REPORT OF

THE ARMY SCIENTIFIC ADVISORY PANEL

Ad Hoc COMMITTEE

FOR

PROJECT MASSTER

JANUARY 1973

TABLE OF CONTENTS

PREFACE

During the Army Scientific Advisory Panel meeting held at Fort Hood, Texas on October 5-6, 1970, Major General J. Norton distributed a set of scientific challenges which were later presented to the Panel as the basic terms of reference for an Ad Hoc Working Group on Project MASSTER. These proposed terms of reference were provided to the chairman, Dr. Vincent S. Haneman, Jr., Associate Dean, Oklahoma State University by letter of transmittal of the Assistant Secretary of the Army, R&D on 8 March 1971. During the spring and early summer of 1971, meetings were held in consideration of the terms of reference. An interim letter report was prepared and submitted on 19 August 1971 concerning instrumentation for the work being conducted at Fort Hood. In mid-July, an additional set of terms of reference were provided the Group concerning Night Low Level Operations of Helicopters. Emphasis was then placed upon the helicopter problem in light of the depth of study and time required to resolve the original set of scientific challenges concerning Project MASSTER.

It was decided in early 1972 that the additional terms of reference concerning the Night Low Level Operations did not directly affect the original terms of reference and therefore could be and were separated to form a second project of the Army Scientific Advisory Panel. These will be reported upon separately.

Of the original membership of the Working Group, Dr. Finn Larsen died in September 1971 leaving a great void in both the working group and the Panel, and Professor Lawrence O'Neill had to withdraw due to the pressure of other ASAP matters. Dean Kenneth Clark was added in November 1971 as a result of an identified need for professional review in his area of expertise.

The Working Group operated from inception March 1971 until April 1972 without a liaison officer. Project MASSTER provided military support at Fort Hood but the lack of these skills in Washington were felt in attempting to bring this work to a rapid conclusion. Major John Moore, OCRD, was appointed in April 1972 and his efforts materially assisted in this work.

The Ad Hoc Working Group has held 12 meetings and wishes to express a deep appreciation to all who have put forth so much effort. Of special note are those at MASSTER and Fort Hood, those at CDCEC and Fort Ord, and their science advisors, Dr. Philip Dickinson and Mr. Walter Hollis. The quality of work at both stations has been outstanding with very few exceptions. The transient nature of the programs has imposed great burdens on officers and civilian leaders in obtaining well defined test objectives, in maintaining the motivation of players and in the operation of new equipment with its shortcomings. Specific thanks are due to LTG G. P. Seneff, MG R. Shoemaker, and MG J. Singlaub. Additional thanks are due to CDC personnel and LTG John Norton for their interest and time.

Time has brought many changes, and this is true for MASSTER's mission and title since the study was initiated. "Project MASSTER" is now "Headquarters MASSTER" and the terminology in this report alludes to the former, as represented the original "terms of reference."

SECTION I

SUMMARY

The primary consideration of complex evaluations is the establishment of a measurement of achievement. The task facing the armed services is that of conducting tests such that the resulting evaluations will stand the scrutiny of soldier, student and politician. The terms "absolute and relative measures of effectiveness" are used to describe the hoped for final norm. These data are needed not only to consider how well the Army could accomplish its mission in combat but to provide data for optimal assignment of the limited procurement budget in peacetime. The ultimate descriptions of a perfect testing technique would provide information on how to sort new items and technologies that would accelerate the test process and achieve maxima in these "measures of effectiveness", thus eliminating unproductive equipment.

The Ad Hoc Working Group on Project MASSTER was requested to consider the "scientific challenges" distributed by Major General John Norton during the ASAP meeting at Fort Hood on October 5-6, 1970 as the basic terms of reference, the full text of the request is in Annex A. The Ad Hoc Working Group membership is listed in Annex B; a chronological summary of the Group's activities is contained in Annex C; the interim letter report is Annex D; and special papers felt to be germane to the overall problem area are included as Annexes E, F, G, and H.

The addition of "terms of reference", relating to the "Night Low Level Operation of Helicopters" (NLLO) in July 1971, caused considerable confusion and diversion of effort.

In early 1972 it was decided to make this a totally separate effort although the membership of the Working Group would remain identical and the task would be reported upon separately. At the January 1972 NLLO meeting it was further concluded that the major item of the Project MASSTER consideration was Scientific Challenge I, parts A and B and this group would restrict itself to respond to this.

The resultant considerations then of

> "IA - Derive useful measures of unit effectiveness in a test environment" and
>
> "IB - Develop techniques to sort new items and techniques in order to accelerate the test process,"

were developed as subquestions that follow.

1. Is there such a measure which can be described for all situations without consideration of the process, parameters or conditions?
2. What is the character of test in terms of known, unknown; measurable, unmeasurable; predictive, unpredictive parameters?
3. How are the present tests being conducted?
4. Are the results valid; do they have substantial meaning?
5. How is the information, knowledge, experience being used?
6. What is the status of measurement devices? Are they adequate?
7. Was the equipment, tactic or combination ready for test? Did the condition further affect the test procedure?
8. What techniques are currently used to determine nonproductive trends? How is this scheduled throughout the whole system?

In the following section of the report further development is provided relative to certain items considered as a result of the group's study of the above questions.

During the course of the Panel's studies, it became evident that the present Army research and development program is not taking full advantage of the knowledge and experience gained from MASSTER and to a lesser extent CDCEC. The skills and education gained by the engineering and scientific cadre of Army officers does not appear to be formally programmed into the system.

As a result of the study and visits to various posts, forts, et al, for both the Project MASSTER and for Night Low Level Operations, serious conflicts seem to appear relative to the testing process and the division of labor. This has been presented as Annex E, a Special Paper by Professor Enoch Durbin.

During the course of the effort Dr. Paul Kruse was requested to study and recommend an instrumentation system which could establish when a line-of-sight existed between two points. This was done as a portion of the effort on Terms of Reference - IB and is presented as Special Paper Annex F.

Finally, two techniques are described that have merit under testing situations which, although the number of situations are limited, may establish more valid measures of effectiveness, Annex G and H.

SECTION II

CONCLUSIONS

1. It is the opinion of this Working Group that absolute measures of effectiveness can be established only for certain types of problems where the object is well defined in terms of a finite number of measurable parameters. There are a significant number of situations where no valid numerical value can be established within the "scientific process." Many of the experiments being conducted both under Project MASSTER and CDCEC are of this latter type. Under these conditions, reliance must be placed upon judgment, and while numbers are generated and are in terms of relative measures of effectiveness, there exists no panacea in "measures of effectiveness" which will sanctify the relative values that they are endurable for all time. The data cannot be taken out of context without destroying their validity. At present, the national rationale that "everything can be quantified into immutable numbers" has forced reasonable evaluations into nightmares of unqualified, unsubstantial, unrealistic accounts, especially when the data is used without the limitations imposed by full knowledge of all the test environments and conditions.

The tests at MASSTER and at CDCEC are necessary to provide the Army with a basis upon which to make rational decisions regarding development and procurement. When material, personnel and tactics are combined, as is necessary in the above evaluations, the test situation is ever changing, never static nor the same even on two successive tests. The ability to measure and record such items as stress level, personnel learning curves, and troop motivation is not within the "state of the art" for these field exercises and, therefore, judgment must be applied. Judgment is virtually impossible to divide more finely than good, bad, or indifferent and, therefore, even "relative measures of effectiveness" demand an unusual level of expertise in testing to assure that the number of basic parameter changes will not totally destroy the comparisons being constructed.

2. It is generally conceded that system effectiveness, particularly combat systems, is a measure of the extent to which a system may be expected to achieve a set of specific mission requirements. The basic approach of assessing this achievement may be judgmental, analytical, empirical or a combination of these. The effectiveness of a combat system is related to the property of the system output which was the basic reason for acquiring or requiring the system; i.e. the performance of some specific functions. If the system is effective, it performs these functions well. If it is not effective, the deficient attributes must receive special attention in the effectiveness evaluation. It follows, therefore, that the effectiveness measures are subject to the conditions on the battlefield (or simulated test battlefield environment) and are potentially sensitive to changes in the field situations. The effectiveness, at one time, may be measured in terms of damage inflicted on enemy ammunition supplies, while at another time it may be measured in terms of time to intercept an intruding tank column.

If mission profiles have been specifically defined, the problems of stating measures of effectiveness can be expressed as a probability of accomplishing all or a given part of the system's mission. In other cases, however, where missions cannot be clearly associated with the system, or where the variety of missions exceeds the practicality of evaluating all of them, the measures of effectiveness should be related to the physical parameters of the systems, such as, range, channel capacity, resolution, processing time, warhead effectiveness, etc.

The quantization of the measure of effectiveness is, indeed, desirable and is often approached by assigning specific "values" to performance of subsystems or general characteristics; however, in most cases it is not possible to aggregate these "values" into a combined figure of merit. This is not necessarily a disadvantage, since the decision maker will still be in a position to select options even if the effectiveness is stated in separate categories; granted that the decisions will not be as simple or clear-cut as in the case of a single effectiveness measure.

In Summary, the expected outcome of test and evaluation must be carefully analyzed in terms of the measurable parameters, the judgmental decisions and the interrelationships which exist between the various factors to be evaluated. From this type of analysis, the appropriate measure or measures of effectiveness can be derived with more deliberate choice.

3. From the many meetings held in various locations during this effort, it appeared to the committee that the experience gained from

MASSTER, and to a lesser extent CDCEC, is not being assimilated as rapidly as it could be by the developing agency. This includes not only the actual information gains with respect to equipment but also the knowledge and experience gained and hopefully retained by the officers who conducted and led the test. The experience gained both at Fort Hood and Fort Ord would be of real value to the AMC laboratories if the officers could be so utilized by assignment to the laboratories as an information source.

4. During this study the Ad Hoc Group has been privileged to review a very large number of tests performed. These tests met all the criteria that anyone could demand for reproducibility, player motivation, and quality of measurement under the field conditions existing. Several times, though, the item to be evaluated was not in a form which could be properly tested in the MASSTER environment and the item's performance was severely degraded. It was concluded that a few tests were made on "bread board" equipments where failures or shortcomings were due to the nature of their condition and not a failure of concept or potential value to the Army. It appeared to this working group that, in some instances, the test reports have been so terse that vital information has been lost, and that the shortcomings have not fully been disclosed with suggestions for repair or modification.

5. Any large scale test of the type used by MASSTER to evaluate a device or concept is in effect a simulation of an actual engagement. Simulations are effective to the degree that they sample the actual test modes of response to be expected in the field. Any such tests

must, therefore, be judged upon two separate dimensions in order to provide independent answers to the following questions:

(a) Was the simulation of field conditions actually achieved? If not, were the deficiencies of such magnitude that the results of the exercise must be disregarded or limited in their evaluation and utilization?

(b) Did the device or concept under study produce any measured changes in effectiveness of the unit?

Obviously, the first question must be answered satisfactorily before any interest in the answers to the second question can be generated. Should the simulation be only a poor approximation of battle conditions, unit effectiveness may vary in either direction. A particular concern in the MASSTER exercises under review are those conditions which affect the operators' attitudes and emotional responses. For example, if the operator has less than total motivation to survive, his performance in the test will be below his potential. This was observed in tank crew operations when searching for attack helicopters. The reverse may also be true where the operator may be under less emotional stress than in combat and he may well perform at a level superior to his potential battlefield performance. This latter case may occur in the situation of the helicopter pilots in their mock attacks on the tanks.

6. From the early phases of this study, it was concluded that a system of instrumentation was needed for Project MASSTER. The data reviewed disclosed that it would not be possible to conduct

experiments, exercises or tests which would provide the information required without a sophisticated time, location and event measurement system. Further, the data must be capable of field use, direct computer analysis, and not interfere in the test being conducted. The total system should meet the suggestions presented by LTG John Norton at the ASAP May 1971 meeting at Fort Ord, i.e.

(a) Design a portable instrumentation system for monitoring forces in the field, and

(b) Simulate the stress of combat in test situation.

After consideration of the proposed Project MASSTER instrumentation plan, an interim letter report Annex D was submitted on 19 August 1971. A special requirement, based upon CDCEC experience, was the need for a gun simulator to provide strike/kill information which would produce a more realistic situation for 6 (b) above.

7. As the study progressed there were a number of sessions when the working group or individual members were requested to consider specific, well delineated problems. These interfaces provided the basis for the conclusion that real value can accrue to the Army by fostering interplay between science advisors and panel members.

As an example, Annex E was prepared in response to a request from the science advisor to CDCEC, Mr. Walter Hollis by Dr. Paul Kruse. It is understood that this suggestion is being implemented by Harry Diamond Laboratories for CDCEC. While this is the most notable of the items in support of Scientific Challenge IB, a number of other suggestions have been made directly to MASSTER personnel and it is understood that several of these have been implemented. The close relations with both MASSTER and CDCEC have made this method of operation most covenient for the Ad Hoc Working Group.

8. During the course of review the Working Group concluded that a number of analytical techniques have serious merit for the types of tests considered by MASSTER. In the following discussions, some of the inherent problems or limitations are presented to emphasize that while each technique has value, it is not a panacea and must be applied only with a clear understanding of all of its limitations.

(a) Evaluation of complex systems may be accomplished by design of a weighted sums computation (Performance Index). While such a method does not assure the validity of the output numbers it does provide a method of evaluation which may be applied to a large category of problems. The limitation is apparent, and must be borne in mind when applying the results. The choice of the weighting factor of the different facets of the test may have to be based upon experience, upon command opinion, upon intuition, or it may be based upon hard, sound logic. The validity of the results will hinge upon the weakest of the weighting factor choices.

As an example of the simplest form of the problem and the limitations, suppose that a report is requested on the utility of a new viewing device. The test to be considered may have three facets, each with a desired outcome. Item one may be the maximum distance at which an average soldier may recognize a tank with the equipment. Item 2 may be the length of time the equipment may be used before failure. Item 3 may be whether or not to repair the unit at the operating level when it fails. In this case Item 2 is fairly sound and easy to measure, using normal statistics and use by troops in the field to determine a "mean time" to failure. A problem might be determination of the "using time" or "running time" in the field, but this can be resolved. Item 1 may be more difficult to assess since it depends on the individual soldier--his attentiveness, his motivation to see the tank, his eyesight under test conditions--it also depends upon test conditions such as light, contract, clarity of the air, other distracting conditions, imposed by tactics or battle simulation. Item 3 is the most difficult to determine since it is a question of judgment as to the degree of complexity as seen by the individual required to repair the unit. Is a high school education required? must the soldier be mechanically inclined? does it require special tools and/or training? What weight should be applied to each of these facets--is recognition distance the prime requisite, with failure rate next and repairs a poor third? The agency requesting the test must keep those subsets of requirements in mind when testing is desired. The testing agency then must also report its findings in the same fashion since the "found to be best" depends on the weighting applied to each facet. In this example, problems not covered include such basic questions as: Was the device used? was it used at the proper time? and did it add to the Unit effectiveness, however measured? would troops carry it?

While this example is simplistic, the technique is, nonetheless, used often and with good results. A more detailed discussion may be found in Annex G.

(b) Due to the random variables introduced by personnel and other factors it is often necessary to repeat a test several times to obtain a valid mean or average. With large scale tests, such as those conducted at MASSTER, it is not economically sound to have a number of replications to obtain significant statistical substantiation. CDC Headquarters has provided an innovation to the testing routine which has merit. The basis for the technique is that the large test is in fact a composite of smaller tests all performed at the same time. Therefore, any set of the small tests, run in isolation or by themselves, may be repeated a sufficient number of times to obtain a reasonable statistical "standard." When the large test is performed, the "standard" obtained earlier may provide the basis for comparison--this then may be a "measure of effectiveness." It is important that the implications be fully understood with regard to the comparison made by this technique. The variables of tactics and personnel still are present and will influence the result so that the item measured is not just the quality of this piece of equipment but it is the combination of the equipment with tactics and personnel. The changes due to these items may be such that reproducibility of the results may be so low that the analysis is useless.

(c) The application of a modification of the Delphi technological forecasting system has the possibility of overcoming some of the questions where "best mix" of equipment and motivation are concerned. This is presented in Annex H. The primary problem is that of motivation of units in trying new techniques and equipment as opposed to known methods of operation. The major limitation is that, regardless of the sophisticated title, this is a public opinion poll, even though the public is limited and their opinion may have high credence value.

SECTION III

RECOMMENDATIONS

1. The Working Group recommends that the concept of an overall single measure of effectiveness, either absolute or relative, be viewed with skepticism. The "simple" test of equipment in the field used by the "troops" in some simulated battle condition does not lead to "simple" answers which can be taken out of context and removed from the inherent limitations of the situation. The measurements are vital and valid within the restrictions of the test, but the data demands the reverence that it deserves.

2. It is recommended that the Army laboratory system of AMC provide a positive organizational structure to capitalize on the information gains from MASSTER and CDCEC tests. The recently added AMC support group at MASSTER may provide the interface needed between all facets of the developer community, from laboratory, to test, to troop use. Close attention must be paid to assure achievement of this goal. Utility of MASSTER to input and to develop requirements should be maximized.

3. It is recommended that a study be made of the naval system in which officers are assigned to laboratories and in particular the Naval Ordinance Laboratory, China Lake, to be used as a source of user experience. It is understood by this Working Group that these officers have no duties other than to provide their backgrounds and experience for the engineering and scientific cadre of the laboratory with the development of new weapon systems. The Army could further modify this to directly use the skills of officers

who have been involved in either MASSTER and CDCEC since many of these officers have field experience and have engineering backgrounds.

4. This Working Group recommends that all reports of tests include definition of the shortcomings disclosed by the tests, where they are located, and make recommendations for repair or modification so that they will more nearly meet the originator's and Army's requirements. These disclosures must not be limited to the test areas defined in the experiment requirements but include all those found.
5. It is recommended that, when "bread board" equipments are submitted for tests under the rigorous environment imposed by troop use, adequate spares and technological support be provided by the requesting agency to assure that the equipment is in a condition such that the tests will in fact measure the potential of the item or the concept under consideration.
6. While there is no clear and simple solution to the problem of simulation and operator motivation in stress, it is most strongly recommended that each test incorporate evidence of the degree to which battlefield conditions were approximated so that later review may include allowances for this item.
7. It is recommended that the gun simulator and complete information gathering system for Project MASSTER should receive high level emphasis. The early procurement of these items will directly affect the quality of the test results at Fort Hood. Development of these items will improve the test process throughout the Army.

8. A recommendation of this Working Group is that some portion of the relationship of the Ad Hoc Working Group and the science advisors of MASSTER and CDCEC be continued. This might be best accomplished on an individual basis, in support of items Ia and Ib of the Scientific Challenges.

9. It is recommended that consideration be taken of the various analytical techniques discussed in previous sections and the appropriate annex, as well as promote development of other systems. In all methods, users of the data must be alert to the inherent limitations of the analysis and results. There is great danger in taking information out of context.

10. It is further recommended that a working group within the Army be established to study the theory of testing as applied to these general categories of problems. Studies should include but not be limited to the lines suggested in this report and Annexes.

11. This Working Group has not directly tackled here items IIA and IIB, Annex A of the Scientific Challenges, and recommends that a review be made of these items to determine whether a new effort is warranted.

ANNEXES

A. Proposed Terms of Reference

B. Ad Hoc Group Membership

C. Chronological Summary of Group Activities

D. Interim Letter Report - 19 August 1971

E. Special Paper - Army Testing Hierarchy

F. Special Paper - LOS Existence System

G. Special Paper - Weighted Sums

H. Special Paper - Modified Delphi

ANNEX A

Proposed Terms of Reference

ASAP Ad Hoc Working Group on Project MASSTER

The proposed terms of reference follow:

I. Introduction and General Comments

Project MASSTER is a comprehensive program for the testing and evaluation of components and systems related to sensing, target acquisition, and night observation which are for the purpose of making the individual soldier and Army units more effective in combat. Like all test and evaluation enterprises, it also serves as a source of advice on means of improving components and systems within its province and as a source of proposals for new components and systems. Although these ancillary functions are important, they are not primary ones in MASSTER.

The materiel tested and evaluated by MASSTER cannot be divorced from soldiers and military units because it achieves, or fails to achieve, its purpose insofar as it makes or fails to make soldiers and units more effective. Moreover, an assessment of this materiel cannot be made without considering alternative means of performing a given function and estimating the "costs" (not necessarily exclusively in financial terms) of the various means of improving individual and unit effectiveness.

It is clear, therefore, that the task of Project MASSTER is important, basically consequential for the Army of the future, and extremely difficult.

The central task of the ASAP Working Group on Project MASSTER is to develop guidance, on the basis of the scientific and technical qualifications of its members, the advice of military and other experts, and, most basically, the insights provided by Project MASSTER personnel, to make the work of Project MASSTER as fruitful as possible.

II. Specific Comments

Emphasis has been given in the preceding section and in the comments of Project MASSTER officers to the necessity of measuring individual and unit effectiveness when using the components and systems with which MASSTER is concerned. Measures of effectiveness are needed in both an "absolute" sense and in a "relative" sense. That is, measures that appeal to common sense in describing the ability of a soldier or a military unit to do his or its job in combat are needed, and similar measures that allow sensible choices among alternative components and systems are also needed. These measures are needed both because of the obvious importance to the Army of knowledge of how well it can do its duty in combat in the future and because of the need to spend a limited budget as wisely as possible.

No task placed upon the ad hoc group is more important than that of suggesting measures of absolute and relative effectiveness that can stand the scrutiny and win the approval of the soldiers, scientists, engineers and managers who will be called upon to make Army procurement decisions in the future. The soundness and usefulness of suggested measures must be demonstrable in the test environments that MASSTER will provide.

MASSTER could dissipate its resources with inadequate return, by testing and evaluating the nearly unlimited quantity and diversity of components and systems placed before it from various sources. There is a need, therefore, to develop sound means of choosing among candidate components and systems which will yield results of greatest utility in the shortest time consistent with reliability of results, and at lowest achievable cost. In this respect, the scientific and technical experience of the ad hoc group, influenced by military advice, may be extremely useful. For example, if a costly technique is in principle, or as a matter of practicality, certain to suffer from high false alarm rates or very low detection rates or both, it clearly does not merit consideration and effort by MASSTER until after more promising alternatives have been tested and evaluated.

The task of developing practical advice for MASSTER to permit the project to use its funds and its other resources best is an important one for the ad hoc group.

In combat, the detection of "something" in the vicinity of a soldier or a unit may be of some worth. However, such a detection may also be of no worth at all because if many detections are of "non-targets", they will lead to a burdening of the communications and information transfer and processing system without improving the unit's combat effectiveness. Such a "non-target" might be better undetected. It is, therefore, clear that "detection" must be associated with "target identification" if it is to improve individual or unit combat effectiveness. It is also clear that in general "identification" requires the transfer of more information to a human or automatic data evaluator than does mere "detection" (e.g., of a seismic, magnetic, or acoustic disturbance). Accordingly, there is a need to "trade" the additional cost, complexity and (presumably) reduced reliability of detection and identification systems, as contrasted with mere detection systems, against the increments in individual and unit combat effectiveness that are achieved through the use of detection combined with target identification techniques.

Advice based upon the scientific and engineering experience of the ad hoc group, on the basic "price" that must be paid for target identification (e.g., increases in information channel bandwidth, improvements in signal-to-noise and signal-to-interference ratios, and additional component costs) and on the nature of the "trade" previously mentioned, is sought. This is an important task of the group.

Finally, the proliferation of target detection and identification devices and systems throughout our Army in combat poses the problem of devising means of processing a large amount of information rapidly, reliably, and at acceptable cost, and of "displaying" the results of such processing in a way that will permit responsible officers and enlisted men to make their decisions correctly and with confidence. To solve this problem requires that a set of decisions be correctly made concerning such things as how to use multiple or redundant detections, how to use correlated detections, how to distribute information throughout a combat force, and how to "filter" it at various points in the force. Advice based on the scientific and technical background of the ad hoc group, with the guidance of experienced military personnel, is desired concerning the problem here described.

III. Assignment

In the light of the preceding comments, the following "scientific challenges" distributed by Major General Norton during the ASAP meeting at Fort Hood on October 5-6, 1970, are offered as the basic terms of reference for the ad hoc working group on Project MASSTER.

Project MASSTER

Scientific Challenges

I. To Improve Test Process

  A. Derive useful measures of unit effectiveness in a test environment.

  B. Develop techniques to sort new items and new technologies in order to accelerate the test process.

II. To Improve Combat Forces

  A. Combine target detection capabilities with a positive means of determining target description.

  B. Overcome the information processing and intelligence display problem.

IV. Membership

The proposed membership of the ASAP Ad Hoc Working Group on Project MASSTER is as follows:

Dr. Vincent S. Haneman, Jr. - Chairman
Dr. Finn Larsen
Dr. Paul Kruse
Mr. Lawrence H. O'Neill

and, either as members or consultants to the group, two officers chosen by the director of Project MASSTER, one of whom has had substantial combat experience in Vietnam and one of whom has had substantial combat experience in Europe or Korea.

The study effort should not take over a year.

ANNEX B

MEMBERSHIP

Finn Larsen

Paul Kruse

Kenneth Clark

Lawrence O'Neill

Manford Gale

Enoch Durbin

Vincent S. Haneman, Jr.

Chairman

Military

Major John Moore

Lt. Col. Edward Roderick

ANNEX C

ASAP

Ad Hoc Committee -- MASSTER

Meetings:

19 March 1971	Washington, D. C.
30 March 1971	Fort Hood
25/27 April 1971	Fort Hood
2/3 May 1971	Fort Ord*
6/8 June 1971	Washington, D. C.
14/16 July 1971	Fort Hood
23 October 1971	Fort Sill*
2/4 November 1971	MASSTER
13 March 1972	Washington, D. C. *
21/22 March 1972	MASSTER
11/13 April 1972	CDCEC
3 August 1972	Carlisle Barracks*

*In conjunction with another meeting

ANNEX D

Oklahoma State University

Engineering Research/(405)372-6211, Ext 7553/Stillwater, Oklahoma 74074

19 August 1971

Honorable Robert L. Johnson
Assistant Secretary, Army (R&D)
Washington, D. C. 20310

Dear Mr. Secretary:

Following your letter of 8 March 1971, requesting that an ad hoc committee of the Army Scientific Advisory Panel study the Project MASSTER Program at Fort Hood, the committee was formed. The committee members:

ASAP:
- Dr. Paul Kruse
- Dr. Finn Larson
- Mr. Lawrence O'Neill

DA:
- Mr. Manny Gale

PROJECT MASSTER:
- Lt. Col. Edward Roderick

have met a number of times, as listed upon Attachment 1.

From these meetings one item has developed as being a problem of overriding importance for the success of Project MASSTER, and, as an item for consideration with respect to the scientific challenges, Item I A & B of the Basic Terms of Reference. This item is the lack of appropriate instrumentation to measure the fundamental parameters concerned with the major tests being conducted.

The committee has reviewed the concepts of the three major units of instrumentation, which are either under contract or in the process of procurement, to resolve this problem. All three units are vital to the concepts of Project MASSTER, but within these three, precedence should be: (a) Lasser Engagement Scoring System (gun simulator); (b) The Data Collection System; and (c) The Position Reporting and Recording System, -- in that order.

The committee has not had time to review in detail the technological mechanization, but from what material that has been surveyed, it endorses the development and recommends most strongly that procurement proceed with all possible speed.

The tests witnessed at Fort Hood seemed to be well conducted, well motivated, and well run, but they lacked substantive measurement to establish any hypothesis.

With the present schedule of tests at Fort Hood, especially the tank/helicopter efforts, it is imperative that a gun simulator system be used to obtain meaningful results.

The systems proposed seem to meet the most recent challenge presented by Lt. General Jack Norton at the Fort Ord meeting -- i.e. a truly portable instrumentation system capable of monitoring the position and activity of elements engaged in simulated combat.

We therefore respectfully recommend that all possible emphasis be placed upon attainment of this capability in the immediate future.

Respectfully,

V. S. HANEMAN, JR., P. E.
Chairman, Project MASSTER, Ad Hoc Committee

cc: Lt. General Wm. Gribble
Lt. General J. Norton
Lt. General R. Williams
Lt. General G. P. Seneff
Major General Jack Guthrie
Brig. Gen. R. M. Shoemaker
Dr. Phillip Dickenson
Lt. Col. William Boyd
Committee Members

ASAP

Ad Hoc Committee

PROJECT MASSTER

Meetings:

19 March 1971	Washington, D. C.	Haneman
30 March 1971	Fort Hood	Haneman, Roderick
25/27 April 1971	Fort Hood	Kruse, Gale, Larson, Haneman
3/4 April 1971	Fort Ord	Kruse, O'Neill, Haneman
6/8 June 1971	Washington, D. C.	Kruse, O'Neill, Gale Larson, Haneman, Roderick
14/16 July 1971	Fort Hood	Kruse, O'Neill, Haneman, Roderick

ANNEX E

21 March 1972

LOS Existence System

Paul W. Kruse, Army Scientific Advisory Panel

Walt Hollis, CDCEC, has pointed out the need for a system which determines the instant at which a line-of-sight appears between a helicopter and a ground-mounted system, e.g., a Vulcan. This memo presents a first cut at analyzing a cooperative approach involving a modulated infrared beacon mounted on the helicopter and an infrared receiver mounted on the Vulcan. The beacon could be a tungsten ribbon filament bulb mounted on the roof of the helicopter within a rotating cylinder having slots cut in it. A second concentric counter-rotating slotted cylinder is employed such that the radiation emitted through both cylinders is modulated in a quasisinusoidal manner. A near infrared transmitting filter is employed so that the source is covert, emitting between roughly 0.9 and 2.5 microns wavelength. A second approach would be to employ a xenon lamp without the cylinders, the modulation produced by sinusoidally modulating the lamp current. In either case a Fresnel lens or equivalent causes the modulated radiation to be emitted isotropically in the horizontal plane and through, say, ± 60° in the vertical plane. The beacon would be located on the helicopter in such a location that it could be viewed readily from any point within the emitting pattern. If the aircraft structure blocks the view from certain aspect angles, then two beacons at different points on the airframe are necessary. Since the modulation from them should be synchronized, the xenon lamp approach would be the more desirable in that event.

The ground based receiver consists of an uncooled infrared detector, probably lead sulfide but maybe silicon, with a lens and scanning mirror. The axis of the lens would be vertical, with the detector mounted on the axis below the lens at the focal point. Above the lens is a plane mirror inclined at 45° from the horizontal, rotating about the vertical axis and thereby scanning through 360° about the vertical axis. The size of the detector and the focal length of the lens are so chosen that the instantaneous field of view projected into space is a long narrow slit, the long axis vertical, rotating about the vertical axis of the scanner. Therefore this instantaneous field of view, of horizontal extent say, 1°, and vertical say, ±30° from the horizontal, sweeps across the helicopter each scanner revolution, which might be at a rate of, say, 10 revolutions per second.

The analysis which follows was done in a very short time the evening of March 21. It is preliminary only, and may contain errors. The choice of the values of the important parameters, near the end of the analysis, is somewhat arbitrary. It is shown that these choices give rise to a signal-to-noise ratio of unity for a 900 meter range. Assuming the atmospheric transmission and detector detectivity values are acceptable 4- kilometer range with a signal-to-noise ratio of five will require either increasing the optical radiated power in the wavelength region of interest to be greater than the 10 watts used, or else a multielement detector array will be required, with attendant increase in complexity and cost.

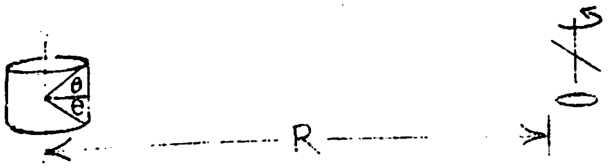
It should be borne in mind that the system described is a clear weather one. However, the helicopter, whether a scout or attack one, is basically a clear weather system. As the weather degrades the line-of-sight existence system so too it degrades the target acquisition and engagement capability of the helicopter. Therefore, the clear weather aspect does not unduly restrict the utility of the system.

An interesting side benefit is that by modulating the beacon rotation rate with input from the altimeter, and with a magnetic pick-off on the scanner at any instant, it may be possible to determine the angular coordinates of the line-of-sight.

Analysis

Consider a beacon mounted on a helo consisting of a modulated xenon lamp or else a tungsten bulb with a rotating can in can. Let the modulation frequency be f_o. Let the input power be Pi and the output power in the wavelength interval from λ_1 to λ_2 be P_o. Then by definition

$$\eta = \frac{P_o}{P_i}$$



Let the geometry be such that the radiation is directed over all horizontal azimuths and, say θ vertically

Let this be intercepted by a lens of diameter D located at a slant range R from the source. Then the intercepted modulated power (sinusoidal) is

$$P_{int} \quad \frac{P_o \pi D^2/4}{2\sqrt{2}\cdot 2\pi R\cdot P\cdot 2\theta} \cdot T_a = \frac{\eta P_i D^2 T_a}{32\sqrt{2}R^2\theta} \quad ;$$

Where T_a is the atmospheric transmission factor between λ_1 and λ_2 which is a function of range, humidity, etc. Then the signal to noise ratio measured in the bandwidth Δf centered at the modulation frequency f_o is

$$S/N = \frac{P_{int}\ T_r}{P_{n_{\lambda_1-\lambda_2}}\ (\Delta f)^{1/2}} \ ;$$

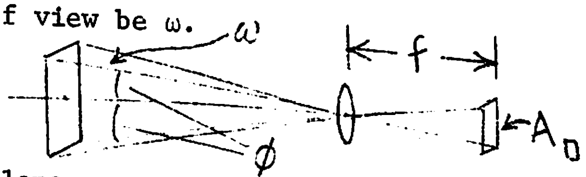
Where P_n is the noise equivalent power of the detector per root cycle measured at f_o over the spectral interval from λ_1 to λ_2 and T_r is the transmission factor of the lens. In terms of the detectivity $D^*_{\lambda_1 - \lambda_2}$ the NEP is given by

$$P_N = \frac{(A_D)^{1/2}}{D^*_{\lambda_1 - \lambda_2}} \ ;$$

so

$$S/N = \frac{\eta P_i D^2 T_a T_r D^*_{\lambda_1 - \lambda_2}}{3 \cdot 2\sqrt{2} R^2 \theta (A_D \Delta f)^{1/2}}$$

Now the size of the detector element is determined by the need for the scanning optical system to sweep out a sufficiently large volume of space. In horizontal subtense the angle can be very small, but in vertical subtense it must be large. Let the vertical angle be 2θ and the horizontal instantaneous field of view be ω.



The detector geometry is thus a long narrow element, or a linear array should the s/n dictate it to be so. This detector actually looks vertically through a lens at a plane mirror inclined at 45° with respect to horizontal, rotating about vertical axis. The value of A_D is thus

$$A_D = 2f^2\phi\omega$$

so

$$S/N= \frac{\eta P_i D^2 T_a T_r D^*_{\lambda_1} - \lambda_2}{G4R^2 \theta f(\phi\omega)^{1/2} (\Delta f)^{1/2}}$$

The bandwith Δf is determined by the modulation characteristics of the source and the dwell time as the instantaneous field of view sweeps across the source. The dwell time T_d is given by

$$T_d = \frac{\omega R}{\alpha} \quad \frac{\omega}{\alpha} \ ;$$

Where α is the angular sweep rate (radians/sec). Then, to a close approximation

$$\Delta f = \frac{1}{2T_d} = \frac{\alpha}{2\omega}$$

So the detection range $R_{s/n=1}$ for s/n=1 is given by

$$1 = \frac{\eta P_i D^2 T_u T_r D^*_{\lambda_1} \lambda_2}{64R^2_{S/N=1} \theta f(\phi\omega)^{1/2}} \quad \frac{2\omega}{\alpha}^{1/2} \ ;$$

$$R_{S/N=1} = \frac{\eta P_i D^2 T_a T_r D^*_{\lambda_1} - \lambda_2}{32\sqrt{2}\theta f(\phi\alpha)^{1/2}}^{1/2}$$

Assume $\eta = 0.1$

$P_i = 100$ watts

$D = 10$ cm

$T_a = 0.5$

$T_r = 0.8$

$D^* = 10$" cm Hz 1/2 watt (uncooled PbS)

$\theta = 1$ rad.

$f = f/\text{no.} \times D = 20$ cm (f/2 lens)

$\phi = 0.5$ rad.

$\alpha = 10$ rev./sec. $= 20\pi$ rad./sec.

Then $R_{S/N=1} = \frac{0.1x10^2x10^2x.5x.8x10''}{32x1.4x1x20(.5x20\pi)^{1/2}}$

or $R_{S/N=1} = .9x10^5 cm = 0.9 km$

An increased range can be obtained with an increase in the lamp power or employment of an array of N detectors. It would be much better to increase the radiated power P_o of the lamp above the 10 watt value chosen. To achieve a 4km range with s/n of one in this way would require a P_o of

$$P_o = \frac{4}{0.9}^{2} \times 10 = 200 \text{ watts}$$

For a value of n = 0.1, the input power Pi would need to be 2 kilowatts, which is somewhat high. The range can also be increased by going to a multielement linear array of N elements. Since the s/n depends upon N, the range to 4km by increasing the number of elements only would require

$$N = \frac{4}{0.9}^{4} = 400$$

This is a complex system.

The range may also be increased by increasing the mirror diameter from 10 cm to, say, 100 cm. However, to keep the f/no. at two, the focal length will have to be increased to 200 centimeters. The net increase is therefore

$$R_{S/N:1} = \frac{10}{10}^{2} {}^{1/2} \times 0.9 = 2.9 \text{ kilometers}$$

In summary, the range can be increased by several ways. The best may be a combination of several approaches. It should be borne in mind that s/n = 1 is unsatisfactory; it would be much better to have a system which provides s/n = 5 at R = 4 kilometers.

ANNEX F

ARMY TESTING HIERARCHY

Enoch Durbin

Much of the difficulty in MASSTER, CDCEC, and NLLO programs have to do with confusion about the division of labor between the various testing processes in the development of new equipment for improving army capability. An independent group should be asked to examine this division of labor in order to sharpen the goal of each testing step.

They could look at the various testing steps normally used in the development of equipment to enhance aircraft military utility, and rank these steps in a _hierarchy of realism_, in order to indicate the type of question which could be answered by testing at that level.

For each level they could indicate the kind of data which should be collected and, therefore, the kinds of instrumentation which one should use.

For each level they should establish the kinds of reasons that would suffice for equipment or system rejection at that stage of testing.

For example in the case of the aircraft night operations, the hierarchy of realism could be ranked from least realistic to most realistic as follows:

1. Laboratory and bench tests to answer questions about _equipment characteristics_ such as field of view of a night vision device, light levels, power consumed, weight, size, etc.

2. The _night vision test range_ would _perform testing of the influence of human operators on equipment characteristics in a static_

environment. They would deal with such questions as the range of detectability by humans using the equipment as a function of target characteristics.

3. The ECOM flight simulator would perform testing of equipment with humans in a loop in a single aircraft to study the degradation of performance when the human performs other duties in a static environment. Here you make equipment systems studies to establish potentially useful aircraft system candidates which are worthy of future evaluation.

4. Flight testing using the single RAVE type aircraft. This includes testing of candidate systems with humans in a loop in a single aircraft to study the degradation of performance when the human performs other duties in a dynamic environment.

5. CDCEC - evaluation of candidate systems with humans in a loop in a dynamic environment in a multiple aircraft situation where the chief addition to realism is the added complexity due to multiple aircraft and tactics.

6. MASSTER - To answer questions about incremental military value of equipment -- i.e. incremental contribution.

In this hierarchy of realism, the instrumentation employed goes from complex in the laboratory bench test to the simplest in the MASSTER test (or perhaps even to none at all at the MASSTER level); where the military judgment is the key element.

Much of the testing that this committee has witnessed in these various testing places suffers from a lack of appreciation by the tester as to steps that fit in the overall process.

ANNEX G

DEFINING MEASURES OF UNIT EFFECTIVENESS IN A TEST ENVIRONMENT - A PROPOSED LONG TERM APPROACH

Paul W. Kruse

Army Scientific Advisory Panel

The task set before the ASAP Ad Hoc Committee on MASSTER is that of defining measures of unit effectiveness in a test environment. According to the MASSTER Dictionary of Terms (Supplement to AR 310-25, Dictionary of US Army Terms ATMAS-OP-OR), measures of effectiveness are defined as "quantified functions of major significance that defines acceptance or rejection of the concept being tested." The Committee considers the modifier "unit" to refer to the organizational structure, such as squad, section, platoon, company, etc. Thus the Committee understands the task set forth to be one of compiling and assessing methods by which MASSTER can determine the increase in effectiveness of a squad, section, platoon, company, etc., in carrying out their various missions when aided by various STANO items, over the effectiveness without the items.

When taken literally, the task is an enormous one because of its general nature. First, the list of STANO items is extremely long, certainly several hundred. When grouped into categories of like items, viz., low light level television, forward looking infrared systems, foliage penetration radar, the list becomes shorter, but certainly numbers some tens of categories. Each category applies to each unit size, of which there are eleven, from squad to Army group. Thus, there is a two-dimensional matrix of some tens by eleven, some hundreds of combinations. But over and above this is the list of representative scenarios applicable to each unit. In the Committee's mind, the question of defining the representative scenarios is a major task in itself. How many scenarios are needed to adequately describe the operations and tasks of each unit? In order to keep the number low, each scenario must be quite different from the others. To the best of the Committee's knowledge, this question has never been answered. The number is probably a large one, much greater than ten per unit. Thus, a third dimension whose scale is at least ten is added to the matrix, so that the resulting number of combinations is now in the thousands.

The Committee cannot possibly answer the task literally. Indeed, were the task somehow reduced from a level of complexity in the thousands to one in, say, the tens, even then the Committee should not answer it literally. For to analyze any one combination of equipment class, unit size, and scenario in the depth required to provide a responsive reply is in itself a very detailed and time consuming task.

The Committee therefore believes it can best carry out its function by observing and commenting upon the manner in which the Army is now attempting to carry out this task. Indeed, this is the normal mode of operation of ASAP ad hoc committees.

In the overview, the Army has taken a first, albeit small, step toward solving this complex question. Through a "measures of effectiveness offensive" established by LTG John Norton at Combat Developments Command, the Scientific Services Directorate of the Systems Analysis

Group at CDC is spearheading a drive to compile all the MOE in use by the Army, as an aid to the action officers overlooking the many program areas. The known MOE, not limited only to the STANO test task but to all Army tasks, are in the hundreds. At its present stage, there will be no defining of new MOE; rather there will be a compilation of the established ones and a presentation of these in a common format.

Thus the MOE offensive is but a first step, the results of which are applicable but in no way solve the task set forth before the Committee. Because the report of the group conducting the MOE offensive has yet to be issued, the Committee has not attempted to assess it.

Because of the complexity and enormity of the task, and the dearth of work within the Army bearing on it upon which the Committee can comment, the Committee considers that it can best carry out its function by presenting a broad overview of the manner in which the Army should address the task.

Consider the action officer who is asssgned the task of ascertaining whether or not a newly developed item of STANO equipment is of sufficient value to the Army for its cost in dollars, maintenance, logistics, etc., to warrant production. It is known that the item conforms to the specifications required by the contract under which it was developed, having passed engineering and service tests. But is it really of use to the squad, section, platoon, etc., in light of the fact that these men are already burdened, specifically by the weight they must carry and generally by the requirement to provide trained operator, maintenance personnel, logistic support, and so on.

A compilation of established MOE would be of some value to the action officer. Yet his ultimate decision must today involve a great deal of personal judgment. Assume that in reviewing the MOE he finds, say, ten applicable to the item. For example, if the item were a new type of night vision device, the applicable MOE might be (a) range at which 50% of individuals detect a tank-type target under quarter moonlight conditions; (b) time required for identification under quarter moonlight conditions after tank-type target enters unit area of responsibility; (c) range at which 50% of individuals identify a tank-type target under quarter moonlight conditions, and so on. He then sets up a field test at, say, the squad level, being faced with a choice of a selection of "representative" scenarios. Following this, he conducts the exercise and obtains numerical values for the MOE.

What then? He is faced with the very difficult task of assessing the relative importance of the various MOE. Assume in the example above that the detection range was high compared to a similar piece of STANO equipment, but that the identification time was low as was the identification range. This might be the case for a night vision device of high photosensitivity but poor spatial resolution. Is it important to have detection at long range if the identification range is short? There is of course no standard answer; it must depend upon the scenario. For example, if the location of all friendly forces is known with high accuracy, so that any intruder can be assumed to be an enemy, then detection range is paramount. But if friendly and enemy forces are thoroughly mixed, then it becomes vital to identify the target before engaging it.

Thus the question of determining the relative importance or weight to be attached to the various MOE is highly sensitive to the scenario. What then are the relationships between the MOE and the scenario? First there is the problem of determining the functional interrelationships between the MOE. This is equivalent to defining an equation interrelating independent variables. One method under investigation is to construct a linear weighted sum. That is to say, the numerical value of each MOE is multiplied by a normalized number, and the weighted values are summed. If the sum exceeds a threshold value, or else if the sum exceeds that of a similar item under test, then the item in question is acceptable or is better than the other one. Here the crucial problem is determining the proper weights. The unusual method is to seek the collective opinions of experienced officers, either through direct questions or by indirect methods such as are employed in utility theory. However, the linear sum is not necessarily the proper interrelationship between the MOE. For a given scenario, there exists some proper mathematical relationship between the MOE which describes the performance of the system under test. It may be, for example, a product of probabilities, as the kill probability of a given weapon is the product of probabilities. Or it may be some other functional relationship.

In any event, the functional relationship between the MOE to satisfy a given scenario amounts to an equation with the MOE as independent variables. It is well known that n equations are required to solve for n independent variables. Therefore, it appears that if there are n clearly different scenarios which describe the possible operations of a given unit equipped with a given device to be tested, then that device must be described by n MOE. Although this seems apparent, the Committee is not aware that it is clearly realized by the Army.

In summary then, the problem of adequately distinguishing the worth of competing STANO items in use by a unit of a given size requires the following steps:

1. Determine a list of scenarios which adequately describe the most important ways by which a unit will employ the items.
2. Choose MOE capable of being experimentally measured which are independent of each other and whose number equals the number of scenarios.
3. Determine the functional interrelationships or equations which connect the MOE or independent variables, one such interrelationship for each scenario.

The Committee wishes to emphasize that this task is a very large and difficult one. There is no simple solution. Thus a recommendation of the Committee is as follows: Establish a working group within the Army whose task it is to study the theory of testing along the lines set forth above. This group might well be an outgrowth of the effort already underway at the Scientific Services Directorate of the Systems Analysis Group of CDC. It should consist of mathematically and theoretically inclined systems analysts, Army officers experienced in testing, and scientists familiar with the technical operation of the

STANO devices under consideration. This group should be adequately funded over a long term, say, five years or more. Much of the effort should be supported by contracts with civilian organizations experienced in systems analysis and testing. Expect no near term results. Yet if such a broadly based long term effort is mounted and carried out, the long term payoffs of this new "measurement of effectiveness offensive" in reducing costly mistakes in the acquisition process may well be orders of magnitude greater than the expenditure.

ANNEX H

THE USE OF A DELPHI SYSTEM FOR ARMY TEST ENVIRONMENTS

V. S. Haneman, Jr.

The Delphi System was originally applied to Technological Forecasting and has had extensive use by such firms as Thompson Ramo Wooldridge, Rand, Ling Temco Vought, Litton Industries, and others, for studies of future markets. Oklahoma State University has participated in one such effort with Ling Temco Vought. The principle involves the use of a large number of qualified experts who are requested to give their estimation of the date of a series of future accomplishments which hinge heavily on technological development in their professional field. An average date and a statistical evaluation is performed upon the data. To eliminate the effect where one person may assume the role of leader and thus influence the others unduly, the experts are provided the statistical picture with no identification of any individual's selections. The same question is asked and a second replication is performed. To achieve a reasonable settled selection, the experience indicates that a total of three or four replications are required before there is only minor change in selection.

The major problems with this technique is that it is a public opinion poll, even though it is recognized that this public is a selected group of experts whose opinions may be given more credence than the average person. Examination of the details of the efforts by the corporation and by Oklahoma State University indicate that individuals and groups of individuals may have specific biases which might override technological data that would deny the choice made.

To relieve this limitation, the examiner or the test conductor may wish to set up control devices to ferret out such effects. If the system is applied, the user must be aware of potential limitations.

To apply this to an Army test environment let us use an example, the evaluation of night viewing equipment assigned to a squad or platoon. The first question is "what is the mix of night viewing devices to be issued?" This should then be followed by the question of: "how to test the squad?" Since each man may have a different experience with the various pieces of equipment, some good, some poor, and some very poor, "how will this effect the results?"

The modified Delphi approach may be applied here in that the users are the members of the squad and are the "experts" since it will be their use of the equipment which will provide the basis of evaluation. A first step might be to allow the members of the test squad to select their own mix. They would then be tested in a scenario which would be replicated for several other squads who would also have selected their equipment. The squads records might be identified by unit and by mix of equipment. The statistical test data would be circulated to each squad who would be allowed to reselect equipment mix for a second test. This would be repeated a number of times over a number of different scenarios to obtain a test of the different problems to be encountered, such as detection range, portability, and meantime to failure.

Limitations would include the situation wherein the squad leader dominated, and regardless of all data to the contrary, he selected what he had known from some previous experience. It is also possible that the members of the test units were not motivated to succeed or to better the record of the other unit.

The end result would be a record of a number of units with different mixes of night viewing devices, hopefully, settled toward a set distribution of equipment, used under a variety of conditions.

This technique may be applicable to situations where choice,uses motivation, and test scenarios would effect the test results. Further examples might include STANO devices (UGS), hand weapons, and even armour units.