OPTIMAL SOLUTIONS FOR COMPLEX DESIGN PROBLEMS: USING
ISOPERFORMANCE SOFTWARE FOR HUMAN FACTORS TRADE OFFS

Robert S. Kennedy*, Marshall B. Jones**,
Dennis R. Baltzley*
*Essex Corporation
1040 Woodcock Road, Suite 227
Orlando, Florida 32803

**Milton S. Hershey Medical Center
Pennsylvania State University
Hershey, Pennsylvania 17033

ABSTRACT

A major application of isoperformance is as a trade-off methodology of the three major drivers of system design; equipment, training variables, and user characteristics. The flexibility of isoperformance allows each of these three components to be nearly any rational variation. For example, aptitude may be military Armed Forces Qualification Testing (AFQT) categories, cutoff scores within a selection procedure, or simply dichotomizing high and low scorers (pass/fail). Equipment may be new versus old, "smart" versus "dumb", high versus low resolution, etc. Training may be short versus long or varieties of media types (lecture versus CAI/CBI versus self-paced workbooks).

In its final computerized form isoperformance lets the user set an operational level of performance (e.g., a jet pilot in a simulated emergency must take prescribed corrective action and clear the plane in several seconds, pilot astronauts will check out all shuttle flight systems within 30 minutes, or Mission Specialists must handle successfully a required number of job elements). At this point the computer program guides the user through any requested trade-offs of the three components while maintaining the specified operational level of performance through "isoperformance curves." A demonstration of the computer program is currently available.

INTRODUCTION

Since the 1950's applied behavioral scientists working in the fields of systems, training, and selection have remained largely independent from each other. Within most organizations the various policy management guides and functional mission statements reinforce this separation. Historically, in systems research work, human factors practitioners have been taught that their role is to gather these human input/output data (transfer functions) and determine how they interact with their equipment (or physical and environmental stimuli). These data would then be used to generate standards and specifications which could then be used by design engineers to improve systems performance. Human factors experts also believed that design engineers were eagerly awaiting these data to incorporate into new systems which would permit efficient allocation of functions between man and machines (Fitts, 1951; Taylor, 1963). This goal was naive; attention must be given to techniques whose goals are to improve decision making in systems research by employing as a strategy the notion of "trade-off technology."

The isoperformance approach (Jones, Kennedy, Kuntz, & Baltzley, 1987) is based on the premise that differing combinations (i.e. trade-offs) of individual differences, training, and equipment variables can lead to the same desired outcome in total operational systems performance. It is called isoperformance (iso meaning same) and is a conceptual approach to systems research in human engineering. The key ingredient of isoperformance is to invert the question of operational performance enhancement by setting a desired level of performance and derive how it can be attained by different combinations of personnel, training, and equipment. The goal is that once these combinations have been determined, choices among them can be made in terms of maximum feasibilities or minimum costs. The program takes into account technology advancement and systems, personnel, and training research. It leaves an audit trail of the decision process.

Development of the Isoperformance Concept. The isoperformance concept originally surfaced out of our scientific human engineering studies and experiences on military flight simulators. The simulator experiments sought to identify which equipment features best promoted acquisition of flying skills (Lintern, Nelson, Sheppard, Westra, & Kennedy, 1981) and followed the holistic design philosophy of Simon (1976). This approach reports the size of main effects of "equipment features" like display resolution, luminance and contrast or scene

content, and permits comparison of these effects (in terms of size) with reliable individual differences like visual contrast sensitivity, dark focus, or video game performance, as well as training improvements. The data from these studies were generally reported in terms of percent of variance accounted for, for each variable, and a meta analysis of the relative contributions has been made over all the studies (Jones, Kennedy, Baltzley, & Westra, 1988) over a nine-year period of our association with the project. The present authors concluded that in these studies, when comparing aptitude, training, and equipment, reliable individual differences (aptitude) explained substantially more (usually twice as much) of the variance in the task performance than either practice or equipment. From this relationship has stemmed the isoperformance concept of trading off the three performance predictors where it appeared that individual differences (reliable differences, not error) played a large part.

In earlier conceptualizations an Omega Squared (Hays, 1977) meta analysis was attempted for a large portion of the published human factors literature in order to identify the relative contributions of individual differences versus training versus equipment features. In this work, over 10,000 citations from the body of literature in the field produced less than 0.1% which possessed and reported data in sufficient detail in order to be able to perform an adequate analysis of effect size (Jones, Kennedy, Turnage, Kuntz, & Jones, 1986). These findings presented a sobering commentary on the state of the existing literature for grounding the development of trade-off methodologies on empirical findings.

However, there are several other alternatives. Expert judgments, based on knowledge of the technical literature, can be heavily constrained to make inferences in the form of estimates (Jones, Kennedy, Kuntz, & Baltzley, 1987). Alternatively, formal experiments can be carried out and implemented under an innovative technical framework such as we (Kennedy, Jones, & Baltzley, 1988) have showed where a considerable amount of the explainable variance in task performance remained after blocking out the results according to dictates of the isoperformance methodology. Other methods also exist when empirical data is unavailable.

Structural Elements in Isoperformance. There are three main elements in the isoperformance methodology. These include individual differences, training, and equipment variations. Several synonyms are used in this report for these three dimensions.

a. Individual Differences. These differences include all of the many identifiable characteristics of people from sensory sensitivities, strength and anthropometric variables to mental capabilities and motor skills. The military, for example, employs the multidimensional Armed Forces Qualifications Test (AFQT) where anyone classified below Category 4 is not accepted (Maier & Grafton, 1980). Nevertheless, even with these restrictions in range (Sims & Hiatt, 1981), individual differences among military personnel are great. For example, in naval aviation, stringent visual examinations are used for acceptance, yet the distance at which one pilot customarily can detect opponent aircraft is sometimes 50-70% better than another, resulting in 2-3 mile advantages in early detection. Moreover, some pilots who are better at visual detection can even "outsee" the poorer ones when the latter use telescopes (Jones, 1981, personal communication). In this example, if equipment factors were evaluated to determine effects on performance in terms of the amount of accountable variance, one could not adequately assess the question without taking into account the differing visual and perceptual capabilities of the individual pilots. Cognitive and other mental capabilities also show wide variation (cf., Schoenfeldt, 1982, for a review). These relations are similarly available in industry and business although perhaps not as well documented.

b. Training. Training or practice can also be viewed under several different rubrics such as the number and length of trials, instructional systems (e.g., lecture, on-the-job, text), simulator vs. embedded training, or the type of practice regimen (massed versus distributed). In the discussions which follow, when we employ these and other denotations to report the specific outcomes, it is always our intention to connote the more general notion of the broader class of the dimension.

Specifically, a recent review of the lawful relationships from the scientific literature related to training has been completed (Lane, 1986). The sheer magnitude of the information in the report defies simple explanation. For example, learning curves vary in their shape. Tasks that are primarily conceptual may show plateaus or large gains with short amounts of practice. However, skill acquisition and procedural tasks generally show the "traditional learning curve." The shape of the learning function is such that the most rapid amount of training effect occurs initially and the best description of the overall relationship is that log performance (or practice) is a linear function of log practice (Newell & Rosenbloom, 1981). Thus, ranges of improvement in performance during formal training can be an order of magnitude of improvement for each epoch of time spent in training (cf. Hagman & Rose, 1983; Lane, 1986; Schendel, Shields, & Katz, 1978). Improvements of as much as 500% are not unusual. Such a

range of improvements can temper any expected change due to equipment factors and, because of their size, must be included in planning for technological design of the workplace.

The problem outlined above is not one which will lessen with time but rather the converse. It is believed that the problem of function allocation becomes more critical with the growing complexity and sophistication of machine systems. Considering the survey of the literature (Jones et al., 1986) it is believed a systematic methodology, such as isoperformance, can be provided to account for man/machine interface problems and present decision aids to create trade-off alternatives from the human side of the combination with no loss of operational proficiency.

c. Equipment. By equipment comparisons we may mean features that can be varied on a single piece of hardware (e.g., brightness, resolution or contrast) or several disparate engineering options (e.g., artificial intelligence versus unaided displays) or different software modifications (e.g., rate aiding, predictor display). Equipment is also a term that can encompass many of the new workplace technologies including office automation systems and computerized manufacturing systems. Under certain circumstances "equipment" could mean two models or versions of a system or it could be simulator versus actual aircraft.

The Isoperformance Methodology. Cost-effective methods may proceed in either of two general ways. The more familiar is to fix costs and maximize effectiveness. One gets, as the popular phrase puts it, "the biggest bang for the buck." The alternate procedure is to fix effectiveness and minimize health, safety, personnel, training, equipment, and manpower costs - to get "the same bang in the least costly and most expeditious way." This latter approach leads naturally to trade-offs among the cost factors and is the approach taken by isoperformance methodology (Jones et al., 1987).

The heart of this methodology is the isoperformance curve. With respect to aptitude levels and training times such a curve looks like the one given in Figure 1. The Y-axis is aptitude measured, for example, by cut-off scores. For this example let's suppose there were five categories within which incumbents or applicants could fall (two high, two low, and a middle category) on a particular aptitude test. The X-axis is training time in weeks. The job might be that of a computer operator. The curve drawn is for 80% proficient. That is, any point on the curve (any of the indicated combinations of aptitude level and training time) will produce personnel 80% of whom are proficient at the job. Thus, if one

has high-aptitude applicants (for example, in the highest cut-off categories) 80% proficient can be reached in roughly eight weeks. With lower aptitude people more training time is needed and for some aptitude levels (The lowest cut-off scores) no amount of training time up to the maximum considered will suffice to produce computer operators 80% of whom are proficient.
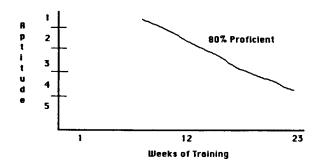

Figure 1. An isoperformance curve for 80% proficient.

Isoperformance curves come in families. A separate and distinct isoperformance curve exists for every level of performance that one specifies. Thus, if one were to specify 50% proficient, for example, one would get a different curve than the one that appears in Figure 1. Note that the second curve (Figure 2) lies to the left and down from the first curve presented. It takes less time to train the same people to the lower level of performance or, in the alternative, for the same amount of training time the lower level of proficiency can be attained with lower aptitude personnel.

A pair of curves quite similar to the pair in Figure 2 can be obtained in a quite different way. Suppose one were to automate part of the computer operator's job by providing him/her, perhaps, with more advanced computer equipment that was itself easy to use (which is done with some regularity). With the new equipment the job becomes considerably simpler so that the same objective results can now be achieved by lower aptitude people or with less training time. The situation is depicted in Figure 3. Again there are two curves, but this time the two curves correspond to two equipment variations and both represent the same level of performance. Any point on either curve suffices to produce personnel 80% of whom are proficient. Using the new equipment the same people can be trained to the same level of performance (80% proficient) in less time. Or, for a given amount of training time, the same level of performance can be achieved with lower aptitude personnel.
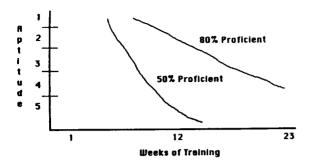
80% Proficient

50% Proficient

Weeks of Training

Figure 2. Two isoperformance curves, one for 80% and the other for 50% proficient.

80% Proficient Configuration A
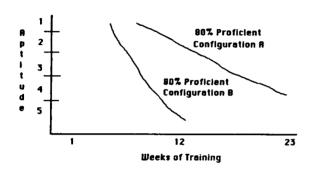
80% Proficient Configuration B

Weeks of Training

Figure 3. Two isoperformance curves, one for each of two equipment configurations, but both for the same job and the same level of performance.

80% Proficient Configuration A
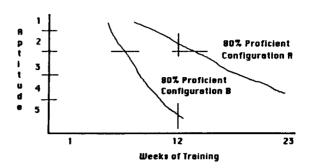
80% Proficient Configuration B

Weeks of Training

Figure 4. Figure 3 marked to indicate that category 1 and 2 people are not available and that training times in excess of 12 weeks are too expensive.

Isoperformance curves must be evaluated before any conclusion can be reached. Any point on either of the two curves in Figure 3 will produce 80% proficient personnel -- but which point is best? To answer this question one invokes other cost considerations. Category 1 and 2 people may be in such demand for other jobs that they must be regarded as unavailable. Training times in excess of 12 weeks may be excessively expensive. Figure 4 re-presents Figure 3 marked to reflect these two considerations. Since category 1 and 2 personnel are excluded by reason of unavailability, and category 3 personnel (or lower) require more than 12 weeks to reach 80% proficient using the original equipment, there is no solution to be obtained using equipment configuration A. The alternative equipment, however, does provide a range of solutions. Any point on the lower curve between the horizontal and vertical bars would be acceptable insofar as personnel availability and training costs are concerned. They might not be equivalent, however, on other counts. It might be, for example, that training schools for computer operators must extend at least eight weeks, shorter lengths of time being impractical for scheduling reasons. The solution would then have been narrowed to the second equipment configuration (B), category 3 and 4 personnel, and a training time between eight and twelve weeks.

Isoperformance is a very powerful aid to decision making and becomes more indispensable as the organization becomes more complex. Thus far, it has been explored in the environment of military systems but the implications are much broader. Moreover, because of the greater requirement to utilize and train available manpower, much of the potential power and flexibility of the model can be taken advantage of in civilian applications as outlined in a later section. The final computer program may be used not only by system design specialists, but also by executive decision makers, human resource and training specialists, as well as human factors engineers doing strategy planning.

In addition to related work like the Army's MANPRINT, the Navy's HARDMAN, and the Air Force program RAMPARTS, some of which have been referred to above, we offer as background some of the work on isoperformance. Isoperformance, in addition to referring to the computer program which is developed for the Air Force, is also a philosophy or conceptual model which structures how one addresses man-machine interactions. All of these trade-off technologies therefore have considerable features in common. What distinguishes the isoperformance work to some extent is the series of interactive software programs which are in the process of completion. The next development in that Air Force work is IsoCore. A preliminary version of this planned computer program will be available in FY88 and is described in some detail below.

IsoCore. Suppose we are given a known or designed piece of equipment and a definition of what "proficiency" means for a task to be performed using this equipment. Suppose further that if no data is available the program's user is able to estimate certain training outcomes for different categories of personnel (which outcomes will be specified shortly). Isoperformance is intended to achieve the following aims:

316

-- It forces the user to make estimates of training outcomes for different personnel categories;

- It provides checks on the internal consistency and logical coherence of these estimates;

- It provides checks on how well the estimates conform to known regularities from research in systems, human engineering, personnel, and training;

- It informs the user as to the results of these checks, together with information about what can be done to make the estimates consistent or bring them into closer conformity with known regularities and facts;

-- It outputs "isoperformance curves," that is, curves in a space defined by aptitude and training time all points of which are estimated to produce the same proportion of proficient personnel;

-- It leaves a hard-copy audit trail of all estimates, feedback, and outputted isoperformance curves.

The isoperformance core subprogram is being written in four phases: specification, input, verification, and output. These phases will be discussed in the order given.

Specification. The first phase of the core subprogram requires the user, in effect, to state the problem. The user is asked to specify:

-- the system under study
-- the task to be performed
-- what is meant by "proficient" performance
-- the personnel population to be considered
-- the aptitude dimension to be used as predictor
-- how that dimension is to be divided into ranges or "aptitude categories"
-- the training program
-- the maximum amount of training time to be considered.

With two exceptions, the main purpose of these specifications is to provide a basis for checking the user's input estimates against known training outcomes and predictive validities for similar tasks and personnel categories.

The isoperformance core subprogram provides for a single predictor dimension. This one dimension does not, however, have to be unitary in a factor-analytic or any other sense. In the usual case it will be a maximally predictive linear composite of aptitude variations related to performance on the task under consideration. The specified personnel population is divided into ranges by cut-off points on the one predictor dimension. The cut-off points themselves are not important. What matters is the proportion of the subject population that falls into each category. The maximum amount of training time to be considered must be specified because, otherwise, training time becomes indeterminate. The instructions to the user are to specify the maximal amount of time that could, taking cost considerations and other demands on instructional personnel and facilities into account, be considered feasible for training people to perform the specified task.

Input. For each aptitude category the user is required to make either two or three estimates. The first of these estimates is the amount of training time necessary for 5% of category one personnel to become proficient; the second is the proportion of persons in the category who will be proficient given the maximum amount of training time considered feasible; and the third estimate is the amount of training time necessary to make 50% of category one personnel proficient. Plainly, this third estimate is needed only if the second is greater than 50%. These two or three points define a rough skill acquisition curve. If the second estimate is greater than 50%, however, the negative acceleration of the mean performance curve is intensified once the median category member becomes proficient. At that point the distributional effect also becomes negatively accelerated making for a relatively sharp "turn" at 50% proficient.

Verification. The third phase of the ISOPERFORMANCE core subprogram checks to make sure that input estimates are "reasonable." Doing so involves three kinds of checks: formal, general, and specific. A formal check is analytic, that is, a matter of logical necessity. The estimates, for example, should increase or remain the same with decreasing aptitude category. The second kind of check (general) is for conformity with known regularities from personnel and training research. The third kind of check (specific) involves comparing the input estimates and extracted correlation coefficients with known training outcomes and predictive validities for similar tasks and personnel categories.

Output. Output consists of easily understood isoperformance curves on a graph of aptitude on the ordinate and training time on the abscissa. This background outlines the extensive groundwork which has been done formulating the isoperformance concept for military applications. This work will also serve as the basis for which to empirically ground the transition to civilian industry as a useful decision making tool for organizational and personnel cost assessment (in time and dollars).

COMMERCIAL APPLICATIONS

Isoperformance methodology has broad application in systems research for government and private industry. Five major areas of application are: (a) as a management decision aid for human factors engineering design; (b) as an adjunct to aid executives in organizing manpower, personnel, and training (MPT) applications, particularly where "what if" questions need to be answered and where an audit trail of the solution adopted is useful; (c) as a formal system for conducting trade-offs where cost analyses are conducted for existing systems; (d) as a way for industry to meet the functional specifications and requirements in an RFP; and (e) as a way for industry to be responsive to governmental contracts, especially those adhering to new guidelines on using NDI procurement strategies. A brief example is provided for each of these areas to demonstrate the utility of isoperformance methodology.

Isoperformance kinds of estimates are already required in the military in the form of MANPRINT analyses. The data available from the MANPRINT requirements for systems will work well as data for explicit trade-offs in isoperformance analyses. From these current data "ground-up" HFE design work could be pursued with maximally efficient systems as a result.

Within the MPT arena isoperformance methodology permits trade-offs for each component and provides immediate feedback for forcasting efficiency and selection/placement. The current IsoDemo program developed for the Air Force (Jones & Jones, & Essex Corporation, 1987) provides a constrained example of using the isoperformance methodology. At the end of the program the manager can tell what the lowest aptitude category is within the training time allotted and equipment constraints available. Conversely, he/she can also find the minimum training time necessary if the very best people were available which is seldom the case in the military; however, this will differ for private industry selection.

The third major area of isoperformance has the broadest application. This area is using isoperformance methodology for existing systems. Isoperformance can be used to evaluate and suggest improvements in any system where there is a man/machine interaction or the various costs of the different parts can be compared. This is especially useful with emerging technologies. In private industry technology changes weekly, isoperformance allows the decision-maker to evaluate each potential upgrade or changeover from a complete systems viewpoint and to make better informed choices from a organizational cost/benefit perspective.

Finally, industry may use isoperformance methodology to meet the functional specifications and requirements in an RFP or simply to be responsive to a customer's needs. Suppose the government or another large organization calls for updating or replacing an in-place piece of equipment. A company may propose to modify the system by upgrading it to make it "state-of-the-art," or it can trade off the complexity through longer training time or selection of higher aptitude personnel. The company may propose to replace the equipment with a less complex system with no development cost associated. In this way the company cannot only lower the unit cost but could provide isoperformance verification for shorter training time and broader use of the labor pool. This would result in substantial lowering of total system costs in training, personnel, and support. The benefits are obvious; the company may elect to pursue a technological advantage or an overall cost advantage. Both are defensible and may be suggested to a manager for overall preference. If the system is a simulator, state-of-the-art may be required. If it is a vehicle, an overall cost approach may be chosen. The Army, for example, adapted the Chevy Blazer to meet their light truck requirements.

As a computerized decision aid in design, the isoperformance program may be used to trade-off the aptitude, equipment, and training dimensions which are known or can be estimated for a prospective system. In this way overall utility as well as cost/benefit considerations may be assessed. For example, in a new weapons system, the projected manpower of the target service as well as the allowable minimum and maximum training times may be reasonably estimated. This will form a "window" within which the equipment (man/machine interface) must stay. Many questions about which elements to emphasize can be answered almost immediately by framing the question within the context of the isoperformance model.

Additionally, as a fifth point, the isoperformance approach is well suited for application of the recent policy mandating the use of Non-Developmental Items (NDI) in the military acquisition process. This NDI procurement plan is a direct result of the President's Council on Defense Acquisition, The Packard Commission. Governmental agencies are required to evaluate the ability of an "off-the-shelf" item for satisfying their functional needs. An NDI may be entirely off-the-shelf needing no development, or the item may require a dedicated R&D effort by the contractor to modify the item for current governmental needs. A major principle in NDI acquisition is that less than full compliance with a programs performance objectives is insufficient reason not to use NDI. In other words, if an NDI does not meet all specifications and requirements set forth in the Request for Proposal (RFP), it is not

318

disqualified; cost/benefit trade offs can be made. Here lies the isoperformance strong point. Industry which deals in government contracting may invoke these NDI concerns and use isoperformance to trade off any weaknesses in their "off-the-shelf" products to maintain a more flexible position in competition.

REFERENCES

Fitts, P. M. (Ed.). (1951). Human engineering for an effective air investigation and traffic control system. Washington, DC: National Research Council.

Hagman, J. D., & Rose, A. M. (1983). Retention of military tasks: A review. Human Factors, 25, 199-213.

Hays, W. (1977). Statistics for psychologists. New York, NY: Holt, Rinehart, & Winston.

Jones & Jones, & Essex Corporation. (1987). A brief introduction to Isoperformance methodology [Computer program]. Orlando, FL: Essex Corporation. (Armstrong Aerospace Medical Research Lab. Contract No. F33615-86-C-0553)

Jones, M. B., Kennedy, R. S., Baltzley, D. R., & Westra, D. P. (1988, in preparation). Omega square analyses and reevaluation of a meta-analysis for equipment features and individual differences in an aircraft simulator. Orlando, FL: Essex Corporation.

Jones, M. B., Kennedy, R. S., Kuntz, L. A., & Baltzley, D. R. (1987). Isoperformance: Trading off selection, training, and equipment variations to maintain the same level of systems performance. Proceedings of the 31st Annual Meeting of the Human Factors Society (p. 634-631). Santa Monica, CA: Human Factors Society.

Jones, M. B., Kennedy, R. S., Turnage, J. J., Kuntz, L. A., & Jones, S. A. (1986). Meta-anlysis of human factors engineering studies comparing individual differences, practice effects, and equipment design variations (Final Rep. No. F33615-85-C-0539). Wright-Patterson AFB, OH: Department of the Air Force, Air Force Systems Command, Aeronautical Systems Divisions. (NTIS No. AD A168 790)

Kennedy, R. S., Jones, M. B., & Baltzley, D. R. (1988). Empirical demonstration of Isoperformance methodology preparatory to development of an interactive expert computerized decision aid (Final Report for ARI). Washington: Army Research Institute.

Lane, N. E. (1986). Skill acquisition curves: An analysis of applicability to the time course and sequencing of military training (Final Rep., Contract No. 5540, Institute for Defense Analysis). Orlando, FL: Essex Corporation.

Lintern, G., Nelson, B. E., Sheppard, D. J., Westra, D. P., & Kennedy, R. S. (1981). Visual Technology Research Simulator (VTRS) human performance research. Phase III (NAVTRAEQUIPCEN 78-C-0060-11). Westlake Village, CA: Canyon Research Group, Inc.

Maier, M. H., & Grafton, F. C. (1980). Renorming ASVAB 6/7 at Armed Forces examining and entrance stations (Technical Memorandum 80-1). Washington, DC: Office of the Assistant Secretary of Defense.

Newell, A., & Rosenbloom, P. S. (1981). Mechanisms of skill acquisition and the law of practice. In J. R. Anderson (Ed.). Cognitive skills and their acquisition. Hillsdale, NJ: Erlbaum.

Schendel, J. D., Shields, J. L., & Katz, M. S. (1978). Retention of motor skills: Review (Technical Paper 313). Alexandria, VA: Army Research Institute.

Schoenfeldt, L. F. (1982). Intra-individual variation and human performance. In M. D. Dunnette and E. A. Fleishman (Eds.), Human performance and productivity: Human capability assessment. (Vol. 1, pp. 107-135). Hillsdale, NJ: Lawrence Erlbaum.

Simon, C. W. (1976). Analysis of human factors research engineering experiments (TR-CWS-02-76). Westlake Village, CA: Canyon Research Group.

Sims, W., & Hiatt, C. M. (1981). Validation of the Armed Services Vocational Aptitude Battery (ASVAB) forms 6 and 7 with applications to ASVAB forms 8, 9, and 10. Alexandria, VA: Center for Naval Analyses. (NTIS No. AD A110 025)

Taylor, F. V. (1963). Human engineering and psychology. In S. Koch (Ed.), Psychology: A study of science (Vol. 5, pp. 831-907). New York: McGraw-Hill.