EVALUATION OF COMPUTER SOFTWARE IN AN OPERATIONAL ENVIRONMENT (U)

UNCLASSIFIED

OEG-PP-290

END

DATE FILED: 12-80

DTI:
EVALUATION OF COMPUTER SOFTWARE IN AN OPERATIONAL ENVIRONMENT

Gregory H. Watson
EVALUATION OF COMPUTER SOFTWARE IN AN OPERATIONAL ENVIRONMENT.

Gregory H. Watson

This Professional Paper was delivered at the joint national meeting of The Institute for Management Science (TIMS) and the Operations Research Society of America (ORSA) in Washington, D.C. on May 5, 1980.
ABSTRACT

This paper examines general procedures for testing military real-time operational software from the user's perspective. A summary of industrial software testing is given with an evaluation of its applicability to the military's requirement for operational testing. The operational test environment is examined to determine the extent of verification, validation or certification of computer software that is possible given the constraints of this environment.
SLIDE 1

Good afternoon, this presentation is a discussion of the steps which will be necessary to implement the software testing requirements of Department of Defense Directive 5000.3 of December 26, 1979. Specifically, the requirements for operational test and evaluation of computer software will be the point of focus. Since many of you may not be acquainted with DOD's test and evaluation procedures, a brief review may help.

SLIDE 2

There are two types of testing which follow a system through its acquisition cycle: developmental test and evaluation and operational test and evaluation. These are commonly referred to by their acronyms -- DT&E and OT&E. DT&E is technical testing which is performed to ensure that a system meets its design specification. DT&E includes the evaluation of hardware/software integration, system software and compatibility/interoperability with existing and planned equipment and systems. DT&E differs from OT&E in at least two other areas: DT&E is conducted by engineers in a controlled environment such as a laboratory or instrumented test range. OT&E is performed to determine if the system meets the required operational characteristics which are specified in the Test and Evaluation Master Plan (TEMP). There are two major areas of operational testing: operational effectiveness and operational suitability. The effectiveness tests are concerned with how well the system performs while the suitability tests are
EVALUATION OF COMPUTER SOFTWARE
IN AN OPERATIONAL ENVIRONMENT

GREGORY H. WATSON

CENTER FOR NAVAL ANALYSES

SLIDE 1
TYPES OF TESTING

- DEVELOPMENTAL TEST AND EVALUATION (DT&E)
  - VERIFY ATTAINMENT OF TECHNICAL PERFORMANCE SPECIFICATIONS
  - ASSIST ENGINEERING DESIGN AND DEVELOPMENT PROCESSES
  - CONDUCTED PRIMARILY IN LABORATORIES OR CALIBRATED RANGES

- OPERATIONAL TEST AND EVALUATION (OT&E)
  - ESTIMATE A SYSTEM’S OPERATIONAL EFFECTIVENESS AND SUITABILITY
  - EFFECTIVENESS TESTS: HOW WELL DOES IT PERFORM?
  - SUITABILITY TESTS: IS THE SYSTEM SUPPORTABLE?
concerned with the system's supportability. In addition to assessing these two areas, OT&E also identifies areas in which the system should be modified and provides information on requirements for tactics, doctrine, organization and personnel. OT&E is accomplished in an environment which is operationally realistic. Typical operational and support personnel will be used to obtain a valid estimate of the user's capability to operate and maintain the system when deployed.

SLIDE 3

DOD Directive 5000.3 specifies several objectives for the test and evaluation of computer software. The OT&E agencies are required to participate in the early stages of software planning and development. OT&E is to be conducted as part of the total system evaluation in the operational environment and shall include tests for both full-system and casualty mode operations. The questions which OT&E and DT&E must answer are related to quantifiable or demonstrable performance objectives. Has the software met an appropriate level of maturity for its present phase of testing?

SLIDE 4

This leads to several questions which must be answered in the testing of software. What procedures are required for the testing? (These procedures include the test design and specific test objectives as well as the procedures for implementing the test.) What is the benefit which will accrue from early involvement
TEST AND EVALUATION OF SOFTWARE

• EARLY INVOLVEMENT OF OT&E AGENCIES

• TOTAL SYSTEM EVALUATION FOR:
  - FULL-SYSTEM MODE OPERATIONS AND
  - CASUALTY MODE OPERATIONS

• PERFORMANCE OBJECTIVES FOR EACH STAGE OF T&E:
  - QUANTIFIABLE/DEMONSTRABLE
  - DETERMINE RESULT FOR FURTHER DEVELOPMENT/IMPLEMENTATION DECISION
  - USED TO DEMONSTRATE SOFTWARE MATURITY
QUESTIONS

- WHAT PROCEDURES SHOULD BE USED IN TEST DESIGN?

- WHAT SHOULD HAPPEN AS A RESULT OF EARLY PARTICIPATION IN TESTING?

- WHAT QUANTIFIABLE CRITERIA SHOULD BE USED TO EVALUATE SOFTWARE MATURITY, EFFECTIVENESS AND SUITABILITY?

- HOW ARE DT AND OT TO BE INTEGRATED?
of the OT&E agency? How will the OT and DT be integrated? What performance criteria should be used to evaluate the software? These questions address several issues: testing procedures, performance criteria and managerial responsibility for the testing and evaluation process. Perhaps a good way in which to gain a perspective on these issues is to take an overview of testing that is conducted in industry to ensure software quality.

SLIDE 5

These ideas are a summary from computer science literature on software testing and may not reflect any particular testing program but they do reflect some general trends. Software quality testing, sometimes called software engineering, is conducted using a well-documented test program which thoroughly exercises an algorithm and provides an accurate indication of major system errors. These software testing procedures are similar to the type of testing conducted in DT. Technical testing of this sort leads to an assessment of the software's accuracy. It may not, necessarily, evaluate the psychological factors in the user's response to the program which is influenced by the military mission, tactics and training procedures used. For this reason the military has a second testing method, OT.

SLIDE 6

DT provides a technical evaluation of several areas of the software's performance. The second requirement for testing method helps to ensure the operational applicability of the software in
SOFTWARE QUALITY TESTING IN INDUSTRY
- HIGHLY CONTROLLED I/O IN WHICH TO EVALUATE THE DESIGN-TO-SPECIFICATION
- USES BENCHMARKED PROGRAM
- SOFTWARE TESTING BY SYSTEM SPECIALISTS

ADVANTAGES
- THOROUGH
- REPRODUCIBLE
- WELL-DOCUMENTED

INDUSTRIAL METHODS ARE CLOSELY RELATED TO DT&E OF SOFTWARE
TECHNICAL EVALUATION

- PROGRAMMING LANGUAGE
- ALGORITHM CORRECTNESS
- THROUGHPUT ANALYSIS
- TIMING ANALYSIS
- NUMERICAL SENSITIVITY
- RESOURCE REQUIREMENTS
- OUTPUT ANALYSIS
- SOFTWARE STRUCTURE

SLIDE 6
addition to its technical correctness. The OT&E testing is a special requirement levied by the military which is user-controlled. If a system does not successfully pass the OT&E testing then it will not be recommended for service use. Who are the users for real-time operational software?

SLIDE 7

There are at least two categories of "user" for mission-related software. The first category is obvious: the tactical user for whom the system is designed. The primary concern for the tactical user of a system is the effectiveness of the software in carrying out the mission. The considerations which weigh most heavily with this user are twofold. Human factor considerations relate the software to the task to be performed. He is also concerned with the accuracy of the algorithm in achieving a problem solution. The second "user" is the Software Support Agency who is responsible for the maintenance of the software for the remainder of its service life (after it has been released for service use). This second user is interested primarily in the supportability of the software: program structure, language specification, details of the code, etc. Operational testing must serve both of these users.

SLIDE 8

Operational Test and Evaluation uses several standard tests to assess the suitability of a system. There are no standard tests to determine system effectiveness. These tests are system specific.
- USERS OF REAL-TIME OPERATIONAL SOFTWARE
  - TACTICAL USER - INTERESTED PRIMARILY IN SYSTEM EFFECTIVENESS
  - SOFTWARE SUPPORT AGENCY - INTERESTED PRIMARILY IN SYSTEM SUPPORTABILITY

- OPERATIONAL TESTING MUST SERVE BOTH OF THESE USERS
### OPERATIONAL TESTING MEASURES

<table>
<thead>
<tr>
<th>STANDARD TESTS (SUITABILITY)</th>
<th>QUANTIFIABLE MEASURE</th>
</tr>
</thead>
<tbody>
<tr>
<td>• RELIABILITY</td>
<td>• MEAN TIME BETWEEN FAILURES</td>
</tr>
<tr>
<td>• MAINTAINABILITY</td>
<td>• MEAN TIME TO REPAIR/MEAN TIME TO FAULT LOCATE</td>
</tr>
<tr>
<td>• AVAILABILITY</td>
<td>• RATIO OF OPERATING TIME TO OPERATING TIME AND REPAIR TIME</td>
</tr>
<tr>
<td>• LOGISTICS SUPPORTABILITY</td>
<td>. . . . . . . . . . .</td>
</tr>
<tr>
<td>• COMPATIBILITY</td>
<td>. . . . . . . . . . .</td>
</tr>
<tr>
<td>• INTEROPERABILITY</td>
<td>. . . . . . . . . . .</td>
</tr>
<tr>
<td>• TRAINING</td>
<td>. . . . . . . . . . .</td>
</tr>
<tr>
<td>• HUMAN FACTORS</td>
<td>. . . . . . . . . . .</td>
</tr>
</tbody>
</table>

**NOTE:** EFFECTIVENESS TESTS ARE SYSTEM SPECIFIC AND ARE GENERALLY QUANTIFIABLE IN TERMS OF PROBABILITY OF SUCCESS

SLIDE 8
and generally result in a quantifiable measure of effectiveness such as probability of detection, etc. The standard tests for suitability have few quantifiable measures and it is difficult to determine their application to software. For instance, the reliability of a hardware system may be measured as Mean Time Between Failures. This makes sense because hardware needs to be periodically repaired. However, software should be repairable once and for all for any particular fault. The reoccurrence of that fault should not be expected. The Mean Time to Repair a system is a good measure of maintainability for a hardware system, but it makes little sense when one is dealing with the world of programmer-wizards. There is a need to reassess the quantifiable measures of suitability used in OT&E to determine if they are appropriate for both hardware and software. Although the system is tested as a whole in military software testing, one must be able to determine which portion of the hardware/software system contributes to a suitability failure so that the appropriate recommendation can be made. The individual contributions of both hardware and software must be tracked while evaluating the system as a whole.

SLIDE 9

The results of DT and OT are presented in this conceptual flow diagram. The technical evaluation is conducted by the DT agency in its controlled environment to ensure logical correctness and suitability. Both the DT and OT agent must agree that the software
is of sufficient maturity to be operationally tested. The OT agent conducts an independent assessment of performance effectiveness and suitability in the OT&E environment and certifies the software as ready for issue to operating units upon completion of a successful evaluation.

SLIDE 10

Since the software testing requirement is recent, there remains much work to be conducted in this area. The standard OT&E tests must be evaluated to determine their adequacy for application to hybrid systems. Measures of effectiveness for both sets of testing (effectiveness and suitability) should be developed and evaluated. The interface between the DT and OT testing should be clarified to determine appropriate areas of responsibility for each community based on their capability to support the testing. Some of these areas, particularly test design and MOE development, are ripe for operations research applications. Others will require managerial efforts by the military services. The only alternative to these proposed efforts is relax the T&E requirements for software; however, this is not an alternative.
WORK REQUIRED

- DETERMINE THE ADEQUACY OF PRESENT STANDARD SUITABILITY TESTS AND MEASURES FOR HYBRID HARDWARE/SOFTWARE OR EMBEDDED SOFTWARE SYSTEMS

- DETERMINE APPROPRIATE PERFORMANCE OBJECTIVES FOR THE EFFECTIVENESS OF SOFTWARE AND MEASURES WHICH QUANTIFY THIS PERFORMANCE

- DETERMINE LEVELS OF SOFTWARE MATURITY WHICH ARE DEMONSTRABLE IN TERMS OF EFFECTIVENESS AND SUITABILITY WHICH MAY BE USED TO ADVANCE SYSTEMS THROUGH THE PROCUREMENT PROCESS

NOTE: THE ALTERNATIVE IS TO RELAX THE REQUIREMENTS FOR SOFTWARE TESTING WHICH IS NOT AN ALTERNATIVE.