



Society of Automotive Engineers, Inc.
TWO PENNSYLVANIA PLAZA, NEW YORK, N. Y. 10001

AEROSPACE INFORMATION REPORT

AIR 1054A

Issued 12-1-68
Revised 12-13-72

THE MEASUREMENT OF TRAINEE PERFORMANCE IN SIMULATORS AND PART-TASK TRAINERS

1. INTRODUCTION

1.1 Increasing attention is being given to measuring trainee performance in synthetic training devices, as evidenced by the recent Air Force, NTDC, and NASA-sponsored efforts in this area cited in the bibliography. A fair summary of well-informed opinions in this area would be:

- a) Such measurement can be most useful (see par 2).
- b) Subjective measures, such as the traditional instructor ratings, do not meet the current requirements for reliable, valid, and practical measurement.
- c) Synthetic training equipment, especially of the digital variety, makes feasible the achievement of needed measures more expeditiously than in operational equipment.

1.2 This report intends to acquaint the non-technical reader with the potentialities and limitations of measuring trainee performance in synthetic training devices, and outline the steps needed to achieve these measures. As an introduction, the report intends to be limited in depth; for those desiring to explore the subject in greater depth, publications cited in the Bibliography should be consulted. Much of the information cited herein is derived from the reports listed in the Bibliography (par. 10); the report by Smode et al. was especially useful.

2. IMPORTANCE AND PURPOSES OF MEASUREMENT

Performance measurement can serve a wide variety of functions. The optimal kind of measurement depends on the functions to be served, but most measures should be relevant for a number of different functions. The specific purposes which can be served by measurement of trainee performance can be stated as follows:

- a) Feedback for Training - It has been well established that furnishing trainees with knowledge of the results of their efforts enhances their learning. Feedback is of most value when it is prompt, accurate, and relevant.
- b) Trainee Motivation - When objective standards of performance exist, each individual has an immediate and concrete goal to strive for, and also meaningful and healthy competition between individuals or crews is possible.
- c) Prediction of Future Success - Measurement may involve the collection of data from which can be esti-

mated (preferably with a specified success probability) how an individual or team will perform in some future context or universe of events. This prediction of future performance may be an evaluation of aptitude for training or an evaluation which uses measures of training achievement as the basis for prediction of subsequent operational performance.

d) Evaluation of Present Performance - Measurement may involve collecting data which can specify the knowledge, skill level, or performance level of an individual or team. The measures may reflect present performance level in both part-tasks and/or more comprehensive segments.

e) Evaluation of Learning Rate - Measurement data may be collected at several points in a training program in order to indicate the rate at which knowledge and skills are being acquired. Such measures provide a basis for judging an individual's or a crew's present stage of learning and readiness for the next phase in a training program.

f) Identification of Areas of Proficiency and Deficiency - Measurement data, particularly of a diagnostic nature, may be used to determine in what areas or tasks either individual crew members or an entire crew are proficient and in what areas they are deficient. Such measures pinpoint the need for and nature of further training and suggest task or environmental modifications to achieve a specified level of proficiency. In addition they can provide information to the trainee which will speed his learning.

g) Evaluation of Training Effectiveness - Measurement data may be used to determine the nature and extent of changes resulting from a training experience. Another important area subsumed here relates to training research and the evaluation of the differential effectiveness of alternative training methods, the contribution of component proficiencies to overall mission accomplishment, schedules, simulators and simulator features.

h) Selection and Placement of Individuals and Teams - Measurement data may assist in identifying individuals more likely to achieve a given level of proficiency: i.e., the identification of persons who either will require less training or will profit the most from a given amount of training. Included here are the assignment of individuals to crew positions and the assignment of crews to special missions.

SAE Technical Board rules provide that: "All technical reports, including standards approved and practices recommended, are advisory only. Their use by anyone engaged in industry or trade is entirely voluntary. There is no agreement to adhere to any SAE standard or recommended practice, and no commitment to conform to or be guided by any technical report. In formulating and approving technical reports, the Board and its Committees will not investigate or consider patents which may apply to the subject matter. Prospective users of the report are responsible for protecting themselves against liability for infringement of patents."

i) Refinement of Criterion Information - Measure-
ment data may provide the basis for refinement of the
criterion by helping to define further what constitutes
successful or proficient performance.

j) Definition of Requirements - Measures of per-
formance may permit statements of functional require-
ments for training equipment to be specified more
precisely. Training standards can be made more pre-
cise and objective, and training equipment more
effective.

k) Evaluation of Equipment and Procedures - This
can include determination of whether given equipment or
procedures permit attainment of required standards, as
well as determination of the better of two equipments or
procedures. For example, total training time to a pro-
ficiency criterion can be used as a measure of effective-
ness of a specific equipment or procedure. Items to be
evaluated include instruments, control feel, personal
equipment (e. g., seating), and task sharing between
captain and first officer.

l) Evaluation of Instructor Capability - The best
measure of an instructor is the kind of student he turns
out. Reliable and valid measures of trainee performance
permit instructors to be evaluated directly by their prod-
uct, rather than indirectly by their knowledge of the
subject and teaching manner.

3. CLASSIFICATION OF PERFORMANCE VARIABLES

3.1 There are seven basic classifications of meas-
ures useful in evaluating trainee performance:

- 1) Time: Measures dealing with time periods in
production of performance.
- 2) Accuracy: Measures dealing with the correctness
and adequacy of production of performance.
- 3) Frequency of occurrence: Measures dealing with
the rate of repetition of behavior.
- 4) Amount achieved or accomplished: Measures
dealing with the amount of output or accomplishment in
performance.
- 5) Consumption or quantity used: Measures dealing
with resources expended in performance in terms of stand-
ard references.
- 6) Behavior classification by observers: Measures
dealing with classifying more complex behaviors into
operationally defined subjective categories. Observa-
tions are placed into discrete classes on a continuum for
the event observed.
- 7) Condition or state of the individual in relation to
the task: Measures dealing directly with the state of the
individual which describe behavior and/or results of acts
that have occurred.

3.2 These classes of measures are graded on a
quantative-qualitative continuum, with precise quan-
tities (time, accuracy, frequency) at one end and more

qualitative interpretations (categorization, descriptive
reports) at the other. Each class or group includes a
variety of subgroups and specific measures. These are
listed in detail in Table 1.

3.3 As in other areas of psychometrics, the more
objective, easier-to-obtain measures usually reflect but
a single facet of the trainee's performance; more global
measures tend to be unreliable, difficult to obtain, or
both. In general, combinations of discrete scores are
required; the task of combining separate scores into a
useful overall score is often a demanding one. Since a
given overall score can be achieved in a variety of ways,
it is usually necessary to utilize sub-scores as well as the
overall score in interpreting trainee performance.

Table 1

A Classification of Measures

TIME

1. Time to Initiate an Activity from the Onset of
a Signal or Related Events
Time to perceive event
Reaction time
Time to initiate a correction
Time to initiate a subsequent activity (fol-
lowing completion of a prior activity)
Time to initiate a course of action
Time to detect trend of multiple related
events
2. Time to Complete an Initiated Activity
Time to acquire, to lock-on, to identify
Time to complete single message
Time to complete a computational problem
Time to make an adjustment/manipulation/
control positioning
Time to reach a criterion
3. Overall Time from Signal Onset to Activity
Completion
Percent time-on-target
Time spent in an activity (communicating,
repairing, computing, etc.)
Time to complete a sequence of activities
Build-up of time (cue length)
4. Distribution of Part Task Times in Completing
an Activity
Time-sharing among events

ACCURACY

1. Correctness of Observation or Perception
(Discrete/sequential)
Accuracy in identifying display readout
Accuracy in identifying extra-cockpit
objects (environment, ground terrain,
celestial navigation objects)
Accuracy in estimating distance, direction,
speed

