COGNIMETRICS
innovative assessment solutions

corporate office: 211 Jardin Vista
San Antonio, TX 78258
phone: (210) 481 – 0261
email: info@cognimetrics.com
web side: www.cognimetrics.com

# Army Alpha Technology Was Developed Ninety Years Ago.

# Anything New?

Martin J. Ippel, Ph.D.

CogniMetrics, Inc.

San Antonio,TX

Paper Prepared for the First Workshop of the

Military Aviation Careers Consortium

at

Arlington, Virginia

July 28 –29, 2005

## Introduction

In the course of the 20th century technological developments have completely changed the landscapes of education, healthcare, work, and warfare. Yet, the mental test technology still solidly rests on the foundation laid by the US Army Alpha Test developed in World War 1. Sure, the market is now flooded with "computerized" tests, but these are simply paper-and-pencil tests adapted for the computer. Test items that previously were presented one-by-one in test booklets are now presented one-by-one on a computer screen and the examinee's performance still consists of selecting one of the multiple answers to these questions (Ippel & Hurwitz, 1998). Test designers continue to work within a testing paradigm that is constrained by the limitations of the old medium (printing on paper)[1] and do not apprehend the full potential of information technology in realizing new measurement methods and opening up new domains of measurement.

## Constraining effects of the old medium

One example of the constraining effect of the printing medium can be found in the measurement of technical aptitudes. Declarative knowledge is the preferred type of knowledge to be represented in print.[2] Even when technical knowledge is the measurement objective, it is exclusively treated as declarative knowledge (e.g., ASVAB subtests: Auto/Shop, Mechanical Comprehension, Electronics Information). However, the relationship of technical knowledge as declarative knowledge with an aptitude for technology is - at best – indirect (i.e., it is likely that individuals with a high technology aptitude are interested in technology and thus acquire a relatively large (declarative) knowledge base of technical domains compared to individuals who lack that interest).

Although the traditional test paradigm has shown to be adequate for many purposes, it limits cognitive testing to certain domains at the exclusion of others. Technological developments, in particular the advent of Information Technology, have made the ability to deal with systems that have a dynamics of their own increasingly important. However, within the traditional test paradigm it is very difficult, if not impossible, to capture the intricacies of procedural skills such as involved in technological

[1] This also includes (hand) written texts, and figures.

[2] Procedural Knowledge is the knowledge of "how to" (e.g., how to replace a car battery). Declarative Knowledge is the knowledge of "what" (e.g., what agent in a car accumulates and disperses electricity).

aptitude(s). Procedural skills typically consist of sequences of actions, often in response to state changes in a physical (or conceptual) system. To measure just the final outcome of these action sequences misrepresents the nature of procedural skills.

### Testing Procedure Acquisition Aptitude

I founded CogniMetrics, Inc.[3] with the explicit purpose to research and develop innovative assessment tools. Our flagship product, the Information Technology Aptitude Battery (ITAB), is a good example of innovation in the technology of mental measurement.

The purpose of the ITAB tests is to measure the aptitude to learn procedural skills. In order to do that the tests provide a task environment in which the examinee has to develop procedures (or algorithms) to achieve a goal. The test measures how examinees incorporate feedback from the system into their follow-up actions and how quickly this leads to the build up of a more or less efficient algorithm. The test scores reflect the efficiency of these procedures and estimate how much exposure (i.e., training) the examinee would need to be able to develop a maximally efficient procedure.

Two basic innovations are: (1) the test provides an interactive environment, and (2) actions of the examinee are not scored as singleton answers to distinct problems, but are analyzed as sequence patterns.

- Complete interactivity is achieved by creating an internal representation of the task-environment. Artificial Intelligence technology is used to compute the "intelligence" of each step taken by the examinee.
- Unlike the present generation of computerized tests, the ITAB tests *do not* consist of items with a standard set of response alternatives. Within the task-environment created by the Hidden Target test, the examinee is free to act. The examinee must produce sequences of actions to achieve a certain goal. These examinee actions should use the feedback from the system. Performance assessment is based on an analysis of these action sequences.

Each ITAB test provides an interactive task-environment. An examinee action (user action) changes this environment (system action) and the examinee receives infor-

[3] See: www.cognimetrics.com.

mation about the new state of the task-environment (system output). The cognitive diagnosis component of the tests measures these changes per user step and analyzes sequential aspects of the user action sequence (Markov model) and the informational content of each new system state.

For both ITAB tests the many measurements generated by the respective diagnostic components are organized into **three scores** that are intuitively accessible, viz., INTELLIGENCE (efficiency in extracting information) and SYSTEMATIC APPROACH (measuring whether the individual follows a steady approach and is careful in his/her inferences), and PERFORMANCE (a multifaceted performance score for procedural aptitude). The scientific basis of structuring of test information in this way is an empirically tested structural equations model (Figure 1).

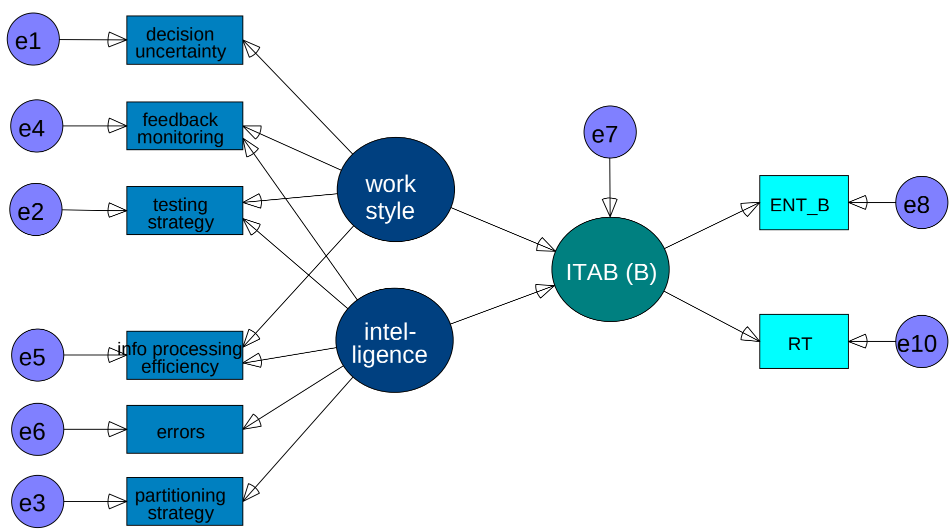


**Figure 1.** Structural Equations Model Generating ITAB Scores. In this case the SE model is for the Battery Test.

## Aviation constitutes a high-performance work environment

The following is based on my research on remotely controlling airplanes (UAVs) at the AF Research Laboratory at Brooks AFB in the late 90s (Ippel & Watson, 1998; Watson and Ippel, 1998).

UAVs operate under stringent time and fuel constraints. Missing a target in a reconnaissance mission often represents a failure that is hard to correct during the same mission. UAV teams operate in a high-performance work environment. To maintain workload at manageable levels, technology has to be developed to either empower human operators to handle the workload, or to distribute the workload between human operators and intelligent decision aids.

In this section I would like to focus on the general characteristics of those empowering tools. The following observations may help to better understand what is at issue here.

- A UAV is a tool itself that empowers human operators to precisely target objects, or persons, over large distances.
- It has a relatively large (output) action repertory that is controlled by a relatively small set of different user actions.

The latter observation is particularly important. Many empowering tools derive their versatility from a relatively complex internal structure that maps the same user actions on different system (output) actions dependent upon its current (internal) state. An example that we are all familiar with is the cruise control system in cars, a standard feature of cars in the US nowadays. The purpose of cruise control is to lower the workload of the driver.[4]

What causes a typical danger in driving on cruise control? Often the driver is mistaken with regard to the current (internal) state of the cruise control system. We call that "state confusion". The choice of user action has to be derived from the required system action given traffic conditions. For example, accelerate or slow down. It is clear that state confusion can be dangerous if a particular user action would result in different system actions in different states of the system, or fail to have an effect when the state is in a state on which the user action is not defined.

[4] Somewhat odd is the fact that cruise control systems lower the workload safely only when the workload is at its lowest already (i.e., on quiet Interstate Highways).

The UAV control task is highly susceptible to state confusion risks. In a typical UAV control system one of the operators is sitting in front of a monitor that displays a 2-D map with a north-up alignment, also showing the current position and direction of the UAV. Now, let us assume the UAV is flying alongside the *Y*-axis of his computer monitor and the UAV camera is in the same direction as the UAV roll axis. Then, the effect of a user action, say, *forward*, is to move the scanned spot further up alongside the *Y*-axis. However, when the UAV is flying alongside the *X*-axis, say, eastward, the same user action now causes the camera to move along the X-axis. Thus, the same user action corresponds with more than one output action. The problem for the UAV operator is to translate his perspective (north = up) into the UAV perspective.

What is (or probably: was)[5] exactly the task of a UAV camera operator? Any time a camera operator takes action to cause a particular camera action, he faces an uncertainty of which user action should be chosen. Given the present direction of the UAV camera (which is the direction in which the UAV flies) and the location of the object to be scanned a particular camera operation is expected. Let's assume for the sake of simplicity that the camera operator works with four arrow keys (i.e., possible user actions) and only on of these keys can cause the required camera operation (i.e., system action) given the task situation.[6] The amount of selective work required in performing the Camera Directory Task can be divided into two stages. First, the camera operator has to extract information from the task situation, that is, he has to map the task situation onto a required camera operation. Second, the camera operator has to decode the expected camera action into an arrow key (i.e., a camera operator action) to cause that particular action. Discrepancy between the operator perspective and camera (or UAV) perspective – most likely – affects does the first stage (for details see: Ippel & Watson, 1998).

Steve Watson and I reported a study (Ippel & Watson, 1998) with a prototype of a test to assess a person's capacity to handle the various degrees of discrepancy between the two perspectives. Figure 1 presents the average performance of a group of 173 Air Force recruits on this test. The testees had to shoot targets while flying a fighter jet in various different directions (under 0, 30, 60, …, 300, 330 degrees with a north-up alignment.

[5] GPS systems could significantly simplify the camera operator task.
[6] This reflects the situation in our prototype test.

When the operator selects the appropriate arrow keys to cause the required camera actions all the time, we speak of maximal information transmission. Figure 2 shows that information transmission is close to maximal when the two perspectives coincide. As the disparity between the perspectives increases the number of state confusions increase and errors occur. Note that:

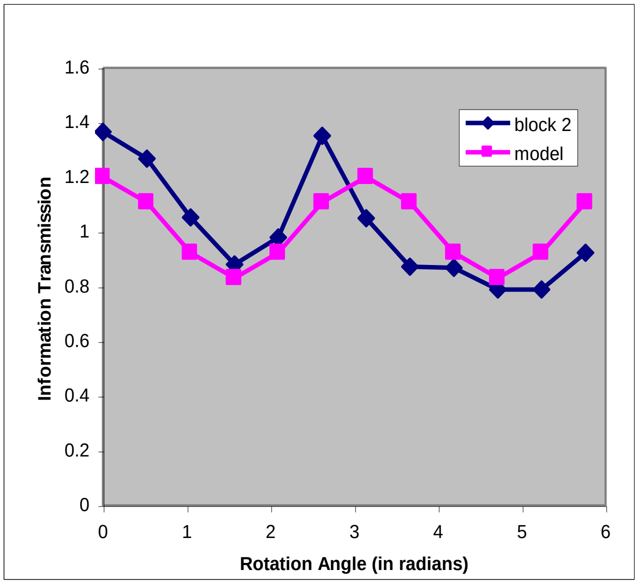


**Figure 2.** Information Transmission as function of the disparity in operator and UAV perspectives.

1. This does not mean that the ***maximum*** information transmission has become lower, but that the ***actual*** information transmission has decreased, because of information loss (as indicated by performance errors).
2. The linear functional relation between perspective disparity and information transmission does not hold for the entire domain between 0 and 180 degrees of disparity. Surprisingly, the number of errors is largest at a disparity of 90 (and 270) degrees.

**General Conclusions:**

For now the details of this study are not of importance. I will make a few remarks from the perspective of the theory and technology of cognition testing.

1. Traditional approaches to mental testing (classical test theory, generalizability theory, item-response theory) would use a single score (e.g., the mean) to

characterize this UAV performance. Variation around the mean would be regarded as random error. Figure 2 shows that this might not be a valid assumption. The performance pattern can be fairly accurately described by a cosine function.

2. Part of the R&D for new tests will be the development of models of task performance, mathematical models that adequately represent the cognitive processes involved in task performance. Given the complexity of the aviation domain as work environment these models may be fairly complex, that is, non-additive and non-linear (see also: Ippel, 1986, 1991; Lohman and Ippel, 1993).

   I am convinced that this UAV Payload Camera task can be model for a whole range of empowering tools. The general characteristics of task performance with empowering tools are:

   - Extract information from the task environment so as to be mapped onto the required system action.
   - Decode the required system action into a user action.
   - Factors determining the workload are: (1) domain-general: the size of the sets of system actions, the set of user actions, and the complexity of mapping those two sets using the intermediary of a set of internal states. (2) domain-specific: factors involved in extracting information from the task-environment (e.g., is it a spatial task, a symbolic task, etc.).

   The mathematics to pursue the development of measurement instruments to measure the individual's capacity to deal with those tools is available.

3. Assessment of this performance took an exorbitant number of trials: 960 per individual. I expect that a great deal of sophistication have to be invested in designing more parsimonious observation designs for this new generation of measures. Bayesian approaches may provide workable solutions.

**References:**

Ippel, M.J. (1986*). Component-testing: A theory of cognitive aptitude measurement.* Amsterdam: Free University Press (119 pp). (monograph). (ISBN 90-6256-129-2 CIP)

Ippel, M.J. (1991). An information processing approach to item-equivalence. In Peter Dann, Sidney Irvine, & Janet M. Colis (Eds.). *Advances in Computer-based Human Assessment.* (pp. 377-396). Dordrecht: Kluwer Academic Publishers.

Ippel, M.J., & Watson, S.J. (October, 1998). *The UAV Camera Directory Task As A Finite-State Machine.* Paper presented at the Symposium "Getting UAV Research off the Ground" (co-chairs: M.J. Ippel and S.E. Watson) at the Human Factors and Ergonomics Society 42st Annual Meeting: Chicago (Il).

Ippel, M.J., & Hurwitz, J.B. (1998). A New Generation of Tests. *Individual Differences in Performance*, Human Factors and Ergonomics Society, 4 –6.

Lohman, D.F., & Ippel, M.J. (1993). Cognitive Diagnosis: From Statistical based Assessment toward Theory-based Assessment. In N. Frederiksen, R.J. Mislevy, & I.L. Bejar (Eds.) *Test Theory for a New Generation of Tests.* (pp. 41- 71). Hillsdale (N.J.): Lawrence Erlbaum Associates.