A program of theoretical and empirical research focusing on the ability determinants of individual differences in skill acquisition is reviewed. An integrative framework for information-processing and cognitive ability determinants of skills is reviewed, along with principles for ability-skill relations. Experimental manipulations were used to evaluate the cognitive ability demands associated with information processing parameters of skilled performance. Experiments described include basic information processing (a noun-pair look-up task) and complex problem solving decision making (a terminal radar approach control simulation task—TRACON). Generalization and applications to the public sector (development of an ability-based selection battery for Federal Aviation Administration air traffic controllers) are described briefly. Finally, new theoretical and empirical developments regarding the construct of intelligence as typical performance, and general personality-ability relations are described. Ten figures and one table summarize study results. A list of 80 references is included. Appendix A contains a table of noun-pair results (one figure), and Appendix B described the TRACON task (one figure). (Author/SLD)
Final Report
Contract N00014-80 - 1974

Learning and Individual Differences: An Ability/Information-Processing Framework for Skill Acquisition

Phillip L. Ackerman
University of Minnesota

The research summarized in this report was sponsored by the Cognitive Science Program, Psychological Sciences Division, Office of Naval Research, under Contract No. N00014-89-J-1974, Contract Authority Identification Number. NR 4422-543.

Approved for public release; distribution unlimited.

Reproduction in whole or in part is permitted for any purpose of the U. S. Government.
Learning and Individual Differences: An Ability/Information-Processing Framework for Skill Acquisition

15 May 93
Final Report (8/1/89 - 1/31/93)

C N00014-89-J-1974

University of Minnesota, Department of Psychology
N218 Elliott Hall, 75 E. River Rd.
Minneapolis, MN 55455

Report 93-01

Office of Naval Research
Cognitive Sciences Program (Code 1142CS)
800 N. Quincy Street
Arlington, VA 22217-5000

Unclassified/Unlimited

This final report reviews a program of theoretical and empirical research focusing on the ability determinants of individual differences in skill acquisition. An integrative framework for information processing and cognitive ability determinants of skills is reviewed, along with principles for ability-skill relations. In addition to a review of previous theory and data, experimental manipulations are used to evaluate the cognitive ability demands associated with information processing parameters of skilled performance. Experiments described here include basic information processing (noun-pair lookup task) and complex problem solving decision making (terminal radar approach control simulation task -- TRACON). Generalization and applications to the public sector (development of an ability-based selection battery for Federal Aviation Administration air traffic controllers) is described briefly. Finally, new theoretical and empirical developments regarding the construct of intelligence-as-typical-performance, and general personality-ability relations are described.
I. Abstract

This final report reviews a program of theoretical and empirical research focusing on the ability determinants of individual differences in skill acquisition. An integrative framework for information processing and cognitive ability determinants of skills is reviewed, along with principles for ability-skill relations. In addition to a review of previous theory and data, experimental manipulations are used to evaluate the cognitive ability demands associated with information processing parameters of skilled performance.

Experiments described here include basic information processing (noun-pair lookup task) and complex problem solving decision making (terminal radar approach control simulation task -- TRACON). Generalization and applications to the public sector (development of an ability-based selection battery for Federal Aviation Administration air traffic controllers) is described briefly. Finally, new theoretical and empirical developments regarding the construct of intelligence-as-typical-performance, and general personality-ability relations are described.
Table of Contents

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>V. Abstract</td>
<td>1</td>
</tr>
<tr>
<td>II. Background</td>
<td>3</td>
</tr>
<tr>
<td>A. Review of skill acquisition principles</td>
<td>3</td>
</tr>
<tr>
<td>B. Review of structure of cognitive/intellectual abilities</td>
<td>4</td>
</tr>
<tr>
<td>C. A Theory of Cognitive Ability Determinants of Skill Acquisition</td>
<td>5</td>
</tr>
<tr>
<td>III. Theoretical and Empirical Progress I: Cognitive Ability Determinants of Skill Acquisition</td>
<td>12</td>
</tr>
<tr>
<td>A. Extensions to Learning/Performance Processes</td>
<td>12</td>
</tr>
<tr>
<td>B. Generalization to Complex Task Acquisition - Terminal Radar Approach Control (TRACON)</td>
<td>16</td>
</tr>
<tr>
<td>C. Generalization and Application in a Complex Task Environment - TRACON II</td>
<td>24</td>
</tr>
<tr>
<td>IV. Theoretical and Empirical Progress II: Intelligence-Personality Relations</td>
<td>25</td>
</tr>
<tr>
<td>A. Intelligence and Typical/Maximal Performance</td>
<td>25</td>
</tr>
<tr>
<td>B. Empirical Investigation of Typical Intellectual Engagement</td>
<td>34</td>
</tr>
<tr>
<td>V. References</td>
<td>41</td>
</tr>
<tr>
<td>VI. Appendix A -- Noun-Pair Lookup Task</td>
<td>48</td>
</tr>
<tr>
<td>VII. Appendix B -- Terminal Radar Approach Control (TRACON) Simulation Task</td>
<td>50</td>
</tr>
</tbody>
</table>
II. Background

A. Review of skill acquisition principles

The starting point for a consideration of ability-performance relations regarding skills is not with individual differences, but rather with general theories of skill acquisition. Fortunately, at the level of analysis needed for this discussion, most current theories of perceptual-motor learning are in agreement about the nature of the learning process. Although learning is typically thought of as a continuous process, without breaks or hierarchical plateaus (e.g. see Newell and Rosenbloom, 1981), theorists have often settled on a description of learning that is broken into phases or stages of skill acquisition. More details can be found in the respective descriptions of the theories, but for the current level of analysis, there is substantial similarity in the descriptions of three major phases of skill acquisition, as they are described by Fitts and Posner (1967), Anderson (1982, 1985), and Shiffrin and Schneider (1977).

Phase 1 -- Declarative Knowledge. The first phase of skill acquisition starts with the learner's initial confrontation with the task. During this phase, the learner begins to understand the basic task requirements, the rules for task engagement. In other words, the learner formulates a general idea of what is required of him or her. The term "declarative knowledge" refers to "knowledge about facts and things" (Anderson, 1985, p. 199). For a word processing operator, information that would have to be encoded as declarative knowledge would presumably include rules of starting the program, organization of commands (e.g. menu structure), effects of command implementation on the text (e.g. understanding the difference between "single occurrence" and "global" search and replace commands), location of command keys, and the use of macro (or multiple) commands. The essence of declarative knowledge is that it is represented in a form that generally allows conscious-mediated retrieval. Thus, a test of knowledge at this stage of skill acquisition often takes the form of a written examination of facts and procedures. As the research of Schneider et al. has pointed out, performance at this phase of skill acquisition is slow, attentionall- effortful, and error prone (e.g., see Fisk, Ackerman, & Schneider, 1987).

Phase 2 -- Knowledge Compilation. Following the development of the necessary declarative knowledge base that is required to move beyond a trial-and-error task engagement, the learner can proceed to the next phase of skill acquisition. This phase has been termed the "associative" stage of skill acquisition by Fitts and Posner (1967) because the key element of performance improvements that take place during this phase come about through increasing the strength and efficiency of associations between stimulus conditions and appropriate response patterns. In this sense, the "associations" are formed in the first phase of skill acquisition, but are strengthened in the second phase of skill acquisition. Anderson has coined the term "knowledge compilation" from a computer programming analogy. That is, performance at the first phase of skill acquisition can be thought of as writing a computer program interactively. As each line of code is written, the system indicates whether the code is syntactically correct or incorrect. However, at the "compilation" phase, the program has
been completed and is compiled to be run more quickly and efficiently. Regardless of the analogy adopted, this phase of skill acquisition is typically marked by substantial increases in the speed of task accomplishment, with a concomitant increase in the accuracy of performance. While conscious mediation often takes place during the early part of this phase, with additional practice, the conscious mediation often becomes an epiphenomenon -- it no longer plays a role in determining the speed and accuracy of performance.

**Phase 3 -- Procedural Knowledge.** When the learner has reached a level of skill such that performance is characterized as requiring minimal attentional effort, but at the same time is fast and accurate, the knowledge required to perform the task has become "automatized" (Shiffrin and Schneider, 1977). Anderson chose the term "proceduralized knowledge" to refer to "knowledge about how to perform various cognitive activities" (Anderson, 1985; p. 199). In contrast to declarative knowledge, proceduralized knowledge does not require conscious mediation. In fact, if the task is sufficiently procedualized, so that declarative knowledge is no longer involved in task accomplishment, declarative knowledge of "how" the task is performed can be forgotten, with no decrement to task performance (Schneider & Fisk, 1982). Numerous common examples of such situations can be found, from the person who can effortlessly dial a familiar phone number sequence (but has difficulty writing down the actual number), to a person who operates a complex piece of machinery (but has difficulty in reporting what it is that he/she does).

**B. Review of structure of cognitive/intellectual abilities**

In order to discuss individual differences in skill acquisition, and particularly the role of abilities in predicting performance at the three phases of skill acquisition, it is necessary to explicate a structure of cognitive/intellectual abilities. Although the literature contains numerous competing theories of the structure of human intellectual abilities (e.g. Guilford, 1982; Horn and Cattell, 1966; Vernon, 1961), there is a familial similarity among most current perspectives. These theories have been termed "hierarchical."

Explicit in most of these theories is a general intellectual ability factor (or g). The g ability factor represents the highest node in a hierarchy of ability factors. The influence of such a factor has been estimated by Vernon as reflecting anywhere from roughly 20-40% of the variance in a population of "all human abilities." However, the theories diverge when it comes to identification of factors that constitute the nodes below g. However, all theories appear to be in agreement about the nature of the hierarchy. That is, the general factor represents the broadest ability, and factors at the next level represent broad, or major group factors (e.g. Verbal: Educational, Practical: Mechanical, as in Vernon's theory). Each of the abilities at this broad group factor node may be further fragmented to reveal their constituent abilities. For example, at the next node, the Verbal ability factor might fragment into Vocabulary, Reading Comprehension, Associational Fluency, and so forth. These lower ability nodes may, in turn, be further subdivided to allow representation of the different test formats for assessing the specific abilities, and so on.
A particularly useful heuristic approach to describing the structure of abilities that is also consistent with the hierarchical approach is one developed by Marshalek, Lohman and Snow (1983; see also Snow, Kyllonen, and Marshalek, 1984) called the radex. The graphic representation of the structure of abilities is with a circle, as shown in Figure 1.

Two salient characteristics of this structure should be noted. First, the proximity of a test or measure to the center of the circle is determined by the "complexity" of the material being tested. As complexity of material increases, the measure shares more in common with g or with tests of general intellectual abilities (e.g. reasoning). As the complexity of test material decreases, the measure has less in common with g, and indeed, more in common with narrower abilities and skills. The second salient characteristic of this model is that "content" abilities represent different segments, or slices of the circle. This particular aspect of the structure illustrates the nature of "group" factors of spatial, verbal, numerical (and, perhaps, mechanical) abilities. That is, tests that share the same or similar contents, also have an increased degree of common variance. Thus, the location of a particular test in this structure is a function of the test's complexity of required information processing and the type of content upon which the test is constructed.

While the Marshalek et al. (1983) model is useful for capturing the relations among a great many ability tests, a shortcoming of the structure is that it does not take the speed of processing into account. Ackerman (1988) has proposed a modification of this structure that adds speed of processing as a third dimension, orthogonal to the dimensions described by content and complexity. This model is illustrated in Figure 2.

As shown in the Ackerman model, adding the speed dimension has the advantage of delineating unspeeded types of information processing (such as are required in power tests), from speeded information processing that is required in prototypical perceptual speed and psychomotor ability tests. While the Marshalek et al. model would relegate both highly speeded tests and tests that involve specific content (e.g. narrow tests of spatial or verbal knowledge) to the periphery of the circle, there is now an explicit separation of these two classes of ability measures. As will be discussed below, the Ackerman modification enables linkage between ability constructs and phases of skill acquisition. It is important to keep in mind that this structure of abilities not only is consistent with the Marshalek theory but is consistent with the broad hierarchical theories of intelligence discussed previously.

C. A Theory of Cognitive Ability Determinants of Skill Acquisition [A full report of this investigation is in Ackerman (1988, Journal of Experimental Psychology: General).]

Using the basic outline of skill acquisition reviewed above in conjunction with the ability structure delineated by Ackerman, it is possible to build a representation of the determinants of individual differences in performance during skill acquisition. The structure adopted here is based on the theory proposed by Ackerman (1938) that maps classes of cognitive/ intellectual abilities to the phases of skill acquisition. The theory states that a
Figure 1. Structure of intelligence map proposed by Marshalek, Lohman, & Snow (1983).
Figure 2. A modified radex-based model of cognitive abilities. Complexity is represented as in the Marshalek, et al. model. However, the dimension of level/speed is added to represent perceptual speed and psychomotor abilities. (From Ackerman, 1988).
direct mapping of ability classes and phases of skill acquisition can be established. From this mapping, it then becomes possible to predict the association between individual differences in abilities and individual differences in task performance during skill acquisition. The principles are as follows:

**Phase 1.** Given the fact that the first phase of skill acquisition has substantial demands on the cognitive system (mostly in the domain of attention, memory, and reasoning; but also in the area of broad content knowledge), this phase of skill acquisition closely represents the types of information processing that are sampled in tests of high complexity and minimal speed demands. Thus, general and broad content abilities (namely, spatial, verbal and numerical) illustrated at the top of the ability cylinder (shown in Figure 2) are implicated in predicting individual differences at the declarative knowledge phase of skill acquisition.

**Phase 2.** Once the learner has developed appropriate strategies for successful task accomplishment, and moves to the knowledge compilation (or associative) phase of skill acquisition, the demands on the general and broad content abilities are reduced in a fashion analogous to the reduction in demands on the attentional and declarative knowledge systems. Superior performance at this phase of practice is indicated when learners can efficiently develop streamlined productions for task accomplishment, those that improve speed and accuracy. Similarly, as one moves forward on the speed dimension of the proposed ability structure, from general and broad content abilities to perceptual speed abilities, marker measures require fewer demands on attention and memory and greater demands on the refinement of relatively simple productions (as in a digit-symbol test, or a clerical checking test -- both such tests are prototypical exemplars of Perceptual Speed ability). In fact, perceptual speed ability can be thought of as "the capacity to automatize, by means of practice" (Werdelin and Stjernberg, 1969, p. 192). Thus, individual differences in performance at the second phase of skill acquisition will be well-predicted by individual differences in the perceptual speed ability.

**Phase 3.** When the final phase of skill acquisition, procedural knowledge, is reached, the learner can often effectively meet the task demands with little or no attentional effort (e.g. see Schneider and Fisk, 1982). Performance limitations at this phase, especially when the task is critically dependent on motor operations, are primarily determined by differences in the speed of encoding and responding, with very little involvement of declarative knowledge. The phenomenological essence of such levels of skill acquisition is that task performance is "automatic." The highly deterministic nature of the task has a number of striking similarities to ability measures at the extreme level of speededness, that is, the family of psychomotor tests (such as simple reaction time, or tapping speed). In both task and test scenarios, the individual knows exactly what response (or response sequence) needs to be made once the stimulus is detected. As such, the theory specifies that individual differences in psychomotor abilities will well-predict individual differences in performance at this last phase of skill acquisition.
The overall implication of the theory is that the determinants of individual differences in performance are dynamic, that is, different classes of abilities are most highly correlated with performance at each phase of skill acquisition. Figure 3 illustrates the anticipated pattern of ability-performance correlations across a sequence of task practice, leading to proceduralized knowledge. As indicated in the figure, individual differences in performance at Phase 1 are best predicted by general and broad content abilities; at Phase 2, the best predictions come from measures of perceptual speed ability; and at Phase 3, the most salient correlations are found for psychomotor abilities.

A series of eight experiments (Ackerman, 1988) provided the initial empirical foundation for the theory. These experiments examined the role of task factors of consistency and complexity/novelty in moderating ability-performance relations during skill acquisition. The following conclusions were supported from the basic exposition of the theory and from the empirical results:

1. The novelty and complexity of a task both determine the initial degree of association between general/broad content abilities and individual differences in task performance. (The complexity finding has been previously demonstrated by other researchers, most notably by Kyllonen, 1985.) Tasks which have little in the way of complexity, or those that allow positive transfer of training from other tasks, may by-pass the first phase of skill acquisition and thus be more highly associated with perceptual speed abilities than general/broad content abilities.

2. The consistency of task components, a requirement for transition from declarative to procedural knowledge, is also a requirement for the transition of general/broad content ability determinants of individual differences in performance to perceptual speed and psychomotor abilities. When consistent information processing components are not the dominant elements of the task, performance will continue to have substantial associations with these broad abilities, regardless of the amount of task practice.

3. For novel, moderately complex, but consistent tasks that importantly depend on motor behavior, ability-performance relations are well described by the theory. That is, initial performance is best predicted by general and broad content abilities, intermediate levels of skilled performance are best predicted by perceptual speed abilities, and late, asymptotic performance levels are best predicted by psychomotor abilities.

D. Abilities and Skill Specificity: Test of the Theory Extensions [A full report of this investigation is in Ackerman (1990, Journal of Experimental Psychology: Learning, Memory, and Cognition).]

A subsequent investigation (Ackerman, 1990) demonstrated that some prototypical perceptual speed and psychomotor ability tests are best thought of as microcosms of skill
Figure 3. Hypothetical ability/skill relations derived from the framework. The hypothetical task is moderately complex, involves a moderate amount of broad transfer, and provides for consistent information processing. (From Ackerman, 1988).
acquisition, in agreement with the theory described above. That is, the critical determinants of individual differences in performance on perceptual speed tests in their canonical instantiation are associated with rapid and efficient development of proceduralized knowledge. However, if examinees are given substantial practice on the perceptual speed tests themselves, performance on such tests after practice has more in common with psychomotor abilities than with perceptual speed. A similar effect is found with some predominantly psychomotor tests. That is, post-practice performance on psychomotor tests has less in common with perceptual speed abilities and is increasingly determined by individual differences in asymptotic psychomotor abilities. It was further found that there is substantial underlying heterogeneity to the nature of validity changes associated with practice on prototypical perceptual speed ability tests. Tests that were contingent on proceduralization of rules that were somewhat novel, and mostly consistent, showed the expected shift in predictive validities with practice (i.e. post-practice test performance is more highly associated with post-practice performance on the criterion task). Tests that lacked consistency of information processing components (e.g. number checking), or those that had high initial memory demands (e.g. digit-symbol substitution), showed quite different patterns of initial performance and post-practice validities for a criterion task.

The results show that there is still much work to be done in delineating the predictor space, especially when the additional dimension of test practice (familiarity) is included. On the one hand, single administrations of perceptual speed ability tests showed sufficient communality to "hang together" in a factor solution. On the other hand, the dynamic properties of predictive validity for perceptual speed tests showed too much diversity to provide a coherent taxonomic identification of such tests. Similarly, the results clearly indicated that a distinction between perceptual speed ability and psychomotor abilities can only be drawn in an arbitrary fashion. The dimension of speededness, as with the dimension of complexity, is continuous and multifaceted.

This research was laboratory-based, but there appear to be at least two major implications for applied concerns. The first implication pertains to the method for assessing skill-learning referents of predictor measures. That is, by examining the practice-based validity changes of predictor tests, the current predictor space can be better mapped out, and modified, when the result of such an analysis is an indication of a more cost-effective stage of skills to be predicted (e.g. training success vs. on-the-job performance). The second implication concerns the increased predictive power that might be obtained by using predictor measures that are administered in a repeated fashion. In this way, improved prediction may be obtained at both intermediate levels of skilled performance (from initial performance on the predictor measures), as well as late. asymptotic performance levels on the criterion (from later, post-practice performance on the predictor measures).
III. Theoretical and Empirical Progress I: Cognitive Ability Determinants of Skill Acquisition

A. Extensions to Learning/Performance Processes [A full report of this investigation is in preparation, by Ackerman & Woltz.]

Previous research on individual differences in learning (Ackerman, 1987) indicated that for consistent tasks, interindividual variability nearly always declines with practice, when reaction time is the critical performance variable. Moreover, the general theory of the cognitive determinants of individual differences in skill acquisition (Ackerman, 1988) specifies a reduction in general ability-performance correlations over practice on simple, consistent information-processing tasks. However, pilot study research at AFHRI by D. Woltz suggested an anomalous set of findings regarding both interindividual variability and ability-performance correlations, on a task with perceptual scanning and associative memory requirements. Given the importance of individual differences in working memory for knowledge and skill acquisition (Woltz, 1988), and the potential exceptions to the theory reviewed above, we decided to investigate the nature of learning, performance, and individual differences with this paradigm, called the noun-pair lookup task.

Use of the Noun-Pair lookup paradigm started with an initial AFHRI contract and has continued under the ONR contract with a series of studies conducted at the University of Minnesota (in collaboration with Prof. Dan J. Woltz of the University of Utah; Ackerman & Woltz, in preparation). The noun-pair lookup paradigm is described in Appendix A. To date, five experiments have been conducted that focus on both general issues of learning and performance, as well as focus on issues of the cognitive ability determinants of performance. Although this work is on-going, initial results of the study series are described below:

Learning vs. Performance. One of the salient characteristics of the noun-pair lookup task, is that performance may be accomplished without learning (when a subject simply scans or "looks-up" the matching noun-pair at the top of the display). For varied mapping (VM) versions of this task (where noun-pairs change from one trial to the next), learning of word pairs is impossible, and thus the only method for obtaining correct responses is the look-up method. However, in consistent mapping versions (CM) of the task, a "learning" strategy may be used. This strategy involves committing the word pairs to memory. When a probe is presented, the subjects retrieve the appropriate word-pair from memory, compare the pair with the probe pair, and make a response. When subjects have completely learned the noun-pairs, they no longer need to refer to the list of word pairs at the top of the display. Under these circumstances, response times may be substantially reduced (in comparison to the scan/look-up strategy). Given that the VM version requires a look-up strategy, the difference between CM and VM response times is a good reference measure of the degree of learning that has taken place (especially for data that have been aggregated over subjects).
Individual Differences in Learning. The most striking results of the studies conducted to date involve the joint manipulation of task consistency (consistent vs. varied mapping of stimuli and responses) and the relatedness of word sets. When word pairs are changed from one trial to the next, little overall performance improvement takes place. Regardless of the relatedness of words, the magnitude of individual differences in task performance is slightly reduced with practice, that is, the standard deviation of task performance shows modest decreases as practice proceeds. However, when the word pairings are consistent, the relatedness of words has a substantial impact on both the average learning curve, and on the magnitude of individual differences (see Figure 4). When unrelated words are used, most learners develop associative skills with the task. As the subjects learn the task, between-subject standard deviations reduce substantially (as with most consistent information processing tasks, see Ackerman, 1987 for a review). However, when the task is made more difficult by using word sets that are related (the top row of words representing a single category – such as occupation, and the lower row representing another single category – such as article of furniture), the task becomes more difficult to learn. In this condition, individual differences in task performance are exacerbated (standard deviations increase) until after substantial task practice has been provided.

Examination of individual subject performance clearly indicates that some subjects continue to perform the task as they would an inconsistently mapped task -- that is, they fail to learn the consistent word pairs. In this sense, the task is believed to capture a very important component of 'propensity to learn' that, subject to further investigation, may be general to other situations where either (a) learning is more difficult than 'adequately' performing the task, or (b) task consistency is not readily apparent, even though it underlies the task constraints.

Ability - Performance Relations. Aspects of task performance noted in the analysis of means and between-subject standard deviations are similarly illustrated and also validated by an analysis of ability - performance relations (see Figure 5). When the task is administered with inconsistent mappings of word-pairs, the task is predominantly a perceptual-scanning task. Individual differences in performance of inconsistent versions of the task are associated with Perceptual Speed abilities. However, when consistent versions of the task are utilized, the task is only initially associated with Perceptual Speed abilities. With practice in this condition, the association between Perceptual Speed and individual differences in performance attenuates. When related words are used, consistent versions of the task have a higher association with Reasoning ability and a higher (and more persistent) association with Perceptual Speed abilities. These results are consistent with other data indicating that the related word condition leads to slower learning (especially for some subjects), and a concomitant increase in associations with Perceptual Speed.

The experiments in this series are continuing. The most recent investigation has focused on the volitional nature of learning vs. performance in this task. In two studies, we have demonstrated that there are general effects associated with making learning salient to the subjects (either by embedded memory tests or by self-regulatory interventions). With these
Figure 4. Response time means and standard deviations for Noun-Pair Lookup Task. Left panel -- Consistent mappings (CM) of stimulus pairs. Right panel -- Varied mappings (VM) of stimulus pairs. Both sets include related word and unrelated word conditions.
tools, it has been possible to map out how volitional processes affect the nature of individual differences in performance. Moreover, we have effectively manipulated the aptitude-performance correlations for consistent (CM) task conditions in a way that compresses the individual differences distribution -- specifically by aiding subjects who are otherwise poor learners.

Also in the analysis stage (in collaboration with Dr. Astrid Schmidt-Nielsen, from the Naval Research Laboratory) is an experimental series that attempts to generalize performance on the Noun-Pair task to a similar real-world task (that of learning shortcuts on menu-driven/command-driven computer systems). We believe that this paradigm is relevant to predicting propensity for learning in other situations where performance can be accomplished without extensive learning (such as using a mouse and pull-down menus), but where extensive learning (such as memorizing the command-level language) leads to vastly improved performance.

B. Generalization to Complex Task Acquisition - Terminal Radar Approach Control (TRACON) [A full report of this investigation is in Ackerman (1992, Journal of Applied Psychology).]

This study represented an attempt to extend and test the Ackerman (1988) theory of the cognitive ability determinants of skill acquisition in the context of a task with both high complexity and substantial demands for controlled information processing. The study provided evidence concerning the acquisition of a relatively high-fidelity simulation of many tasks performed by Air Traffic Controllers (the tasks pertain to monitoring, decision making under time-stress, spatial information processing, and perceptual/motor responding). For a full description of this task, see Appendix B. In addition, ability-based prediction of individual differences in performance during skill acquisition was provided in a theoretically-driven framework that includes both general ability (e.g., g), content-based abilities (e.g., spatial ability), and perceptual/psychomotor abilities. From this framework, both task component skill and ability component predictor perspectives can be applied to general and specific aspects of skill acquisition.

For such a complex simulation environment, a comprehensive task analysis was a critical prerequisite for clarity of empirical results. An additional advantage of adapting the TRACON task to the laboratory has been the substantial literature on task analysis for Air Traffic Control (e.g., Alexander et al., 1990; Ammerman, et al., 1987; Means, et al., 1988). These materials have been abstracted and integrated with our own cognitive task analysis, to gain a full understanding of the components of TRACON performance (see discussion in Appendix B. and in Ackerman, 1992; Landon, 1991).

Within the context of broadening the scope of the Ackerman (1988) theory, the goals of the research with TRACON were as follows: (1) to validate the theory of the cognitive ability determinants of performance during skill acquisition; for a complex task that requires a substantial degree of controlled processing resources throughout task practice, and (2) to
Figure 5. Correlations between ability factors and task performance, for consistent mapping of stimulus pairs (CM) and varied mapping (VM) of stimulus pairs. Top panels, correlations between Reasoning Ability and performance (Reaction Time). Lower panels, correlations between Perceptual Speed ability and performance (response time). Both sets include related word and unrelated word conditions.
provide the experimental basis for developing tailored aptitude batteries for predicting the success of trainees for complex tasks, at various stages of skill acquisition.

Procedure. This experiment was completed in 6 sessions (four, 4-hour sessions, and two, 3-hour sessions). The first session was devoted to ability testing and to viewing an instructional videotape specifically designed for this task version. The videotape, 110 min in length, described the major task components, the rules regarding operation of the computer interface (keyboard and mouse), display characteristics, and procedures for accomplishing the controller task. Additional ability testing and TRACON task practice was performed during Sessions 2 - 6. Each of these practice sessions included four simulation scenarios. Each scenario (trial) lasted about 42 min. The entire experiment was completed in 22 hours, including 5 hours of ability testing, 2 hours of task-based video instruction, and 15 hours of total time-on-task in TRACON. Total sample size was 102 subjects.

Predictions and Results. Predictions and associated results for this study were as follows:

**Prediction #1.** Overall task performance will be highly, and consistently predicted by general, and task-relevant broad content abilities (i.e., spatial ability) over task practice.

**Prediction #2.** Overall task performance will show stable correlations with perceptual speed ability and perceptual/psychomotor ability predictors over practice.

**Overall Performance.** To assess the global predictability of performance over practice, the correlations between the general ability composite and overall performance were calculated, and are shown in Figure 6. With the exception of Trial 1, the correlations between general ability and performance show remarkable stability over practice, consistent with a priori predictions (and data from less complex inconsistent information processing tasks) regarding individual differences in performance on tasks that require a great deal of controlled information processing (e.g., see Ackerman, 1987, 1988). These data are entirely consistent with Prediction 1 -- correlations between overall task performance and general ability are high and stable, throughout task practice.

However, general ability - performance correlations don’t tell the whole story of the dynamic nature of ability - performance relations during task practice. Figure 7 shows the correlations between the four classes of ability and performance over practice. Although perceptual speed and perceptual/psychomotor abilities show stable correlations with performance over practice, as predicted, the reasoning and broad content ability (spatial) show increasing correlations. The divergence of reasoning ability and spatial ability correlations supports the inference that once the general rules of TRACON are acquired, spatial abilities account for an increasing amount of performance variance, and so are implicated in limiting individual performance levels (a test for the difference between the regression slopes for the spatial and reasoning ability curves supported this point t(34) = 2.61, p < .05). These data are consistent with both Predictions 1 and 2, that is: (1) Reasoning and Spatial abilities (general and broad content) have consistently high correlations
Figure 6. Correlations between the unit-weighted ability test composite (general ability) and number of planes handled (overall performance) in each TRACON trial, over practice on TRACON.
Figure 7. Correlations between planes handled (overall performance) on each TRACON trial and four ability composites: Reasoning Ability, Spatial Ability, Perceptual Speed Ability, and Perceptual/Psychomotor Ability.
with task performance throughout practice; and (2) correlations between overall performance and perceptual/psychomotor abilities were stable throughout practice.

For task components with consistent information processing components (e.g., airport landing procedures, rules of safe plane separation -- see Appendix B), the following ability-performance relation prediction was made:

**Prediction #3.** Task components with predominantly consistent information processing requirements that rely significantly on motor responses will show diminishing correlations with general and spatial abilities, and increasing correlations with perceptual speed and perceptual/psychomotor abilities.

Ability-performance relations are shown for the arrival and overflight scores in Figure 8. At the level of general ability prediction, both arrival and overflight measures show clearly increasing correlations with general intellectual ability. There was no significant difference in slopes for the general ability - arrival/overflight curves ($t(36) = 1.48, ns$). However, the correlations between spatial ability and performance show a decided differentiation for arrivals and overflights. A simple contrast between the slopes of the curves for the spatial ability composite show marked significant differences ($t(36) = 4.17, p < .01$). Consistent with a priori expectations (i.e., once learned, overflights require far less intervention and spatial reasoning than arrivals -- Ackerman, 1992), these data indicated that while the "arrival" and "overflight" components of TRACON involve similar ability demands at the start of task practice, the spatial demands of the arrival component increased as skills were acquired, relative to the spatial demands of the overflight component.

Perceptual speed ability-performance correlations showed some divergence as well. A comparison between the regression slopes revealed significant differences for the Perceptual Speed ability composite ($t(36) = -3.626, p < .01$). That is, the greater slope of changing ability-performance correlations is found for Overflights. Since overflights have a diminished demand on spatial abilities, and a proportionally greater demand on perceptual scanning and response components, these two sets of results are consistent with the a priori specifications of task requirements, and are consistent with the theoretical perspective (Ackerman, 1988). No significant differences in correlation patterns for the other ability composites were found. These results are generally consistent with Prediction 3, that is, consistent task components (overflights) with motor responding demands, showed increasing correlations with perceptual speed and psychomotor abilities. Although correlations with spatial and reasoning abilities did not decrease, the spatial ability demands were diminished for overflights, late in practice, relative to arrivals (where arrivals had greater demands of inconsistent information processing).

**Prediction #4.** For consistent task components with minimal motor response demands (e.g., declarative knowledge), initial correlations with general abilities will be high, but will attenuate as the information is learned.
Figure 8. Correlations between number of planes accepted (Arrivals and Overflights) in each TRACON trial with five ability composites: General Ability composite, Reasoning Ability, Spatial Ability, Perceptual Speed Ability, and Perceptual/Motor Ability.
Information Requests. Ability correlations with airport information requests are plotted against TRACON trials in Figure 9. It is clear across all ability classes, that frequent information requests are associated with higher ability levels early in practice and either attenuate as practice proceeds, or become negatively correlated with each of the ability classes later in practice. Again, these results are consistent with the inference that the more able, higher-performing subjects access the critical airport information early in practice and acquire the declarative knowledge necessary to perform this task component without further prompts. Lower performing, and lower-ability subjects tended to continue to access this information far into task practice. Moreover, these data are consistent with Prediction 4; consistent tasks with minimal motor response demands showed initially positive correlations with general abilities, but as task practice proceeded, the correlations with all ability classes were diminished.

Summary of findings. The results of this experiment address both theory and application issues in the prediction of individual differences in performance during complex skill acquisition. From a general perspective, this experiment showed that individual differences in complex task performance are indeed predictable, at various stages of skill acquisition, when a theoretically-derived battery of ability measures is administered.

Earlier studies (e.g., Ackerman, 1988, 1990) have shown how general abilities, perceptual speed abilities, and psychomotor abilities are differentially important at each of the stages of skill acquisition when the criterion tasks allow for predominantly consistent information processing. This study extends these findings to show the substantial involvement of general ability throughout task practice on a complex task characterized by a moderate degree of inconsistent information processing demands. In addition, by examining components of ability and components of task performance, interactions of ability-performance relations were observed, in accordance with principles of skill acquisition (e.g., Anderson, 1983; Fitts & Posner, 1967; Shiffrin & Schneider, 1977), and principles of the cognitive ability determinants of skill acquisition (Ackerman, 1988).

The specific results of this study give a snapshot of the nature of the air traffic controller simulation task and the relations between several major class of abilities and task performance during skill acquisition. Some task components are consistent (e.g., airport handoff information, and procedures for issuing commands to planes), and thus can be learned to a level of diminished attentional demands. However, several critical task components require novel or inconsistent information processing demands, and thus overall task performance involves a substantial degree of attentional effort to perform well, at all stages of task practice. As a result, a global general ability composite accounted for a substantial degree of individual differences in performance throughout task practice.

The results supported all four theoretically-derived predictions. Overall task performance, which was greatly dependent on inconsistent information processing components, was highly and consistently associated with general and broad content abilities. Correlations between overall task performance and perceptual speed/psychomotor abilities
Figure 9. Correlations between number of airport information requests on each TRACON trial with five ability composites: General Ability composite, Reasoning Ability, Spatial Ability, Perceptual Speed Ability, and Perceptual/Psychomotor Ability.
were stable throughout task practice. In addition, task components with predominantly consistent information processing demands (overflights and information requests), showed initially high correlations with general and content abilities. After task practice, however, for task components with significant motor response demands (overflights), increases in correlations between component measures and perceptual speed/psychomotor abilities were found. For the consistent task component with minimal motor response demands (information requests), all abilities showed attenuated correlations as the information was memorized by the subjects.

Our experience with this task has shown it to be an excellent prototype for studying cognitive and perceptual skill acquisition, especially in the sense that there are identifiable task components, each with a requisite set of sub-skills that must be acquired for adequate whole-task performance. In addition, given the high level of complexity surrounding the task, it provides an ideal source for developing an intelligent training system (i.e., for a full instantiation of the task, it typically takes FAA and USAF controllers at least two years to reach 'full-performance-level' performance).

C. Generalization and Application in a Complex Task Environment - TRACON II [Ackerman & Kanfer, in press, *Journal of Applied Psychology.*]

A series of studies mapping our integrated approach for cognitive/self-regulation aptitudes to complex skill acquisition has been recently completed. There are three components of this work: (a) A second laboratory investigation of skill acquisition, using the TRACON simulation (part of which was funded by ONR); (b) A field-study for cross-validation of our approach to predicting the success/failure of Federal Aviation Administration (FAA) air traffic controllers (for a 9-week screening program, and beyond); and (c) development of a combined selection/training program for the Minnesota Air Traffic Controller Training Program (MATCTC).

The TRACON experiment in this series was designed to extend the results of the first TRACON experiment (described earlier, and in Ackerman, 1992), in a fashion that accorded greater task practice (18 hours time-on-task), and more extensive theory-based cognitive and self-regulatory aptitude assessment. In this experiment, new cognitive aptitude tests were created to increase overlap between theory and assessment instruments. Specifically, tests were developed to assess: (a) Reasoning, (b) Spatial Visualization, (c) Problem Solving, (d) Perceptual Speed, (e) Spatial Time Estimation, and (f) Spatial Memory. Additional measures of self-regulatory aptitudes were also created for predicting individual differences in skill acquisition. Specifically, measures were developed to assess: (a) emotion control, (b) action control, (c) typical intellectual engagement, (d) self-regulatory processes (self-monitoring, self-evaluation, self-reactions), and several other related constructs. One hundred and twenty-two subjects participated in this 27 hour experiment. Results from this experiment replicate the basic findings of the earlier TRACON study, and show that the self-regulatory measures provide substantial incremental validity in predicting both global and component processes involved in acquisition of this complex task skill. The combination of cognitive
and self-regulatory aptitude measures predicted performance quite well \((r = .654)\) -- accounting for 43% of the interindividual differences variance.

The second component of this series was a cross-validation of the laboratory study to the real-world environment of FAA selection/training of air traffic controllers. A small subset of the measures developed for the TRACON experiment (about 2.5 hours of testing) and the prediction equations developed from the TRACON experiment were applied to predicting the success/failure of \(N = 204\) air traffic controllers in the intensive 9-week selection/training program that all controllers must pass. The predictive validity of these combined cognitive/self-regulatory aptitude measures was \(r = .595\), which translates to accounting for 35% of the variance in performance. This result is noteworthy in that it was a strong test of our lab-based measures in a long-term real-world complex task environment. Moreover, our short test battery has higher validity than the test battery developed by the Office of Personnel and Management (OPM) for the same purpose (the validity of that battery is \(r = .470\), accounting for only 22% of the variance in performance). Finally, our integrated battery provides information for training interventions. Implementation of these interventions may have the pragmatic effect of reducing the wash-out rate for FAA air traffic controllers, which currently runs about 40% (even after an initial aptitude screening process that eliminates about 90% of the applicants for these positions).

The third component in this series was a project to refine and apply the full battery of cognitive/self-regulatory aptitude measures in an operational environment (MATCTC), with follow-up over an extended period of time (initially 6 months). This part of the project was recently completed. Based on 100 students, the validity of the battery for predicting success in a the six-month training program was \(r = .63\), accounting for 40% of the variance. Moreover, the battery provides a superior predictive validity, when compared with the performance of the test developed by the OPM. The OPM test correlates \(r = .46\) with the criterion, accounting for only 21% of the variance, nearly half as much as the new battery.

IV. Theoretical and Empirical Progress II: Intelligence-Personality Relations [A full report of this theoretical formulation is in Ackerman (in press, D. K. Detterman (Ed.) Current Topics in Human Intelligence; Volume 4: Theories of Intelligence.]

A. Intelligence and Typical/Maximal Performance

The purpose of this section is to consider the potential mis-specification of intelligence as a construct solely delimited by maximal performance. The implicit assumptions regarding the desired conditions for testing and the inferences to be made regarding the construct of intelligence will be challenged, and an alternate view that includes "typical" performance will be proposed. Even though the concept of "typical" performance requires additional specifications for (a) volitional, (b) affective, and (c) dynamic constructs, it is proposed here that a theoretical specification of intelligence from such a perspective will be both more accurate (in terms of generalizability to the "real-world") and effective (in terms of predictive/concurrent validational strategies).
Maximal vs. Typical Intelligence. An original conceptualization by Cronbach (1949), and a series of papers by Fiske & Butler (1963, Butler & Fiske, 1955) contrast ability testing ("Try to do your best" (Fiske & Butler, p.251)) from personality testing ("We are ordinarily concerned with the typical (modal or mean?) strength of this tendency [to respond in a given way] because this provides the best estimate of what a person is most likely to do." (p. 258)). Fiske & Butler presented a two-fold rationale for assessing ability at its "maximum" level:

"First, we want a pure measure, one that is determined almost wholly by one thing, the subject's capacity, rather than a measure which is affected by several influences. Second, we measure maximum performance because it is probably more stable than performance under more lifelike conditions."

(p.253).

A main purpose here is to question these stated reasons for focusing on maximal performance. By rejecting these fundamental assumptions about the testing situation, a more complex view of intelligence will be proposed, one that is more comprehensive and yet applicable to many situations where traditional measures of intelligence have lacked validity.

Capacity. Many intelligence theorists have discussed whether extant tests actually focus on ability-as-capacity or ability-as-developed trait (e.g., Humphreys, 1979; Wechsler, 1944). With the notable exception of the fluid intelligence construct proposed by Cattell & Horn (Cattell. 1963; Horn & Cattell, 1966), most modern theorists and psychometrists find notions of capacity measures cause for considerable uneasiness. The main reasons for discomfort with ability-as-capacity assessment are as follows:

(1) Developmental/experiential/cultural components to abilities. From the earliest efforts of Binet and Simon (1905), psychometricians have made many attempts to minimize the impact of cultural and environment-specific influences on intelligence performance. However, while a goal of determining intelligence through testing independent of prior history can be seen as admissible, it is recognized that experiential effects cannot be partialled out of the equation (nor, indeed, should they be). Part and parcel of any practically useful test of intelligence is the test's efficiency in predicting aspects of current or future performance; performances which are undeniably influenced by the individual in the context of interaction with his/her environment.

The concept of capacity is, in a lexical sense, a concept of permanence. However, a central finding in the assessment of intelligence is that changes in test performance occur as a result of maturation and aging (from childhood to adulthood, or even across the adult life-span) (e.g., see Bayley, 1949, 1968; Honzik, MacFarlane & Allen, 1948). Moreover, experiential differences across individuals (e.g., schooling, life experiences) often result in other changes to test performance (see Flory, 1940: Husen, 1951). With such a data base to address, it makes little sense to refer to
unstable capacities of the individual. It has been repeatedly shown (e.g., Thorndike, 1940) that intelligence test performance changes over time to a greater degree than would be expected through measurement error alone. Similarly, to identify an individual's "capacity" as his/her maximal performance on a particular test, would require specification of what experiential or maturational conditions would have had to be satisfied in order to match capacity with performance. In addition, given some early deprivations, it may not be possible, even in theory, to specify the conditions that would allow an individual to develop performance to a level that would approximate his/her theoretical capacity (i.e., if an optimal environment had been available throughout the lifespan). Clearly, as many modern theorists have maintained, intelligence is much more parsimoniously thought of in terms of current level of functioning, either on an absolute scale (e.g., for memory-span) or on a relative scale (e.g., the Deviation IQ).

(2) Ability cannot be divorced from motivation and/or personality. The instructions to test administrators often emphasize the importance of adequate rapport between the administrator and the test-taker. Such an emphasis is explicit acknowledgement (supported by some empirical investigations, e.g., Gordon & Durea, 1948; see review by Sattler & Theye, 1967) that the motivation for performing well on an intelligence test is a critical ingredient for accurate measurement of maximal performance. However, as numerous investigators have demonstrated, both the physical state of the test-taker (e.g., time-of-day, arousal level; see Revelle, Humphreys, Simon, & Gilliland, 1980) and the psychological state of the test-taker (presence or absence of debilitating anxiety, incentives, motivation to succeed or fail, degree of interest in the testing situation) may significantly impact the level of test performance (Quereshi, 1960). It thus makes little sense to discuss ability-as-capacity, unless the concept of capacity is substantially subverted to allow for changes over the course of the day or in line with the individual's volitional state.

The interdependence of ability and volition constructs have been recently noted by Cronbach (1990): "Typical response and ability are not truly separable. A person's record on typing tests establishes that his ability has reached some level, but the score also reflects his willingness to push himself in that kind of situation. That is to say, [the] distinction does not take purposes into account. Purposes are important, but they have been given little consideration in the psychology of individual differences. . . . Test directions almost never tell the examinee how to approach the task. Theoretically, style of performance falls under the head of typical response: but variations in style affect "ability" scores." (p. 38) This point is concordant with a broader context of intelligence-as-typical-performance.

(3) Stability. Fiske & Butler (1963) argued that measures of maximal performance are apt to be more stable than measures of typical performance. This proposal was predicated at least partly on the basis of their contrast between the stability of personality measures (described as measures of typicality) and the stability of ability.
measures (which they termed measures of capacity). However, given that personality and ability measures differ on many other dimensions (e.g., ability test items have correct answers and personality test items do not), it has never been clear that the stability of ability tests is caused by conditions designed to obtain maximal performance. Or, rather, that relative stability is caused by other sources of variance, or by inherent differences in trait stability, not the least of which are the controversial aspects of situational specificity of personality. It simply has not been established that, when instructed to "do one's best" on a personality test (as might be told to an individual who is taking the test under job screening circumstances), personality test scores become more or less stable. Nor, has it been established that instructions to examinees to "respond as you typically would if this was a puzzle in a magazine" to an ability test, make ability test scores less stable.

It seems no less desirable to a psychology of intelligence than it does to a psychology of personality that a construct measure provide "the best estimate of what a person is most likely to do" (Fisk & Butler, 1963). Even if future research were to show that measures of intelligence-as-maximal-performance turn out to be more stable than measures of intelligence-as-typical performance, one must also keep in mind that there is a potential for simultaneously increasing reliability and decreasing validity (e.g., such as when test items are made more homogeneous).

To date, there apparently have been no empirical data reported to support Fiske and Butler's conjecture that tests of "maximal" performance are more stable than tests of "typical" performance when a single construct is considered. Moreover, given the argument that a desire for stability is not necessarily a vector model (in preference scaling, a vector model corresponds to "more is always better"), it follows that Fiske & Butler's arguments about the superiority of maximal performance measures for the sake of stability are spurious ones for defending the sole examination of maximal measures of ability.

(4) Cognitive Information-Processing Approaches to Intelligence. The same assumptions made by Fiske and Butler (1963) regarding the circumstances for ability testing are ubiquitous in studies of the information processing components of intellectual abilities. In fact, while most of the time such an assumption of maximal performance is implicit to the experiment (usually through instructions to subjects to "do your best" or "work as fast as you can, without making errors"), often the assumption is explicit (as when subjects are excluded from analysis because they didn't 'try hard' on the task). Because most studies of information processing and intellectual abilities test both information processing and abilities in the same laboratory context, they represent a special case of the states of the world (i.e., when motivation to perform on the information processing task is likely to be concordant with motivation to perform on the ability tests, given under identical instructions). A similar argument for examining information processing under "typical" circumstances seems appropriate. It might not be as farfetched as it seems on the surface, to explain what
kinds of information processing people "actually do," rather than only what they "can do." Some inroads have been made in this domain, by investigations of reasoning and problem-solving in everyday contexts (e.g., Sternberg & Wagner, 1989), and in the context of research on human error (e.g., Reason, 1990; Reason & Mycielska, 1982), but much more work remains to be done in constructing ecologically valid theories of information processing.

Summary. The central point of these arguments is that the construct of intelligence has become overly restrictive in the ninety or so years of modern theory and research. By expanding the construct to include a dimension bounded by maximal and typical levels of intellectual engagement, it is proposed that a better understanding can be derived for how intellectual abilities interact with other personal constructs (e.g., motivation and volition). In order to generalize to the world outside the testing room or the laboratory, intellectual abilities must be placed in the context of everyday life.

By considering typical performance and maximal performance aspects of intelligence in light of attention and learning issues, a more comprehensive view may be obtained that has the potential to improve prospects of predictive validity for school and job performance, to predict developmental changes in ability-as-maximal performance, and to integrate ability with personality and volitional determinants of performance.

Determinants of Performance. The proposed framework (see Ackerman, in press for more details) states that optimal prediction of performance on any of a wide variety of cognitive tasks requires knowledge of four critical determinants of performance:

(1) **Level of intelligence-as-maximal performance.** This construct corresponds to the ideal of current intelligence testing paradigms. Intelligence-as-maximal performance is the level of performance by an individual that is obtained under the conditions that are optimal for drawing out that individual’s best efforts, while minimizing the deleterious effects of distractions. (For example, in Atkinson’s theory [1974], this would correspond to moderate levels of arousal; in Revelle’s [1989] theory, it would correspond to different levels of induced arousal for introverts and extroverts). To date, this appears to be obtained in testing conditions where examinees are highly desirous of doing well, and test time is limited to a few hours.

(2) **Basal (typical) level of intellective engagement.** This construct corresponds to the usual level of attentional/intellectual effort put forth in the normal day-to-day environment of the individual. Essentially, intelligence-as-typical can be thought of as an individual’s average level of effort elicited across situations as varied as school, work, recreation, family interactions, and so on. It is expected that there are personality, volitional, and interest correlates of intelligence-as-typical performance (e.g., see Dreger, 1968; Humphreys & Revelle, 1984; Kanfer, 1987).
Self-regulatory skill level. Self-regulation, especially in the form of self-monitoring, appears to determine the recognition of intellective demands, and thus sets the stage for decisions made regarding the level of effort allocated to a task (below) (e.g., see Kanfer, 1987; Kanfer & Ackerman, 1989). Individuals who are low in self-monitoring skills are expected to have a generally lower effort demanded - effort supplied coherence, whereas individuals who are high in self-monitoring skills are expected to have a higher effort demanded - effort supplied coherence. That is, effective self-monitoring brings about accuracy in evaluating attentional/intellectual demands of tasks. Increasing accuracy in these demand evaluations can be thought of as doubly important -- because the individual will be able to allocate effort when performance is resource-dependent, and because the individual will be able to conserve resources over time, when the task is resource-insensitive, so that resources will not be depleted when they are called for during other activities.

Decisional processes and dispositional tendencies that bring about responses to perceived intellective demands. An individual’s recognition of intellective demands is just one element in the determination of how much attentional/intellectual effort an individual will devote to a particular task. Other decisional processes that are made up of metacognitive and metamotivational strategies help determine how effort is allocated across various demands (e.g., on-task and off-task effort), see Kanfer (in press) for a review. Similarly, dispositional tendencies toward intellectual tasks (e.g., achievement motivation, or learning/performance orientation) affect the valences of particular tasks. Although these later processes and are more properly considered motivational and/or affective, they play an important mediating role in the engagement of intellectual effort (Kanfer, in press; Kuhl, 1981; Revelle, 1989).

Toward Assessment of Typical Intelligence. It is proposed that the construct of “intelligence-as-typical performance” must supplement the construct of “intelligence-as-maximal performance.” There are several key reasons for this particular proposal:

(1) Predictions of typical or everyday intellectual performance, especially long-term performance, based on measures of maximal performance result in a construct mismatch. The employment or school system domains cannot normally be expected to impose the same contingencies as are found in the usual selection test environment, or the environment where scholastic aptitude measures are administered. Once the job applicant is hired, or the school applicant matriculates, how the individual is likely to perform is best matched by that individual’s intelligence-as-typical-performance.

(2) Generalization beyond the classroom or the job. One of the most-enduring criticisms of modern intelligence measurement is that it revolves around items that are relevant only to academic performance (and by implication or explicitly, not to everyday life). The construct of intelligence-as-typical performance appears, at least on the surface, to be able to encompass many aspects of intelligence that fall outside the traditional
academic domain (e.g., see Sternberg, Conway, Ketron, & Bernstein, 1981, for a description of lay concepts of intelligence).

Measures derived from the intelligence-as-typical-performance perspective proposed here, might well include items that concern such things as pursuit of knowledge and skills outside the classroom (e.g., reading for pleasure; challenging oneself to learn new skills, keep up with current events, and so on). In addition, such measures of typical intellectual engagement may be more highly predictive of developmental changes in intelligence-as-maximal-performance; in that the intelligence-as-typical-performance measures may capture the essence of the individual’s ongoing engagement with the world at large.

(3) When coupled with measures of intelligence-as-maximal performance, measures of intelligence-as-typical performance may provide more accurate predictions of school or job performance at all levels of the respective systems. That is, the fault of typical SAT/GRE exams is that they, as measures of maximal performance, are only good predictors of initial performance (in school, or in an employment training program). See Lin & Humphreys (1977) and Humphreys & Taber (1973) for examples. However, measures of intelligence-as-typical-performance may provide incremental validity for prediction of performance when the demands for maximal intelligence are attenuated. Placing maximal and typical performance together in a multiple regression equation, may make it possible to accurately predict performance at early and late stages of knowledge/skill acquisition.

(4) A forum for examination of ability/motivation/personality interrelations. Many extant theories of personality and motivation have minimized the relations between these constructs and intellectual abilities. However, with perhaps the exception of some goal-setting studies and some arousal studies (e.g., see Kanfer, in press, for a review of goal-setting; and Revelle, 1989 for a review of arousal and personality), there have been mis-matches between the situations for study of personality and motivation, and situations for assessment of intellectual abilities. Given the potential implausibility of personality assessment under "do your best" instructions, discovery of overlap among constructs of intelligence and personality seems most likely to occur with assessment under "typical" conditions for both.

(5) An opportunity to evaluate the notion that there is a "recognition" or metacognitive component to the typical engagement of intelligence. The present contrast between maximal and typical aspects of intelligence implicitly assumes that intelligence does not function like an engine that is always running at full speed. Instead, individuals are posited to differ in (a) recognition of demands on their intellectual/attentional system, as well as (b) the resultant outcome of a cost-benefit analysis of whether the intellectual/attentional system should be engaged, and if so, by how much. Discrepancies between maximal and typical intelligence might serve as prima facia evidence for the plausibility of such underlying processes.
By focusing on the stimulus constellations that provoke high and low intellectual engagement, it may be possible to provide diagnostic information about potential remedies for inappropriate strategies for information processing and sub-optimal learning. That is, the existence of further ATIs that affect engagement of intellectual processing may be explored, assessed, and possibly used to provide any necessary remediation.

(6) Developmental consequences. The literature regarding changes in intellectual ability-as-maximal performance is rife with examples of consistent gains or losses in relative intelligence from childhood into early adulthood (excluding issues of senescence or adult aging-related changes in intellectual performance) (e.g., Bayley, 1968). Several developmental theories have been put forth that attempt to explain such consistent changes in relative standing on intelligence tests, not the least of which has to do with environmental privilege. However, some researchers, such as Nicholls and his colleagues (Nicholls, Cheung, Lauer, & Patashnick, 1989; Nolen, 1988; Dweck, 1986; Dweck & Leggett, 1988), have suggested that individual differences in orientation toward the changeable or constant nature of intelligence are at least partly responsible for gains and losses of relative standing on ability-as-maximal performance measures (see Kanfer, 1990, for a review). Such concepts have been referred to as implicit "performance orientation" (i.e., ability is believed to be constant, and in order to demonstrate one's ability, performance should be accomplished with minimal effort expenditures) or "learning orientation" (i.e., ability is changeable, especially as a result of continued effort expenditures in learning tasks). Although there is little long-term developmental evidence for such propositions, the concepts of "performance" and "learning" orientations are consistent with the concept that individuals differ, at least in their self reports, in desired effort expenditures during task engagement. This is a factor that would lead to some learners showing small discrepancies between intelligence-as-maximal and intelligence-as-typical performance (those with "learning orientation") while other learners would have potentially large discrepancies between intelligence as maximal and typical performance (those with "performance orientation").

The fundamental proposition here, as with some theories of motivation, is that the individual has some degree of choice in how much effort is devoted to a intellectual activities. Such choices, when aggregated over substantial time and situation, will have potentially a significant impact on subsequent measures of ability-as-maximal performance, given the relationship between effort expended and the derative stages of knowledge and skill acquisition.

Structural Implications of the Proposed Theory

Given the fact that much of the extant theory of intelligence is predicated on the structure of intellectual abilities, it is appropriate to ask how the current notions about attention and intelligence, and the contrast between intelligence-as-maximum-performance
and intelligence-as-typical-performance interact with specific classes of abilities. One implication is that intellectual abilities can be organized along the dimension of attentional demands, from highly resource-dependent to highly resource-insensitive. That is, tests of intellectual abilities that tap highly learned declarative or procedural knowledge will likely involve reduced demands on attentional allocations, and thus be more insensitive to differences between an individual's maximal and typical attentional/intellectual engagement. Depending on the population of test takers, tests that have reduced attentional demands might include those of simple numerical processes, such as addition and subtraction, general information, aspects of perceptual speed ability, psychomotor ability and so on. Under the rubric proposed here, performance on such ability measures will be less influenced by manipulations which yield maximal or typical performance.

Tests that tap reasoning, working memory, dealing with novel or inconsistent stimuli, however, will more likely be attentional resource-dependent. As such, large discrepancies in test performance (for at least some examinees) are expected when an exam is administered under maximal or typical intellectual engagement conditions.

Fluid and Crystallized Intelligence. At a surface level of analysis, there are some sources of overlap between the constructs of resource-insensitive/resource-dependent information processing and the Horn & Cattell (1966) constructs of crystallized/fluid intelligence. Many tests that are primarily resource-dependent would be considered as measures of fluid intelligence. Similarly, many tests identified as being relatively resource-insensitive (such as general information) are considered as measures of crystallized intelligence. However, there are some salient differences between these classificatory schemes, as follows:

According to Horn (1965), "Gc can be measured in tests measuring awareness of concepts, facility and quickness in the use of concept labels and in various reasoning tasks involving cultural concepts and generalized solution instruments." (p.309). By including reasoning tasks, though they are bound to "cultural" information, for example, the concept of Gc is at odds with the construct of resource-limited tasks. Furthermore, much of the construct of resource-insensitive information processing is not subsumed under Gc at all, but rather under a general speediness factor (Gs). Thus, Gc neither completely subsumes, nor is subsumed within the construct of resource-insensitive information processing.

When tests of Gf are considered, most seem concordant with conceptions of resource-dependence. However, the Gf construct does not encompass all types of resource-dependent tests (e.g., the "reasoning involving cultural concepts" listed above, and several aspects of the spatial domain, that Horn & Cattell incorporate into the Gv ability). As such, Gf can best be thought of as a subset of the domain of resource-dependent information processing.
How should intelligence-as-typical-performance be assessed?

Determinants of intelligence-as-typical-performance, as contrasted with intelligence-as-maximal performance, are likely to be varied and multidimensional. From the review provided above, several key personal constructs are seen as potential mediators of both current maximal/typical discrepancies, and prospective developmental changes to the differences between typical and maximal performances. Such constructs might include personality characteristics, such as self-monitoring, performance vs. learning orientation, and achievement motives (e.g., see Kagan, Sontag, Nelson, & Baker, 1958, regarding personality and IQ change).

A second source of influence is probably captured by interest measures, or such measures as are often included in so-called "biographical-data surveys." For example, interests in cognitively-demanding activities (such as reading, debating, problem-solving and reasoning games, cross-word puzzles, etc.), may very well tap into an individual’s preferences for intellectual engagement, and thus help describe potential discrepancies between typical and maximal performance levels.

Another major source of influence for typical-maximal discrepancies is likely to be individual differences in self-regulatory skills (e.g., see Kanfer, in press; Kanfer & Ackerman, 1989). As salient determinants of learning efficacy, individual differences in self-monitoring, self-evaluation, and self-reinforcement processes are likely candidates for influencing a learner’s perceptions of the need to engage attentional processes and for maintaining attention over protracted situations when there is little external feedback.

Finally, as Fiske & Butler (1963) recommended vis à vis personality testing, naturalistic observation may also prove to be beneficial in assessing intelligence-as-typical-performance. Discrepancies between typical and maximal intellectual performance seem much more plausible in capturing the "overachievement/underachievement" characterizations that have been attached by achievement test - intelligence differences, given that many existing achievement tests indiscriminantly tap performance on both resource-dependent and resource-insensitive tasks.

B. Empirical Investigation of Typical Intellectual Engagement

[A full report of this empirical study is in Goff & Ackerman (1992, Journal of Educational Psychology]

The purpose of this study was to operationalize several distinct but related personality constructs expected to be associated with intelligence-as-typical performance, test specific predictions about personality/intelligence relations and explore the personality construct space defined by the several personality variables.

Typical Intellectual Engagement is proposed as a dispositional construct that, with related constructs, is expected to be associated with intelligence-as-typical performance.
Unstructured interviews with colleagues were used to assist in initially defining the trait construct space related to a core construct of Typical Intellectual Engagement (defined as a personality trait hypothesized to relate to typical vs. maximal intellectual performance). Nine distinct but related personality constructs were defined:

1. **Typical Intellectual Engagement** is the core construct of the hypothesized construct space. Scale items were expected to differentiate individuals in their typical expression of a desire to engage and understand one’s world, interest in a wide variety of things, and preference for a complete understanding of a complex topic or problem: a need to know.

2. **Hard Work** items indicate a preference for tasks or goals that are difficult to attain or require persistence.

3. **Perfectionist** describes persons who tend to value high accuracy or excellence.

4. **Openness** involves preference for, or high tolerance of, new experiences.

5. **Absorption** describes the situation when people find themselves so deeply involved in an intellectual activity that cognizance of time and the immediate environment is decreased markedly.

6. **Distractibility** may discourage intellectual engagement by preventing initial or sustained focus on a task or problem.

7. **Extroverted Intellectual Engagement** involves enjoyment of engagement with others in intellectual activities.

8. **Introverted Intellectual Engagement** occurs when a person tends to spend much time analyzing their own thinking or wondering about the thinking of others.

9. **Energy** is expected to represent a person’s general energy level.

Finally, four interest areas were included, namely interest in: Art and Humanities, Science, Social Science, and Technology.

Because crystallized intelligence \((G_c)\) is considered to result from purposeful experience, and fluid intelligence \((G_f)\) is thought to be more biological and incidental (Horn & Cattell, 1966), we predicted that measures of Typical Intellectual Engagement and several related personality constructs would be differentially associated with fluid and crystallized intelligence. We predicted positive correlations between Typical Intellectual Engagement and composite test scores representing \(G_c\) and a zero correlation with \(G_f\). That is, the contrasting hypothesis can be stated that the correlation between Typical Intellectual Engagement and \(G_c\) will be greater than the correlation between Typical Intellectual Engagement and \(G_f\).

Typical Intellectual Engagement as one of several indicators of intelligence-as-typical performance is expected to be largely independent of the broad Conscientiousness factor; where conscientiousness is associated with need to achieve, orderliness, conscience, and will (Digman, 1990). While some of these or related constructs (such as Hard Work and Perfectionist above) may be aspects of the expression of typical intelligence, we expected a measure of Typical Intellectual Engagement of an individual with the environment to have more in common with the Openness factor in the broad personality domain and yet exhibit independent aspects as well.
Subjects. One hundred and thirty-eight undergraduate students (77 female, 51 male) provide the data reported below.

Ability Tests. Sixteen paper and pencil ability tests were administered to the subjects to allow testing of predictions regarding ability-personality relations. Four of these tests (Figure Classification, Diagramming Relations, Letter Sets, and Number Series) were markers of general reasoning ability or fluid intelligence ($G_f$) (Horn and Cattell, 1966; see also Horn, 1989). Three tests were markers of general knowledge and verbal ability or crystallized intelligence, ($G_c$) (Analogies, Controlled Associations, and Opposites). Three tests were markers of associative memory (First and Last Names Memory, Picture-Number Memory, and Object-Number Memory). Six tests were markers of perceptual speed ability (Clerical Speed and Accuracy, CA-2, Letter/Number Substitution I & II, Number Comparison, and Name Comparison).

Academic Performance. Available academic performance data for each subject were gathered from official university records. These data included high school grade point average, high school class rank, SAT scores and/or ACT scores, college level hours completed, and college level GPA.

Self-Report Items. A large pool of items that appeared to reflect the constructs involved, broadly defined, were generated or adapted from existing sources. Scales were developed to measure the following constructs:

- Typical Intellectual Engagement (TIE). "I prefer my life to be filled with puzzles I must solve."
- Hard Work. "I would prefer to do a job that was very hard for me rather than one that was very easy."
- Perfectionist. "I enjoy work that requires conscientious, exacting skills."
- Openness. "I prefer activities I've never tried to ones I know I will enjoy."
- Absorption. "When I am concentrating, sometimes it is as if other people just don't exist."
- Distractibility. "When faced with a tedious job, I tend to notice other things I could do."
- Extroverted Intellectual Engagement. "I enjoy involved discussions, even those that last for hours."
- Introverted Intellectual Engagement. "You like analyzing your own thoughts and feelings."
- Energy. "I am rarely too tired to read."
- Interest in Arts/Humanities. "I enjoy trying to figure out what a poet was saying in a poem."
- Interest in Sciences. "I like science."
- Interest in Social Science. "You would enjoy investigating the cause of social unrest in the cities."
- Interest in Technology. "I would be a good person to ask about information on technical problems."

In addition, an existing scale that might be related to the construct of Typical Intellectual Engagement and generally to intellectual performance was included. Academic Comfort is a scale adapted from the Strong Vocational Interest Inventory (SVII) (Campbell & Hansen, 1981). These interest items have been shown empirically to differentiate those successful in academic pursuits from those who are not.

In order to allow investigation of the relations between the personality/interest constructs and the broader personality factor space, the NEO-PI (Costa and McCrae, 1985) was administered. The NEO-PI is a personality inventory designed to tap the five-factor
model personality dimensions: extroversion, neuroticism, openness, agreeableness, and conscientiousness. These "Big Five" dimensions are believed to be generally orthogonal and appear to well describe the dimensions on which persons vary in the personality domain (Digman, 1990, Goldberg, 1990). The Crowne-Marlowe Social Desirability Scale (Crowne & Marlowe, 1964) was administered to detect socially desirable responding, and a small number of items from the other scales were duplicated during administration to detect random responding. Examination of the data did not indicate any extreme Social Desirability scores or random responding.

Analysis of TIE-Ability Correlations

Abilities that were previously selected as markers were standardized and summed to create composites for this and subsequent analyses. The $G_f$ composite includes Figure Classification, Letter Sets, Number Series, and Diagramming Relations tests. The Analogies, Controlled Associations, and Opposites tests comprise the $G_c$ composite.

Predictions were made regarding relations between the nine personality and four interest scales and $G_f$ and $G_c$ ability composites. These relations were evaluated by testing the significance of the difference between these correlations for each of the scales. The procedure suggested by Williams (1959) and endorsed by Steiger (1980) was used to calculate the $t$ statistic for differences between dependent correlations. Results of this analysis are shown in Table 1.

Table 1. Correlations between thirteen personality scales and $G_f$, $G_c$ test composites, with $T$-Test for Differences.

<table>
<thead>
<tr>
<th>Scale</th>
<th>$G_f$</th>
<th>$G_c$</th>
<th>$t$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Typical Intellectual Engagement</td>
<td>-.061</td>
<td>.223</td>
<td>3.32**</td>
</tr>
<tr>
<td>Extroverted Intel. Engagement</td>
<td>-.021</td>
<td>.172</td>
<td>2.20*</td>
</tr>
<tr>
<td>Introverted Intel. Engagement</td>
<td>-.082</td>
<td>.053</td>
<td>1.51</td>
</tr>
<tr>
<td>Absorption</td>
<td>-.052</td>
<td>.137</td>
<td>2.15*</td>
</tr>
<tr>
<td>Interest in Arts and Humanities</td>
<td>.006</td>
<td>.238</td>
<td>2.68**</td>
</tr>
<tr>
<td>Interest in Social Science</td>
<td>-.090</td>
<td>.024</td>
<td>1.28</td>
</tr>
<tr>
<td>Openness</td>
<td>-.060</td>
<td>.143</td>
<td>2.31*</td>
</tr>
<tr>
<td>Hard Work</td>
<td>-.167</td>
<td>.096</td>
<td>3.04**</td>
</tr>
<tr>
<td>Perfectionist</td>
<td>-.150</td>
<td>-.009</td>
<td>1.80</td>
</tr>
<tr>
<td>Lack of Distractibility</td>
<td>.014</td>
<td>.072</td>
<td>.66</td>
</tr>
<tr>
<td>Energy</td>
<td>.048</td>
<td>.039</td>
<td>.10</td>
</tr>
<tr>
<td>Interest in Science</td>
<td>.040</td>
<td>.078</td>
<td>.42</td>
</tr>
<tr>
<td>Interest in Technology</td>
<td>.131</td>
<td>-.138</td>
<td>-3.12**</td>
</tr>
</tbody>
</table>

Note. N = 138. * p < .05, ** p < .01
Significant differences between the individual scale correlations with the $G_f$ and $G_c$ composites were found for TIE, Extroverted Intellectual Engagement, Absorption, Interest in Arts and Humanities, Openness, Hard Work and Interest in Technology. However, unlike the other scales, Interest in Technology was negatively related to $G_c$.

Selection in the college student sample involved in this study significantly reduces the variance in ability as compared to the population at large. In order to evaluate the effects of this restriction of range, a correction for explicit selection (Thorndike, 1949) was applied to the correlation between TIE and $G_c$, based on ACT composite test score national norms (American College Testing Program, 1989). The resulting correlation ($r = .333$, [unrestricted standard deviation = 6.0]) indicates that approximately 11% of the variance in $G_c$ can be accounted for by TIE. The corrected correlation between TIE and $G_f$ ($r = .113$) results in a variance-accounted-for of approximately 1%.

The theoretically expected divergence in association between TIE and related scales with $G_f$ and $G_c$ was supported by the data. This magnitude of association while modest, may underestimate the overlap between the respective constructs. This may occur because $G_c$ measures, although they attempt to assess broad acculturation and experience, are nonetheless administered in a maximal context. It is possible that part of the development of $G_c$ results from individual differences in typical engagement with the environment and culture in an intellectual context. It is, however, impossible at this point to attribute causal meaning to the associations described by the current data. The reverse argument is also possible: $G_c$ abilities may be an enabler for higher typical intellectual engagement. It also is plausible there is a symbiotic relationship -- that is, high Typical Intellectual Engagement leads to development of higher $G_c$ ability which in turn enables higher intellectual engagement.

**Typical Intellectual Engagement in the Personality Construct Space**

Ecscal, a non-metric multidimensional scaling routine (Tellegen, 1988; Waller, Lykken & Tellegen, in press) was used to obtain a multidimensional scaling solution in order to facilitate visualization of the construct space surrounding Typical Intellectual Engagement. Scale intercorrelations were used as a similarity matrix for the scaling. The thirteen personality/interest scales, the NEO-PI scales representing the five-factor model, and the SVII Academic Comfort scale were included. Results are presented in Figure 10.

Those scales correlating positively with $G_c$ and that are shown to differentiate significantly between $G_f$ and $G_c$ in Table 2 are grouped near the center of the solution clustered around the core construct. Typical Intellectual Engagement. NEO-PI Openness, the scale found to be most related to the Intellectual Engagement and Openness factors in the factor solution, is near this grouping. The NEO-PI Neuroticism, Extroversion, Agreeableness, and Conscientiousness are peripheral.
Figure 10. Ecological multidimensional scaling solution for scales of Typical Intellectual Engagement and returnee scales (NEO-PI and SVII Academic Comfort). (TIE = Typical Intellectual Engagement.) (From Goff & Ackerman, 1992).
Conclusions

Typical Intellectual Engagement and the other constructs investigated are but a small subset of many possible determinants of typical intellectual performance. Others include the attentional demands of an intellectual challenge, self-regulatory skills and motivational factors, and maximal ability. When students are faced with choices for educational activities, Typical Intellectual Engagement may influence both the choice of activities and the style of interaction with such activities. Such influences are expected to have long-range implications for educational achievement, especially when student choice is involved. That is, as the educational situation becomes less constrained (e.g., when the curriculum allows students to choose elective courses), individual differences in typical intellectual engagement are expected to be emergent determinants of achievement. Future efforts in developing aptitude-treatment interaction manipulations (where the "intellectual" nature of the activity is manipulated), may yield different levels of attention and persistence for students low and high on TIE-type measures.

An understanding and re-conceptualization of intelligence as typical performance may provide a basis for development of measures for predicting performance in educational and work environments. This personality construct has the potential for giving a concrete representation for some aspects of folk concepts of under- and over-achievement, in a fashion that avoids the pitfalls of difference scores (when aptitude and achievement scores are compared, see Cronbach, 1984).

The current investigation demonstrated that individuals differ in Typical Intellectual Engagement and that those differences are measurable. Typical Intellectual Engagement may be related to acculturative and purposeful development and expression of certain intellectual abilities (such as $G_c$) and to typical intellectual performance. Future refinements to the specification of the construct, and to the assessment of Typical Intellectual Engagement may provide a source for both educational prediction and classification applications.
V. References


Humphreys, L. G. (1979). The construct of general intelligence. Intelligence, 3, 105-120.


VI. Appendix A -- Noun-Pair Lookup Task

The Noun-Pair Lookup Task is essentially a combination of perceptual scanning and associative memory information processing components. (See Figure A1 for an example.) The task involves two sets of words (or word-number combinations), and a probe pair of words. Under ordinary conditions both rows of "memory-set" words are simultaneously presented with a single probe word-pair. The subject's task is to determine, as quickly as possible, whether the probe pair of words matches a paired set of words in the top two rows of the display.

The program designed for presentation of the task and collection of responses is flexible, and allows for several instantiations of the task, the following of which have been studied to date: (a) Consistency of word-pair mappings (consistent vs. inconsistent), (b) Yes/No response vs. Construction of response (where a single probe word is presented, and the subject responds with the associated word); (c) Stimulus manipulation (related word pairs vs. unrelated word pairs), (d) Keyboard response vs. Voice responses (via speech recognition hardware and software), and (e) memory testing (where the top rows of word-pairs are not displayed).
Figure A1. Noun-Pair Lookup Task. The above example uses related word sets, and is given in a Yes/No keyboard response condition.
VII. Appendix B – Terminal Radar Approach Control (TRACON) Simulation Task

The task used for this research is a uniquely modified Professional version (V1.52) of TRACON simulation software, developed by Wesson International. Versions of this program have been, and are, in use in several locations in this country (including the FAA, NASA, DoD, and several colleges and university airway sciences programs) for training of air traffic controllers. Modifications for the current instantiation of the program allowed for the collection of a variety of data, described in more detail below.

Task analysis is a critical component of empirical investigation with any task, but especially with a complex task such as TRACON (or the task used by Donchin, et al., 1989; see also Frederiksen & White, 1989). In fact, one of the salient reasons for adoption of the TRACON task here is that there has been extensive historical work on task analysis for the full range of Air Traffic Controller tasks, including the subset implemented in TRACON. In particular, a five-volume set of task analysis materials has been prepared by the FAA (e.g., Alexander, et al., 1990; Ammerman, et al. 1987). Additional task analyses have been conducted by the U.S. Air Force, and more recently by HumRRO (Means et al., 1988). All of these have been carefully studied in an effort to derive the critical components of the simulation task adopted for empirical evaluation (see, e.g., Landon, 1991).

The task requires that subjects learn a set of rules for positive air traffic control, including (a) reading flight strips, (b) declarative knowledge about radar beacons, airport locations, airport tower handoff procedures, en-route center handoff procedures, (c) plane separation rules and procedures, (d) monitoring strategies, and (e) strategies for sequencing planes for maximum efficient and safe sector traversal. In addition, subjects are required to acquire human-computer interface skills: including issuing mouse-based commands, menu retrieval, keyboard operations, and integration between visual and auditory information channels. Although the task represents a substantial reduction of rules and operational demands in comparison to the real-world job of an Air Traffic Controller (ATC), it represents an excellent simulation vehicle for study of skill acquisition, within a time-frame that can be handled in a laboratory-based research environment.

Display. TRACON presents the controller (subject) with a simulated color radar screen, depicting a region of airspace, radio navigational tower locations (VOR), airports, sector boundaries, and range rings. Planes are identified by an icon on the radar scope, with a data tag that indicates plane identification and altitude information. In addition, two sets of "Flight Strips" are presented at the right side of the display, a "Pending" and an "Active" set. Each flight strip contains information about a particular flight, including identification information, plane type, requested speed and altitude, and sector entry and exit destination information (See Figure B1). Finally, at the bottom of the screen is a "Communications Box" -- which shows commands issued to planes (and responses by the pilots), along with the controller’s "score" for the current simulation.
Communications

BA93 heavy! Switching to center frequency.
App/Dep: N66CD, Contact center. Good day.
N66CD: Switching to center frequency.
App/Dep:
Figure B1 (Preceding Page). Static copy of TRACON® screen. There are three major components to the display. The right hand side of the screen shows Pending (not under control) and Active (under control) flight strips. Each flight strip lists (a) plane identifier, (b) plane type, (c) requested speed, (d) requested altitude, (e) Radar fix of sector entry, (f) Radar fix of sector exit (including Tower or Center). The lower part of the screen shows a communications box that gives a printout of the current (and last few) commands issued by the subject, and the responses from pilots or other controllers. The main part of the screen shows a radar representation of the Chicago sector. Planes are represented by a plane icon and a data tag (which gives the identifier, the altitude, and an indication of current changes in altitude). The sector is bounded by the irregular dotted polygon describing a perimeter. Radar fixes are shown as small (+) figures on the radar screen. Airports are shown with approach cones, and a circle indicating the facility proper. A continuous radar sweep is shown (updating at 12 o'clock, every 5 sec). Range rings are also displayed, indicating 5 mile distances.
When planes are about to enter the subject's sector (at a boundary or on the runway of an airport) this information is announced over the headset (e.g., "Northwest 123 ready for takeoff," or "Delta 123, with you, level at 9,000"). However, no flight is allowed to cross the sector boundary or take off from an airport without explicit authorization.

**Task Controls and Knowledge of Results.** Subjects interact with the TRACON simulation in several ways. A mouse is used for the majority of input activities, although the keyboard was also used alone, or in conjunction with the mouse. (Additional pilot experiments have been conducted in our laboratory, using voice recognition hardware/software under these conditions, no other source of interaction is used.) For each plane command, a menu of command choices is displayed on the screen. For turns (Left or Right), a small wheel was shown so that a direction was selected by pointing to the appropriate place on the wheel, and depressing the mouse key. Altitude and Speed commands resulted in the display of a small linear display from which the subject selected a particular value. Direct and Hold commands require the subject to move the mouse-cursor to a specific VOR fix or airport, and select a location. Resume and Handoff commands have no additional menus, but are initiated directly.

Additional commands for information (Flight Path, Plane Type, and Plane Current Heading (in degrees) and Airspeed (in knots) may also be obtained. Information pertaining to the sector constraints (Map of VOR/Airport fixes; and Airport Information, including final approach heading and altitude requirements) may be called up with keyboard commands.

Knowledge of results is provided visually (by text in the communications box) and auditorially with a read-back by the pilot or other controller (using digitized speech broadcast over the subject's headset). If a command is not allowed (e.g., asking a pilot to increase or decrease speed beyond the limitations imposed by the type of plane), the visual and auditory response indicated a failure to comply with the command (e.g., "Sorry, but that is below my 'stall' speed!"). Handoff commands differ from the other commands, in that a handoff to another sector is only accepted when the plane is within 5 miles of the sector boundary. All other requests for handoff are refused by Center Controllers.

In addition, planes follow (as nearly as possible) the commands issued by the subject. Turn, altitude change, and speed change commands are processed by the computer, and are carried out in accordance with the limitations imposed by each aircraft type (e.g., smaller planes turned in a smaller radius than Boeing 747's, but 747's climbed more quickly than the smaller planes). Each plane performs within the constraints that were displayed when a subject calls up the information for that plane type.

Finally, when errors occur (e.g., separation conflicts, near misses, crashes, missed approaches, handoff errors), additional information is presented to the subject. In each of these cases, an alert circle around the plane(s) in question is presented on the screen, and a series of tones are presented over the headset. If two planes crash, a message appeared on the screen indicating which of the planes crashed. (In normal training, the simulation is
immediately halted under such conditions. However, because it is not desirable to minimize learning opportunities of subjects who have crashes, the simulation continues under such circumstances.)

**Points.** Subjects are told to perform the task so that they maximize successful disposition of all flight paths, but that safety is a critical component of the task. Points are given for successful accomplishment of each plane's flight plan, and penalty points are deducted for both commission or omission errors. Points assigned are based on a priori judgements of task component difficulty (e.g., arrivals were more difficult to accomplish than overflights, so arrivals received three times as many points). The point assignments are used to encourage subjects to develop an appropriate strategy for task component emphasis.

**Trial Description.** Trials for the task are created and pretested to be roughly equivalent in difficulty. Each trial contains planes that are divided into three basic categories (Overflights, Departures, and Arrivals). Overflights are planes that enter and exit the subject's airspace at cruising altitudes. Subjects are required to acknowledge these airplanes as they approach a boundary VOR fix, monitor progress through the sector, and handoff to a "Center" controller. Departures are planes that originate at one of the four airports, climb to a cruising altitude and are handed off to a "Center" controller. Subjects are required to release departures from airports, evaluate and remediate potential conflicts as the planes climb to a cruising altitude and turn to intercept their intended flight paths, and then handoff planes to the appropriate Center controller. Arrivals enter the subject's airspace from one of the boundary VOR fixes, and have to be landed at a designated airport. Subjects are required to direct arrivals onto an appropriate heading and altitude to provide an acceptable handoff to the appropriate Tower controller, then these planes can land. Practice flights, which originate at an airport, but have to be correctly vectored to be landed again at the same airport are classified as "Arrivals," because demands of these flights are most similar to other arrivals. For all flights, the subject is required to maintain legal separation (at least 1000 ft in altitude, or 3 mi horizontally).

Each trial is comprised of 16 overflights and departures (with roughly equal frequency), and 12 arrivals. The planes request entry to the airspace at irregular intervals that are constrained to require the subject to be always occupied with at least one active target. The trials are also constrained so that perfect performance (handling all 28 planes successfully) is just beyond the skill level achieved by subject matter experts. Each trial is concluded in 43 - 45 min. That is, in order to provide equivalent practice time across subjects, the trials are ended with time constraints, rather than waiting until all planes are handled -- which otherwise introduces a substantial variance in practice time.

A successful "handle" of a flight is the appropriate accomplishment of the respective flight plan. That is, for a departure or an overflight, the accomplishment was a successful handoff to the appropriate Center controller. For a landing, the accomplishment was the successful landing of the airplane.
Errors in performance take a variety of different forms, as follows: (1) Incorrect speed or altitude for center handoff; (2) A failure to handoff the plane; (3) For arrival flights, errors include "wrong approach altitude," or "wrong approach heading" (which requires the subject to reorient the plane for another landing attempt); and (4) Separation conflicts, near misses, and crashes (for a differing degrees of airspace proximity violations).

**Performance Measures.** After extensive review of the raw data from initial experiments (which include every command issued by subjects during TRACON trials, and a series of summary data for each airplane in a simulation, and for each simulation overall), a general criterion of merit has been selected that reflected overall task performance. This measure, called "Overall Performance" is computed as the sum of all flights accepted into the sector that have a final disposition within the simulation time (minus any planes that are incorrectly disposed of -- e.g., crashes, not-handed-off, vectored off the radar screen). This measure is generally concordant with results from the examination of the criterion space for FAA ATC simulation research (e.g., see Buckley, Debaryshe, Hitchner, & Kohn, 1983). Other measures are also computed, to reflect declarative knowledge (information requests) and task component processing (separate scores for number of arrivals, departures, and overflights accepted into the sector).