Field Test Designs for the Evaluation of Vessel Operators' Use of and Confidence in Various Aids to Navigation

E. Donald Sussman, PhD
Robert DiSario, PhD
Michael J. Sheehan

Jay Corey (EG&G/UNISYS)

PROPOSED TEST DESIGN
December 1995

Prepared for:
UNITED STATES COAST GUARD
RESEARCH AND DEVELOPMENT CENTER
NAVIGATION SYSTEMS BRANCH

This Document Contains Preliminary Information Subject to Change. It Is Considered an Internal Volpe Center Working Paper with a Select Distribution Controlled by the Project Manager. It Is Not a Formal Referable Report.

U.S. DEPARTMENT OF TRANSPORTATION
RESEARCH AND SPECIAL PROGRAMS ADMINISTRATION
VOLPE NATIONAL TRANSPORTATION SYSTEMS CENTER
CAMBRIDGE MA 02142

Table of Contents

1.0 INTRODUCTION

As part of the multi-year Volpe Center project to assist and support the Coast Guard Research and Development Center in evaluating current Coast Guard policies and development plans for Waterways Management, a "Waterways Management Research and Planning" task plan was established as Task Area A in FY '94. The purpose of this task area is to develop decision support tools to assist Coast Guard waterways managers in improving the way they meet the waterway users' requirements for aids to navigation and waterway services. A Waterways Users research topic was established in 1994 and a Baseline Analysis report on Waterway Users was presented in April, 1995. This analysis defined the waterway users to be "vessels of various types, sizes and uses, not the humans operating the vessels", and went on to consider the most important attributes of the vessels. The attributes were defined, for the most part, in physical terms (size and number of vessels, cargo carried, maneuverability, etc.) but consideration was given to the human dimension as well. One of the research tasks proposed in the Baseline Analysis was Task 2-6, User Requirements - Human Factors Critical to Assessment of Effectiveness of Coast Guard Aids to Navigation and Waterways Services. That Task, in turn, was divided into two subtasks, 2-6.1 Define Waterway Users by Navigational Requirements and 2-6.2 Select User Groups for In-Depth Study, which were funded in FY '95.

This proposed test design is a product of subtask 2-6.1. It provides "generic" guidance for field testing mariners' practical use of alternative Aids to Navigation (ATON). It allows evaluation of the utility of the systems relative to the accuracy, ease, and safety of navigation afforded by each. It also considers the confidence of the mariner in the use of the particular ATON. Two alternative approaches are proposed: field tests under operational conditions, using mariners operating their own vessels on operational routes, and field tests under non-operational conditions, where mariners perform navigational tasks in a controlled environment. In both approaches the focus is on obtaining data on the performance of mariners and on the potential acceptance of such systems by mariners. Either approach will require a test team capable of performing the observations, recording data, and interviewing vessel operators. Teams of at least two persons with sufficient understanding of navigation to understand the operators' actions as well as basic familiarity with data gathering techniques will be required to carry out either approach. Subsequent to a review by the Coast Guard, we assume that one of the approaches will be selected and more guidance will be provided on the selected aproach.

1.1 BACKGROUND

New innovative Aid to Navigation (ATON) technologies show significant potential for providing equal or better support to maritime navigation while reducing the cost of operating these ATONs. To determine if these technologies meet this potential for various groups of waterway users, the Coast Guard must systematically and objectively evaluate the utility of these proposed systems relative to currently used systems. The evaluation should consider the extent to which relying on these new technologies will improve the accuracy, ease, and safety of navigation. It should also consider the impact of the confidence of the mariner in the use of the ATON. This can be accomplished through field and laboratory studies. In these studies two classes of data are critical: the performance of mariners and the potential acceptance of such systems by mariners.

Data on the relative performance of mariners using proposed new systems and existing systems:

- Position accuracy achieved through use of the ATON
- Proximity of the vessels to obstructions or hazards
- Workload imposed by the use of the ATON
- Situation awareness of the user, as a function of using the ATON

Data on the potential acceptance of such systems. This data can be achieved through questionnaires to determine the extent to which mariners have confidence in the:

- Usability and accuracy of the navigation information derived from the ATONs and
- The timely availability of the information under all conditions particularly during severe weather conditions

2.0 METHODS

Studies to evaluate the relative performance of new and existing ATON can be conducted under two general paradigms:

- field tests under operational conditions, or
- field tests under non-operational conditions

The following are descriptions of the experimental methods required to gather the types of data listed above.

2.1 Field tests conducted under operational conditions

This type of test offers the "face validity" often required to gain acceptance for a new process, but field tests are costly, difficult to schedule, potentially hazardous and offer limited opportunities for comprehensive experimental designs.

2.1.1 Subjects

The subjects in such studies should be fully qualified licensed mariners. They should be familiar with the waters, routes, vessel, procedures and ATON for the test area. They can be recruited through any or all of the following methods:

- meetings and negotiations with vessel owners
- meetings and negotiations with professional associations or labor organizations
- solicitations in trade newsletters

Because these subjects will be participating in the tests as part of their normal duties they will not require salary payment during testing but will require some form of incentive reimbursement to maintain a uniform and high level of enthusiastic cooperation. Often savings bonds or gift certificates with their magnitude linked to the subjects performance can be used successfully.

2.1.2 Test Equipment

Equipment is generally divided into two groups: equipment needed to run the tests and equipment needed to record and analyze the data.

<u>Equipment required to run the tests</u>

Test vessels - Operational field tests should be conducted using the vessels normally in such service. The vessels should be crewed and operated in the service modes that represent actual revenue operations

The following describes the equipment required to conduct operational field tests of the relative merits, for navigation on the Hudson River, of floating ATON as compared to advanced technology fixed shore based navigational ranges.

As noted above, the studies should be performed with qualified tugboat captains and crews operating tugs as equipped with the navigation and other safety equipment customarily found on such vessel. The vessels should also be operated with the "tows" customary for such service

- Radar Systems (primary and backup) high quality commercial 24 mile range systems capable of maintaining shore fixes
- VHF radio systems (primary and backup)

- Remotely controlled high power spotlight
- Loud hailer
- Horn
- Magnetic compass
- Annotated charts

Some common types of navigational gear are not used on all tugs and therefore are not required for the study: Fathometer, GPS/DGPS, Loran-C, GPS/Loran linked electronic chart.

<u>Equipment required to collect data</u>
A portable Differential Global Positioning System receiver (DGPS) linked with a portable computer. This linked system should have the capacity to acquire, record, and play back the recordings on a graphic depiction of the course. It should be capable of acquiring

- course and position, (<10 meter accuracy)
- track,
- heading,
- course made good (CMG) to a waypoint,
- velocity made good (VOG)
- speed over the ground (SOG)
- time of arrival at each way point, and
- bearing to each waypoint

  (the waypoint may be either the actual position of a floating ATON or a theoretical turning point which was derived from the mariner's expected use of the range)

A portable video / audio recording system. This system is to be used in gauging impact of the use of the ATON on the mariner's workload. It will also be useful in documenting the ways in which the mariner uses the ATON. It will be used to make a taped record of the mariner as he or she operates the vessel.

- The system should be portable and unobtrusive.
- It should be equipped with an infrared camera and light source suitable for recording in almost total darkness without impacting the night vision of the subjects.

Cooperating individuals or organizations (the owners, crew or labor unions representing the crew) may object to such video recordings. If such objections cannot be overcome then data on work load and use of the ATON will have to be taken through recording of direct observation as made by the individuals conducting the experiment.

Questionnaire(s) oral and/or written should be used to assess perceived differences in workload and situational awareness which may result from the use of the new ATON.

2.1.3 Test sites and test conditions

To permit adequate comparisons between the types of ATON appropriate tests sites or venues must be established. In field experiments where conditions are extremely difficult to control it is usually impossible to conduct "full factorial" type experiment. If it is not possible to produce all the conditions under which the ATON are to be used, at a minimum, testing should be conducted in the "worst conditions". In an operational field test, however, the absence of a broad range of features may limit the generalizationality of the study. The most significant variable with regard to the choice of the test site(s) is the availability of each type of ATON. Of course, if a high-technology ATON (e.g., a laser range) is chosen for the test, the cost of temporary installation at the test site must be factored in.

There are two alternative ways to insure that each type of ATON can be tested:

- identify sets of comparable or equivalent test areas which differ mainly in the types of ATONs present.

  In this hypothetical case they would contain predominantly either conventional floating ATONs or the advanced ranges to which they are to be compared.
- use test areas where each type of ATON is used seasonally

Because of the great value of being able to use a DGPS system for data recording the test site should be within an area of good DGPS coverage.

As noted above tests conditions should include challenging or difficult meteorological situations particularly those involving poor visibility. If a choice between testing under excellent conditions and challenging conditions must be made testing should be conducted (consistent with safety) in the poorer conditions because it is under just such conditions that usefulness of the ATONs will be critical. Therefore where possible, (and prudent) tests should be conducted in any or all of the following:

- fog
- darkness
- snow
- glare
- spray
- rain

2.1.4 Statistical Design

Field test under operational conditions, limits the experimenter's ability to control independent variables. The levels of the variables that can be tested are mainly a function of the conditions available in the field test site chosen. Further, some of the independent variables (e.g. visibility, sea state, or wind velocity) can not be consistently reproduced from trial to trial. Finally, even though testing should be done under conditions where the use of the ATON will be most difficult, safety must be the prime consideration in determining test conditions.

Independent Variables (factors that are expected to influence the performance of the subjects)

- Type of ATON
- Course to be followed
- Traffic
- Meteorological conditions
  - Visibility
  - Sea state
  - Wind velocity
  - Tide

Field testing under operational conditions limits the ways in which data may be taken. For example, there is currently great interest in *situation awareness*. It is assumed that the subject makes a cognitive map of the conditions under which he or she is operating. Assessment of the instantaneous accuracy of that map can provide valuable insights into how well a source of information such as an ATON system is being used. The conventional method of gathering such data is to stop the subject at or immediately after some critical point in the study and asses his or her understanding of the conditions. There may be few opportunities for this in an operational field test. The acquisition of such data during operations may be too disruptive. Some of this information can be reconstructed by debriefing the subject during periods of relative inactivity, but it will not have the same quality as data taken at the moment that the map has been developed.

Similarly, objective indicators of workload can be obtained by recording performance on secondary tasks, but in operational field testing the ability to manipulate the secondary task relative to the primary navigation tasks is limited. Subjective workload data can be gathered through debriefing and questionnaires during periods of relative inactivity or after the completion of a watch.

Dependent Variables (Measures of the performance of the subject)

- Ability to accurately follow course
- Time required to complete course
- Confidence in ATON system
- Indicators of situation awareness
- Indicators of workload

In order to reduce variance a repeated measures design should be used. In this design the subject acts as his or her own control. He or she experiences each of the experimental conditions but the sequence is counterbalanced between subjects to reduce the impact of *order effects*. In this type of experiment it is likely that there will be limits on the available subject/equipment pool and restrictions on the periods that the ATON will be available. Given such limits, counterbalancing is preferable to randomizing.

A counterbalanced design also provides flexibility, allowing the experimenter to take advantage of particular meteorological conditions. In the simplest case where tugboat masters are being tested on a river course, testing for subjects A, C, E, and G might begin heading up river while testing for B, D, F, and H might begin heading down stream. Where both types of ATON are not available at the same time/season on the same course, it may not be possible to counterbalance the order of testing. In operational testing this is not a major obstacle because it is assumed that the subjects (tugboat masters) are already very familiar with the routes.

Unlike a laboratory experiment a *full factorial experiment* (where subjects are tested under all combinations and levels of independent variables) is neither practical nor mandatory. If we assume that the purpose of these studies is to determine if and under what conditions new and more economical ATON do not allow the mariner to navigate as well as the existing ATON, an experimental design which only permits comparisons between individual treatments (combinations of experimental variables) may be sufficient. In such testing some statistical sensitivity is lost because of the statistical correction required based on conducting multiple tests. A more serious problem would be the inability to distinguish between the effect of ATON type and the interaction between an ATON type and a second factor (e.g., weather). This design will not support the analysis of interactions between variables. The following tables illustrate examples of the basic design.

Table 1. Design for evaluation of ATON type under different meteorological conditions for course 1 as a function of position accuracy.

	Meteorological Condition					
ATON Type	**Fog**	**Darkness**	**Snow**	**Glare**	**Spray**	**Rain**
Conventional Floating						
Range						

Table 2. Design for evaluation of ATON type under different meteorological conditions for course 2 as a function of position accuracy.

	Meteorological Condition					
ATON Type	**Fog**	**Darkness**	**Snow**	**Glare**	**Spray**	**Rain**
Conventional Floating						
Range						

Each of these designs would permit comparisons of the mariners performance in maintaining position accuracy between two ATON system types, to the extent that all of the meteorological conditions occurred during the test period. Another factor that may be considered is the operator's level of experience on the waterway. However, the courses are likely to differ in factors such as geometry, navigational difficulty and traffic density. If so, aggregating the data to obtain a generalizable result will not be meaningful. This is a problem if a general rule is sought but not if the data is to be used to make decisions on ATON for given localities.

As noted above the location for navigational courses must make use of areas where each type of ATON is present. However the use of the Hudson River for operational testing has some disadvantages. Discussions with tug captains indicates that tug barge navigation there is performed on a shore to shore basis. The relatively shallow draft of tug barges allows the operators to use much more of the river than they would if restricted by draft. Further, shoaling does not impact Hudson river operations to the extent that it does on many other river systems. Finally for most of the river traffic is not particularly dense and ATON are relatively widely spaced. On the other hand, there is a seasonal change in ATON availability in the river, since the buoys are pulled in the winter due to icing. This may present the opportunity to structure field testing with and without buoys over the same waterway. Contacts with the cognizant Aids to Navigation Teams along the Hudson indicate that as of Early December, approximately 65 seasonal buoys are replaced by ice buoys, while three are pulled out without replacement.

Realistically, rivers with greater traffic density and a broader mixture of ATON should be given priority consideration for use in operational testing. The Mississippi system is a far more real-world test environment for tug/barge operations than the Hudson.

2.1.5 Analysis

Objective Data:
Analysis of the objective data could be accomplished using parametric techniques as simple as one-tailed T-tests. Course deviation error can be calculated as the average absolute deviation from a desired course or as the average of the squared difference from the course. This statistic should be calculated over the course as a whole and for individual legs of the course believed to be particularly critical and or difficult. Time to complete the course and time to complete each leg can be dealt with similarly.

Subjective Data
As noted above the following subjective measures should be taken: confidence in the ATON system, situation awareness, and workload. It is important to get quantitative data on what are basically subjective feelings. Two straightforward methods are the *Lickert* scale, and anchored scaling techniques.
The following are two examples of questions to determine the mariners beliefs about the navigational difficulty of a course segment (similar questions would be constructed dealing with workload).

A Lickert type question would look like this:
On Today's trip, using ranges only to navigate the segment between the Tappan Zee Bridge and the Bear Mountain Bridge was very difficult.

Strongly Disagree	Disagree	Neither Agree Nor Disagree	Mildly Disagree	Agree

An anchored scaling technique the question would look like this:
On Today's trip, using ranges only to navigate the segment between the Tappan Zee Bridge and the Bear Mountain Bridge was:

Very Difficult	Difficult	Slightly Difficult	Neither Difficult Nor Easy	Slightly Easy	Easy	Very Easy

These are just examples of two ways of asking the same question, each has its advantages. Of the two the *anchored scaling technique* is favored because it provides broader response scale. Individuals have idiosyncratic ways of responding to such scales. Some subjects may be conservative and consistently restrict or cluster their responses, using only some of points. Other subjects may be expansive and use only the extreme ends of the scale Further, even though the scale provides a clear mid-point some respondents will consistently center their responses nearer to the negative (Very Difficult) or the positive (Very Easy) end of the scale. To some extent these tendencies can be corrected for by calculating a Z-score for each subject and dividing his or her scores by the Z-score. In each case it is very important to provide an anchor for the subject. In preparation for the study all subjects should be exposed to segments that for the purposes of the study are clearly labeled as *difficult* or *easy*.

To extent possible the subjective data (confidence in the ATON system, situation awareness, and workload) should be gathered as soon as possible after the completion of a segment. Where practical the subject should be queried in the first quiet period after completing each segment. If this is not possible the subject should be interviewed after completion of his or her watch. The maneuvers of the vessel as recorded using the DGPS-portable computer should be played back to the subject so it can be used as a memory aid.

2.1.6 Cost Elements

Table 3. Cost Elements required for conduct of operational field test

Category	**Cost Element**
Labor	Final experimental design
	Development and installation of recording gear
	Negotiation with participants
	Selection of site(s)
	Scheduling of tests
	Conduct of the studies
	Data analysis
	Report preparation
Equipment	Acquisition of recording gear
	development and refinement of questionnaires
	Installation of temporary laser range(s) or other test ATON
Travel	Accommodation of experimenters at test site
	Meetings with CG staff*
Recruitment and payment of subjects	Advertisements
	Incentive pay (premiums for participants)
	Scheduling

* not required if in-house contractor or CG staff conducts study

2.2 Field tests conducted under non-operational conditions

This type of test offers less "face validity" but far more experimental control than operational field tests. This type of test reduces the schedule difficulties, and potential hazards, offers much better opportunity for comprehensive experimental design but may actually be more costly. Increased costs over operational field tests can be expected from three sources: pay for the subjects, the requirement for preparing a test range, and the cost of the vessels used in the tests.

2.2.1 Subjects

The subjects in such studies should be fully qualified licensed mariners. They should be familiar with the waters, routes, vessel, procedures and ATON. If a special test area is developed for the tests they will require familiarization runs before the conduct of the study.

Subjects can be recruited through any or all of the following methods:

- meetings and negotiations with vessel owners
- meetings and negotiations with professional associations or labor organizations
- solicitations in trade newsletters

Because these subjects will not be participating in the tests as part of their normal duties they will require salary payment during testing which includes some form of incentive reimbursement to maintain a uniform and high level of enthusiastic cooperation.

2.2.2 Test Equipment

Equipment is generally divided into two groups: equipment needed to run the tests and equipment needed to record and analyze the data.

<u>Equipment required to run the tests</u>

Test vessels, non-operational field tests can be conducted without using the vessels normally found in such service. If all tests are conducted using the same vessel or "sisterships" the relative comparisons between ATON systems should remain valid. The vessels chosen should be seaworthy, reliable, offer weather protection and be capable of being equipped with the equipment listed above. Selection of the vessel should also consider the ready availability of back-up vessels to ensure that experiments do not need to be aborted due to the lack of availability of the test vessel. Smaller size vessels, consistent with seaworthiness, offer the advantages of economy both in terms of fuel costs and crew size. In particular vessels with relatively shallow drafts will permit the use or construction of test courses outside of the main traffic channels.

The following describes the equipment required to conduct non-operational field tests of the relative merits, for navigation on a course established on the Hudson River, of floating ATON as compared to advanced technology fixed shore based navigational ranges.

The studies should be performed with qualified tugboat captains and crews operating small to moderate sized vessels equipped with the navigation and other safety equipment customarily found on tugs operating in the area:

- Radar Systems (primary and backup) high quality commercial 24 mile range systems capable of maintaining shore fixes
- VHF radio systems (primary and backup)

- Remotely controlled high power spotlight
- Loud hailer
- Horn
- Magnetic compass
- Annotated charts

Some common types of navigational gear not used on all tugs are likely to be found on many other vessels: Fathometer, GPS/DGPS, Loran-C, GPS/Loran linked electronic chart. Their use could compromise the study and so they should be moved or obscured from the subjects.

Equipment required to collect data
A portable DGPS linked with a portable computer with the capacity to acquire, record, and play back:

- course and position, (<10 meter accuracy)
- track,
- heading,
- course made good (CMG) to a waypoint,
- velocity made good (VOG)
- speed over the ground (SOG)
- time of arrival at each way point, and
- bearing to each waypoint

  (the waypoint may be either the actual position of a floating ATON or a theoretical turning point which was derived from the mariner's expected use of the range)

A portable video / audio recording system This system is to be used in gauging impact of the use of the ATON on the mariner's workload. It will also be useful in documenting the ways in which the mariner uses the ATON. It will be used to make a taped record of the mariner as he or she operates the vessel.

- The system should be portable and unobtrusive.
- It should be equipped with an infrared camera and light source suitable for recording in almost total darkness without impacting the night vision of the subjects.

Questionnaire(s) oral and/or written should be used to assess perceived differences in workload and situational awareness which may result from the use of the new ATON.

2.2.3 Test sites and test conditions

To permit adequate comparisons between the types of ATON appropriate tests sites or venues must be established, prepared, or constructed. In such field experiments conditions are far easier to control. Variability in meteorological conditions will make the conduct of "factorial" type experiments difficult but the greater flexibility (as compared to operational testing) should allow more complete experimental designs. However, as noted above, where it is not possible to produce all the conditions under which the ATON are to be used testing should be conducted in the "worst conditions".

The most significant variable with regard to the test site is the availability of each type of ATON or the capability to install each type of ATON for the purposes of the study.

There are three alternative ways to insure that each type of ATON can be tested:

- identify sets of comparable or equivalent test areas which differ mainly in the types of ATON present,

  In this hypothetical case they would contain predominantly either conventional floating ATON or the advanced ranges to which they are to be compared.

- use test areas where each type of ATON is used seasonally, or identify a test site and equip it with the ATON to be compared.

Because of the great value of being able to use a DGPS system for data recording the test site should be within an area of good DGPS coverage.

As noted above test conditions should include challenging or difficult meteorological situations particularly those involving poor visibility. If a choice between testing under excellent conditions and challenging conditions must be made testing should be conducted (consistent with safety) in the poorer conditions because it is under just such conditions that usefulness of the ATON will be critical. Therefore where possible, (and prudent) tests should be conducted in any or all of the following:

- fog
- darkness
- snow
- glare
- spray
- rain

2.2.4 Statistical Design

Field test under non-operational conditions impose fewer limits on the experimenter's ability to control independent variables. The levels of the variables that can be tested are still mainly a function of the conditions found in the field test site chosen. The experimenter now can develop a test course, specify the vessel type to be used, and arrange specific test sequence orders. While it remains true that some of the independent variables (e.g. visibility, sea state, or wind velocity) can not be consistently reproduced from trial to trial the experimenters' flexibility to make use of conditions as they occur should be much greater. Testing should still be done under conditions where the use of the ATON will be most difficult but the safety problems will be reduced by not having to maneuver loaded barges or vessels with ten foot drafts. This should permit a broader range of test conditions.

Independent Variables (factors that are expected to influence the performance of the subjects)

- Type of ATON
- Course to be followed
- Traffic
- Meteorological conditions
    - Visibility
    - Sea state
    - Wind velocity
    - Tide

Field testing under non-operational conditions sets fewer limits on the ways in which data may be taken. Assessment of the instantaneous accuracy of *situation awareness* (described above) can be accomplished by relieving the subject at the helm during or immediately after some critical point in the test segment and assessing his or her understanding of the conditions through a questionnaire. Opportunities for this can be built into non-operational field tests. The acquisition of such data during operations may be somewhat disruptive but it will be of much greater value than situational awareness estimates reconstructed from debriefings of the subject during periods of relative inactivity occurring well after the critical maneuver.

Objective indicators of workload can be obtained by recording performance on secondary tasks. Non-operational field testing provides the opportunity to develop and manipulate the secondary tasks without compromising navigational safety. Objective measures of workload can be compared with subjective workload data gathered through debriefing and questionnaires during periods of relative inactivity or after the completion of a watch. Confidence in the ATON system cannot be measured directly, it must be assessed through subject interviews or responses to questionnaires.

Dependent Variables (Measures of the performance of the subject)

- Ability to accurately follow course
- Time required to complete course
- Situation awareness
- Workload
- Confidence in ATON system

In order to reduce variance a repeated measures design should be used. In this design the subject acts as his or her own control. He or she experiences each of the experimental conditions but the sequence is counterbalanced between subjects to reduce the impact of *order effects*. In non-operational field tests it is likely that there still will be limits on the available subject/equipment pool and restrictions on the periods that the ATON will be available. Given such limits, counterbalancing remains preferable to randomizing.

As noted above a counter balanced design provides flexibility, allowing the experimenter to take advantage of particular meteorological conditions. In the simplest case where tugboat masters are being tested on a river course, testing for subjects A, C, E, and G might begin heading up river while testing for B, D, F, and H might begin heading down stream. Where both types of ATON are not available at the same time/season on the same course, it is not possible to counterbalance the order of testing. In operational testing this is not a major obstacle because it is assumed that the subjects (tugboat masters) are already very familiar with the routes.

Unlike a laboratory experiment a *full factorial experiment* (where subjects are tested under all combinations and levels of independent variables) is neither practical nor mandatory. If we assume that the purpose of these studies is to determine if and under what conditions new and more economical ATON do not allow the mariner to navigate as well as the existing ATON, an experimental design which permits comparisons between treatments (combinations of experimental variables) may be sufficient. A non-operational field test provides more control of the experimental variables than an operational test, but the control is not absolute. Such an experiment will support testing hypotheses as to the differential efficacy of the ATON systems under various conditions. In such testing some statistical sensitivity is lost because of the statistical correction required based on conducting multiple tests. Conversely where a-priori assumptions are reasonable more sensitive *one-tailed* tests can be used. Even though much better control of independent variables is possible under this condition than in operational conditions the use of techniques such as analysis of variance is limited because the levels of the independent variables must be considered fixed rather than continuous.

The following table illustrates examples of the basic design.

Table 4. Design for evaluation of ATON type under different meteorological conditions and segments as a function of position accuracy.

Segment	ATON Type	Meteorological Condition: Fog	Darkness	Snow	Glare	Spray	Rain
Segment 1	Conventional Floating						
	Range						
Segment 2	Conventional Floating						
	Range						
Segment 3	Conventional Floating						
	Range						
Segment 4	Conventional Floating						
	Range						

This type of design permits comparisons of the mariners performance in maintaining position accuracy between two ATON system types across segments (to the extent that all of the meteorological conditions occurred during the test period). Because the courses will be designed not to differ in factors such as geometry, navigational difficulty and traffic density aggregating the data to obtain more generalizable results will be possible. To be meaningful the courses constructed will have to be made up of segments which include a broad representation of the critical navigation challenges found in the "real world." This technique will support the development of general rules, and to the extent that the segments adequately represent those found in particular local applications will support decisions concerned with changes in ATON for these localities.

2.2.5 Analysis

Objective Data:

Analysis of the objective data could be accomplished using parametric techniques as simple as one-tailed T-tests, however a *Studentized Range Statistic* will provide information about the performance of mariners on segments which differ in critical aspects.

Course deviation error can be calculated as the average absolute deviation from a desired course or as the average of the squared difference from the course. This statistic should be calculated for individual legs of the course. Elapsed time to complete each leg can be dealt with similarly. Workload can be assessed by recording the hit rate for visual signals which occur in the peripheral visual field of the subject. Situation awareness can be assessed by questioning the subject about important aspects about the vessel (e.g. position, course, velocity . . .), other traffic, and sea state (e.g. current velocity, tidal direction . . .) and scoring the number of correct responses.

Subjective Data

As noted above the following subjective measures should be taken: confidence in the ATON system, situation awareness, and workload. Even though in this experimental paradigm it should be possible to obtain direct quantitative data on situation awareness, and workload it is also important to get quantitative estimates of the subjects' subjective feelings on these dependent variables. As noted above two straightforward methods are the *Lickert* scale, and anchored scaling techniques.

To the extent possible the subjective data (confidence in the ATON system, situation awareness, and workload) should be gathered as soon as possible after the completion of a segment. Where practical the subject should be queried in the first quiet period after completing each segment. If this is not possible the subject should be interviewed after completion of his or her watch. The maneuvers of the vessel as recorded using the DGPS-portable computer should be played back to the subject so it can be used as a memory aid.

2.2.6 Cost Elements

Table 5. Cost Elements required for conduct of non-operational field test

Category	Cost Element
Labor	Final experimental design
	Development and installation of recording gear
	Selection of site(s)
	Design of segments
	Scheduling of tests
	Conduct of the studies
	Data analysis
	Report preparation
Equipment	Acquisition of recording gear
	Development and refinement of questionnaires
	Acquisition of ATON
	Rental of test vessels
	Installation / removal of ATON
Travel	Accommodations of experimenters at test site
	Meetings with CG staff*
Recruitment and payment of subjects	Advertisements
	Subject salaries
	Scheduling

* not required if in-house contractor or CG staff conducts study

3.0 Summary Table

Table 6 Summarizes the major factors that differentiate between the Operational and Non-operational field test approaches.

Table 6. Operational and Non-Operational Approaches -Differential Considerations

	Consideration	Operational	Non-operational
Costs	Differential Cost Element	Instalation of specific ATON	Establish ATON range, Salary for subjects, Use of vessels
Independent Variable	Ability to use same course for both classes of ATON	Weak	Strong
	"Control" of Meteorological/sea-state conditions	Minimal	Some
Dependent Variables	Course Tracking	Influenced by unrelated/uncontrolled factors e.g. traffic	Good
	Elapsed time	Influenced by unrelated/uncontrolled factors e.g. cargo	Good
	Confidence in ATON	Good	Good
	Situation awareness	Very weak due to delay	Good
	Workload	Weak due to delay	Fair because of use of objective measures
Overall Considerations	Experimental rigor	Good face validity	Enhanced experimental control
	Applicability of results to specific sites	Good	Fair
	Generalizablity to all sites	Weak	Good