Operator Loading Tasks

W. B. KNOWLES, Hughes Aircraft Company, Culver City, California

The purpose of this paper is: (1) To review the rationale of measuring operator workload in terms of auxiliary, or secondary task performance scores; (2) To summarize the important characteristics of suitable loading tasks; (3) To describe several loading tasks which have been used or which are potentially useful; and (4) To suggest the development of a set of standardized tasks which would be useful in obtaining more nearly comparable measures over a wide range of primary tasks.

INTRODUCTION

In the design of equipment and the development of operating and training procedures it is important to be able to answer questions such as:

How easy is this equipment to operate?
How much attention is required?
How much learning is involved?
How well will the operator be able to perform additional tasks?

All these questions deal with some aspect of what has come to be called operator workload. Essentially, how busy is the operator?

In some instances operator workload can be assessed with sufficient precision by means of task-analysis and related methodologies (Ekstrom, 1962). Siegel and his associates (Siegel and Wolf, 1961) have also developed digital-computer techniques for predicting the effect of different levels of task-induced stress on operator performance that are applicable in situations where sufficient information is available on the detailed structure of the task. But in many cases operator workloads must be assessed empirically and at a stage of equipment development where only part-task simulation techniques are possible.

In these latter situations it has been found that one of the best ways of measuring operator-load is to have the operator perform an auxiliary or secondary task at the same time he is performing the primary task under evaluation. If he is able to perform well on the secondary task, this is taken to indicate that the primary task is relatively easy; if he is unable to perform the secondary task and at the same time maintain his primary-task performance, this is taken to indicate that the primary task is more demanding. Just how easy or how demanding is expressed in terms of loading scores derived from the secondary-task performance scores.

In addition to giving an overall indication of operator loads, secondary-task scores can also be used for other purposes. When properly designed, secondary tasks can be used to indicate how the operator's load varies during the course of a single performance of the primary task. Secondary-task measures can often be used to evaluate the course of learning during primary-task acquisition, since very often primary-task scores will show little appreciable change in performance level while auxiliary-task scores may show progressive improvement indicating better and better mastery of the primary task. Secondary tasks introduced into part-task simulator trainers could also be used to measure the fidelity of the workload aspects of the simulation.

This basic idea of using secondary-task measures to reflect primary-task performance has been used in one or another of the above applications by a number of investigators, but unfortunately the use of different tasks and different measures make it difficult to compare their results directly. These results do indicate, however, that it should be possible to develop a small set of tasks which would be suitable for measuring operator-load with a wide range of primary tasks. Standardization of such a set of tasks would serve to define the concept of operator-load in more operational terms and would provide measures which would be more directly comparable from study to study.

In the following discussion the rationale underlying the use of secondary tasks will be examined in some detail with the view toward describing the characteristics desired in tasks suitable for measuring operator-load, and specifying the research required to develop this powerful and much-needed tool.

RATIONALE

Secondary or loading tasks are used in part-task simulation studies for two related but basically different reasons, and it is important to differentiate between these uses and the reasons behind them in order to focus clearly on the problem of interest here—the mesaurement of operator-load.

In the first place, the results of part-task simulations are often deficient in that the performance appears unduly good because the operator is permitted to focus all his attention on the part-task and is not required to divide it among several tasks as he would in the total job situation. The first use of secondary tasks is to compensate for this deficiency and to simulate aspects of the total job that may be missing. Ordinarily there is little interest in the secondary-task performance *per se*. The secondary task is used simply to bring pressure on the primary task with the idea that as the operator becomes more heavily stressed his performance on difficult tasks will deteriorate more than will his performance on easy tasks.

Garvey and Taylor (1959) demonstrated this use of secondary tasks by carefully designing two tracking systems which could be operated equally well, as measured by tracking-error scores, but which required different degrees of effort, as indicated by operator-opinion reports. Under the stress of any of a number of loading tasks, e.g., mental arithmetic, tracking, simulated warning-light monitoring, etc., performance decrements for the difficult system were greater than for the easier system.

In this first application of secondary tasks, then, the emphasis is upon stressing the primary task; differences in operator-workload are indicated by differences in primary-task performance measures taken under the stress induced by the auxiliary task. The major methodological difficulty with this use of secondary tasks is that it is impossible to specify or control the degree of stress induced by the secondary task. A second difficulty is that severe stress may disrupt the primary task to the point that the operator's mode of behaviour changes completely.

In the second application of loading tasks, and the one of primary concern here, the auxiliary task is used not so much with the intention of stressing the primary task as with the intention of finding out how much additional work the operator can undertake while still performing the primary task to meet system criteria. Secondary tasks are used because primary part-task performance measures, in and of themselves, seldom reflect operator-load. They usually tell how well some functional system criterion is met, e.g., tracking error, miss-distance, decision accuracy, etc., but they seldom tell the price paid in operator-effort in meeting this criterion.

For example, Ekstrom (1962) measured operator-work-loads during an evaluation of two alternative control modes for the X-15 by having the operators perform an additional self-paced pushbutton task. The pushbutton scores were then converted into an operator-loading index to demonstrate differences in primary-task difficulty. The results of the evaluation showed that, in terms of completing the mission successfully, the pilot could use both modes equally well, but that the more automatic system demanded less of the pilot's attention, particularly during a number of highly critical transition periods.

Knowles and Rose (1962) used a similar self-paced pushbutton task to demonstrate differences in operator-load during simulated lunar landings. The loading scores were sensitive to differences in problem difficulty; they reflected increased ease in handling the control task as a function of practice; they revealed differences in work-load between members of a two-man crew; and they showed that the particular control law under consideration was unsatisfactory because of the extreme buildup of operator-load during the last few seconds of the landing. None of these results was available from system performance criteria, i.e., time, fuel, miss-distances.

Birmingham, Chernikoff and Siegler (1962) have used a self-adaptive tracking system to measure what they call "residual operator bandwidth." Bauerschmidt and Besco (1962) used a forced-pace pushbutton task to evaluate differences in attitude control system configurations but found that the task was insensitive when the rate of presentation was too low. Garvey, who used a mental arithmetic task to stress performance in a display-control compatibility study (Garvey and Knowles, 1954), has suggested that a similar task could be instrumented for use in measuring operator-load.

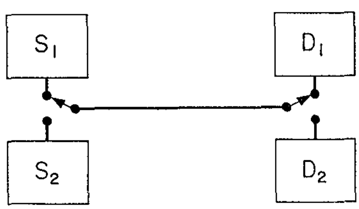

Fig. 1. Simple multiplex model. S_1–D_1 represents the flow of primary-task information S_2–D_2 represents the flow of loading-task information.

In all these examples of the second application of secondary tasks, the actual measurement of operation load, the goal has been to use a task which is subtle in its effect on the primary task; one which is demanding enough so that the operator cannot ignore it, yet not so demanding as to stress the primary task to the point of disruption. In fact, the degree of non-performance on the loading task is taken as the measure of operator-loading during what is essentially normal performance of the primary task.

The basic notions of operator-loading and its measurement by loading-task scores can be summarized by comparing the operator to a multiplex communication system, as shown in Fig. 1.

A multiplex system uses a single channel to transmit messages from several sources to several destinations. So long as the channel is connected to a given source and a given destination, messages from other sources to the same or other destinations cannot be transmitted. The basic channel has a fixed capacity, but within this capacity overall rate of information flow can be maximized by proper coding and switching routines.

In measuring operator-load the primary task, S_1–D_1, represents one set of information to be processed and the loading task, S_2–D_2, represents another set. It is assumed that the system is so structured that priority is given to primary-task information and only as much loading-task information is processed as is possible within the residual channel capacity.

Two or three interesting and important points can be illustrated by reference to this relatively simple model.

First: The switching points at the input and output represent opportunities for interference between the two tasks that must be minimized in designing the loading task and matching it with the primary task. The input switching shown in this model represents essentially the same concept as the gate in Broadbent's (1957) model illustrating division of attention. Priority is assured to primary task inputs by instruction, by sense modality selection, by apparatus configuration, and by controlling signal intensities and timing, etc. Response interference on the output end is usually easily prevented by using different effectors for the two tasks. Furthermore, most tasks of interest in part-task simulations are limited on the perceptual side, not on the response end.

Second: Mastery of most perceptual-motor skills can be thought of in terms of the adoption on the part of the operator of more efficient methods of coding information. As relevant properties of the input signals are selected and combined in terms of the responses and response sequences to be made, i.e. as the coding becomes more efficient, the rate of transmitting information increases. This means that if the rate at which messages are produced is high, more messages per unit time can be transmitted with a given channel capacity; or, if the rate of message production is low, relative to the capacity of the channel, the channel will be open a greater proportion of the time.

This, then, is one way of explaining how loading-task scores can reveal primary task learning in the absence of improvement in system criterion measures. If it is assumed that system performance depends upon the transmission of a certain amount of information at a rate well within the channel capacity, the requirement may be met with very inefficient coding methods. System performance, being dependent simply upon the *effective* transmission of the information, may not show any significant change as the transmission becomes more *efficient*. *But loading-task scores should change* since the more efficient transmission of primary-task messages leaves the channel free to transmit more loading-task information.

Third: Since the loading-task information must be "sandwiched in," so to speak, whenever there is room for it in the channel, this implies that the loading task message units should be short and efficiently coded. The use of discrete stimulus-response units is probably desirable in meeting the first requirement. Efficient coding is obtained by careful attention to factors such as stimulus-responses compatibility and population stereotypes, and, as discussed above, through learning.

In other words, the loading task, if it is to be sensitive to—and reflect accurately—changes in primary-task information-transmission, should be easy, over-learned, and probably composed of discrete message units.

CHARACTERISTICS

In the preceding discussions, allusions have been made to a number of properties that are desirable in secondary tasks used as measures of primary task work-loads. The following list summarizes the more important of these characteristics:

Non-interference

Ideally, the loading task should not physically interfere with, nor otherwise disrupt, primary-task performance. Practically, this is difficult to achieve; some degree of interaction between two tasks is always to be expected. Empirically, it can be determined if the loading-task is too stressful by observing the effect it has on primary-task performance criteria.

Simplicity

Ideally, the task should require very little *learning and should show little inter-subject* variability. Again, although difficult to achieve, both of these properties can be evaluated empirically. Furthermore, learning effects can be minimized by practice on the loading-task alone, and subject differences can be controlled by determining the base-line performance for each subject and then computing the loading index for each individual relative to his own base-line.

Self-pacing

In general, it is preferable that loading-task information be presented at a rate determined by the operator himself. This can be achieved by using a self-paced series of discrete units or by automatic feedback systems which are self-adapting as a function of operator performance. Self-pacing also tends to adjust the attention-demanding properties of the loading task automatically so that, when necessary, the task can *be neglected but not ignored.*

Scoring

The index of operator-load that is calculated from the scores of a given loading task should be comparable from situation to situation. The score obtained on the auxiliary task ordinarily will reflect the average rate at which loading-task information is handled over an entire run. In many applications it is also desirable to be able to evaluate the instantaneous rate of flow of information continuously throughout the run. In either case, it appears that automatic recording of loading-task performance is almost a necessity.

Compatibility

The loading task should be selected so that, if possible, it is different in form from the primary task and, again if possible, so that it will simulate the kinds of tasks that might be required of the operator in addition to the primary task. For example, if the primary task is a tracking task it probably would be unwise to use another tracking task as an auxiliary task because of possible intertask interference. A pushbutton or mental-arithmetic task would probably be less interfering and could be considered to simulate other monitoring or problem-solving functions.

USEFUL TASKS

Several tasks which appear capable of meeting the above criteria, and which warrant further investigation, are the following:

Pushbutton Responding

This is the task used so successfully by Ekstrom (1962) and probably represents the standard against which all other loading-tasks might well be compared.

A matrix of sixteen touch-lights was located outside the operator's central field of view as he performed the primary task.

The expression given for the primary task work load is:

$$W = 100-(W_1+W_t) \quad (1)$$

where

W = primary task load

W_1 = loading task work load

W_t = eye transition time work load.

W_1 was computed from

$$W_1 = \frac{N}{N\,\text{max}} \times 100 \quad (2)$$

N = number of lights handled/sec

N max = calibrated maximum number of lights/sec.

W_t was computed from

$$W_t = \frac{\text{Transitions}}{\text{total time}} \times 100 \times 0{\cdot}14 \quad (3)$$

where 0.14 = average transition time

The index was interpreted as the "per cent' of the pilot's effort or attention that was devoted to the control-task.

Several variations in the basic pushbutton task can also be considered. Separating the display and the control offers the opportunity of manipulating the difficulty of the task by varying display-control compatibility relationships. Verbal responses to the lights is also another possibility which, in some situations, may overcome response interference problems.

Mental Arithmetic

In the arithmetic task used by Garvey and Knowles (1954), the subjects were given two digits to add or subtract. When the subject reported his answer he was given another digit to either add or subtract from the previous sum or difference, and so on through the trial. The task is attractive because, unlike the pushbutton task, it is primarily a verbal task. The use of auditory inputs and spoken responses also avoids the possibility of direct response interference with primarily motor tasks. In this form the presentation of stimuli and the recording of responses are somewhat difficult to control. However, the task can be modified and instrumented so that the digits are presented on light panels and the responses keyed out.

Self-adaptive Tracking

Birmingham *et al.* (1962) have experimented with tracking systems wherein the level of difficulty is controlled by the operator's level of performance. Difficulty may be varied in a number of ways: by varying the complexity of the input signal, by adjusting the effective order of the control system through variations in "quickening" terms, or by adjusting the upper and lower bounds in a nulling task. Regardless of the exact method of varying difficulty, the essential feature of the task is contained in the self-adjusting feature. A predetermined level of performance is established and, whenever the operator exceeds this level, the task is automatically made more difficult; whenever he falls below the prescribed level, the task is automatically made easier. Ultimately the operator stabilizes out at a particular level of task-difficulty. On the assumption that the stabilization level is a function of the channel-capacity consumed by the primary task, Birmingham then calculated an index which he calls a measure of the "residual operator bandwidth." The index is based upon the frequency-amplitude characteristics of the input, the error characteristics of the output, and certain assumptions concerning the characteristics of the human operator regarded as a servo system.

It would seem desirable to have a tracking task to use in conjuction with some kinds of primary tasks, but the best way of instrumenting and scoring it remains to be worked out.

Monitoring

Holland (1958), in a vigilance study, used a meter-monitoring task that might be adapted as a loading task. In this task a meter pointer was driven by a slowly varying signal and the subject was asked to report when the reading exceeded a set of prescribed limits. Monitoring responses were recorded by having the subject press a button to illuminate the meter-face. This kind of "peek and report" task might be very useful. Another possibility might be simply to use several meters which the operator is asked to read as often as he can in a manner similar to scanning an instrument panel. The number of "peeks" or the number of readings could be converted into an index of work-load.

There are undoubtedly several other tasks which also have the proper characteristics for loading tasks. However, these four cover a wide range and one or another should be useful in nearly all simulation situations. With refinement and standardization they would provide a much-needed tool for measuring operator-load.

A SUGGESTED APPROACH

The work required to develop a set of standard loading tasks divide into three phases: selection, evaluation, and standardization.

Selection

A more complete listing of potential loading tasks than has been given here should be assembled with the goal being to find a number of simple tasks, which meet the criteria previously outlined, and which cover a range of types, i.e., verbal, motor, auditory, monitoring, tracking, etc. From the list, a few tasks, three or four, should then be selected for further refinement.

Evaluation

Alternative ways of instrumenting these tasks should be investigated thoroughly. It is of extremely great practical importance that the equipment and scoring problems be worked out well and kept simple. The selected tasks should be used in conjunction with primary tasks of different types and levels of apparent difficulty to evaluate their sensitivity as measuring instruments. Learning and subject variances should be evaluated. The end result of this phase would be the detailed specifications for instrumenting, administering, calibrating, and scoring a small but comprehensive set of loading tasks.

Standardization

The usefulness of the set of tasks would be enhanced if, ultimately, scores on all the tasks could be reduced to a common index of operator work-load, i.e., if the index computed for each task were comparable to the index computed from every other task. There are, of course, severe methodological difficulties to be overcome in achieving this.

First, in a factor analysis sense, the different tasks tap different abilities. This is the basic reason for using different tasks; this is also the basis for subject and learning variances.

Second, although it is assumed that the loading tasks do not interfere with the primary task, it is recognized from the start that, practically, this assumption can only be approximated and, furthermore, that most probably the amount of interference between a loading task and a primary task is a function of the factorial structure of the two tasks. Again, one of the reasons for using different loading tasks is to avoid as much interaction as possible. Presumably, if the structure of the primary and auxiliary tasks were known, an optimum match could be made. But, since the structure of the tasks will probably never be known, it will always be a matter of judgment as to which loading task to use with which primary task.

In the face of these considerations, it is well to look more closely at what may be expected of any measure of operator work-load derived from auxiliary task performance. Fundamentally, such measures yield an ordinal scale; 100 per cent auxiliary task performance does not mean zero operator loading, nor does zero auxiliary task performance mean 100 per cent operator loading. Furthermore, equal increments in loading scores most certainly do not reflect equal increments in work-load. It is therefore most prudent to regard whatever numerical values that are derived with some modesty and to call them what they are—simply indices of operator-load.

The complete standardization of several loading tasks would be an endeavor comparable in magnutide to the cross-cultural standardization of alternative forms of the Stanford–Binet Intelligence Tests, and is certainly not warranted at this time. The immediate problems are (1) to find a few tasks that can be instrumented to give reliable measures, and (2) to demonstrate that these tasks do indeed descriminate between different primary tasks of different degrees of difficulty.

With these tools in hand it should be possible (1) to obtain more meaningful measures of operator-load in the practical context of simulation studies and (2) proceed with such theoretical issues as analysis of the factors which affect operator-load and (3) to investigate the more general problem of task interference.

REFERENCES

Bauerschmidt, D. K., & Besco, R. O. *Human Engineering criteria for manned space flight: minimal manual systems.* AMRL-TDR62-87, Wright Patterson Air Force Base, Dayton, Ohio, October 1962.

Birmingham, H. P., Chernikoff, R., & Ziegler, P. N. *The design and use of "equalization" teaching machines.* Paper presented at IRE, International Congress on Human Factors Engineering in Electronics, Long Beach, California, May 3–4, 1962.

Broadbent, D. E. A mechanical model for human attention and immediate memory. *Psychol. Rev.,* 1957, **64,** 205–215.

Ekstrom, Phyllis J. *Analysis of pilot work loads in flight control systems with different degrees of automation.* Paper presented at the IRE International Congress on Human Factors Engineering in Electronics, Long Beach, California, May 3–4, 1962.

Garvey, W. D., & Taylor, F. V. Interaction among operator variables, system dynamics and task-induced stress. *J. Appl. Psychol.,* 1959, **43,** 79–85.

Garvey, W. D., & Knowles, W. B. Response time pattern associated with various display-control relationships. *J. Exp. Phychol.,* 1954, **47,** 315–322.

Holland, J. G. Human vigilance. *Science,* 1958, **28,** 61–63.

Knowles, W. B., & Rose, D. J. *Manned lunar landing simulation.* Paper presented at IEE, 1963 National Winter Convention on Military Electronics, Los Angeles, California, January 30–February 1, 1963.

Siegel, A. I., & Wolf, J. J. A technique for evaluating man-machine system designs. *Human Factors,* 1961, **3,** 18–27.