtained correlation patterns indicates that task learning difficulty is a strong component of the policy being used to rate task testing importance. It appears that task learning difficulty has a much stronger influence on the assignment of testing importance values for the E-6/7 SKTs than for the E-5 SKTs. The six correlations between testing importance and task learning difficulty for the E-6/7 SKTs were positive and significant beyond the .01 alpha level, and in all cases higher than the correlations between testing importance and task learning difficulty for the E-5 SKTs. These relationships are not surprising since one would expect a preference for more difficult subject matter on promotion tests designed for administration to the higher grades.

The relationship between testing importance and task learning difficulty for the E-5 SKTs is somewhat more complicated. NCOs assigning testing importance values for the E-5 SKTs appear to be using two different strategies depending on the relative "difficulty" of the task being rated. Tasks with below average task learning difficulty tend to have a much higher testing importance rating than tasks with above average "difficulty." Additionally, percent members performing tasks appears to have a greater influence on assigned testing importance values when task learning difficulty is above average. These relationships are also not surprising and can be explained within the context of E-5 promotion test development philosophy.

---

[2]Between 20 and 40 senior NCOs working within each of six AFSs (representing general/administrative, mechanical, and electronics aptitude groups) were asked to rate sets of tasks on a 1 to 7-point scale of "testing importance" for SKTs. Specifically, these NCOs were instructed to provide ratings of how important it was to include on the SKTs for their particular career fields, those job knowledges needed to perform each task. They were told to rate a task "high" in testing importance if it required knowledges that are critical to successful job performance within the career field. Each task was rated twice, once for E-5 SKT testing importance and once for E-6/7 SKT testing importance. Sets of tasks rated excluded all tasks falling within safety, security, supervision, or training duty groups (only "technical" tasks were rated).

E-5 promotion tests are designed to be administered to Air Force personnel in the grade of E-4. The majority of these individuals have between 4 to 6 years TAFMS and do not have a great deal of experience in performing the more difficult tasks in their respective career fields. In order to develop fair promotion tests for these airmen, it would be reasonable to include subject matter associated with the more difficult tasks only when these more difficult tasks are routinely performed by the group for whom the promotion tests are being developed. The strategy being used to assign E-5 testing importance values appears to reflect this philosophy.

Few studies have examined the relationship between task learning difficulty and task performance. In an early effort, Wiley and Hahn (1977), studied the task performance of incumbents as rated by peers, supervisors, and the incumbents themselves within three Air Force specialties. Although the study results were somewhat disappointing in that few variables other than grade were found to correlate with rated task performance, the authors did note that the prediction of task performance was enhanced as a function of task learning difficulty. That is, prediction was better for the tasks with higher task learning difficulty indices. In a more recent study, Lance, Hedge, and Alley (1987) examined ability, experience, and task learning difficulty as predictors of actual task performance by first term Jet Engine Mechanics (AFS 426X2). Study results indicated that benchmark task learning difficulty was a strong predictor of task performance with a significant correlation of -.26 with the criterion.

If task learning difficulty is a measure of difficulty in learning to perform tasks satisfactorily, one would expect that the more senior personnel in terms of TAFMS, time in the career field (TICF), and time in the job would be performing the more difficult to learn tasks; and that the personnel performing these more difficult to learn tasks would exhibit increased job satisfaction in terms of "felt utilization of training." Secondly, one would expect initial skills training course structure and length to vary proportionately with the learning difficulty of task subject matter. Finally, one would expect an inverse relationship between task learning difficulty and actual task performance —— the more difficult a task is to learn, the more

difficult a task is to perform, given everything else is equal[3]. Results of the previously cited studies do provide some support, although sometimes indirect, for all these assumptions. The remainder of this paper will focus directly on the first two assumptions: (a) the relationship between task learning difficulty and various aspects of service time and job satisfaction, and (b) the relationship between task learning difficulty and the structure and length of training.

---

[3]Initial attempts to assess task difficulty were performance based in terms of the operational definitions used. Unfortunately, a great many intervening variables not directly related to the tasks being performed (Madden, 1962) acted to depress measurement reliabilities. The shift to the more narrow definition of "difficulty in learning to perform satisfactorily" was driven by the need for a more reliable measure that could more easily be linked to aptitude. This shift in definition did not imply that task learning difficulty is unrelated to task performance.