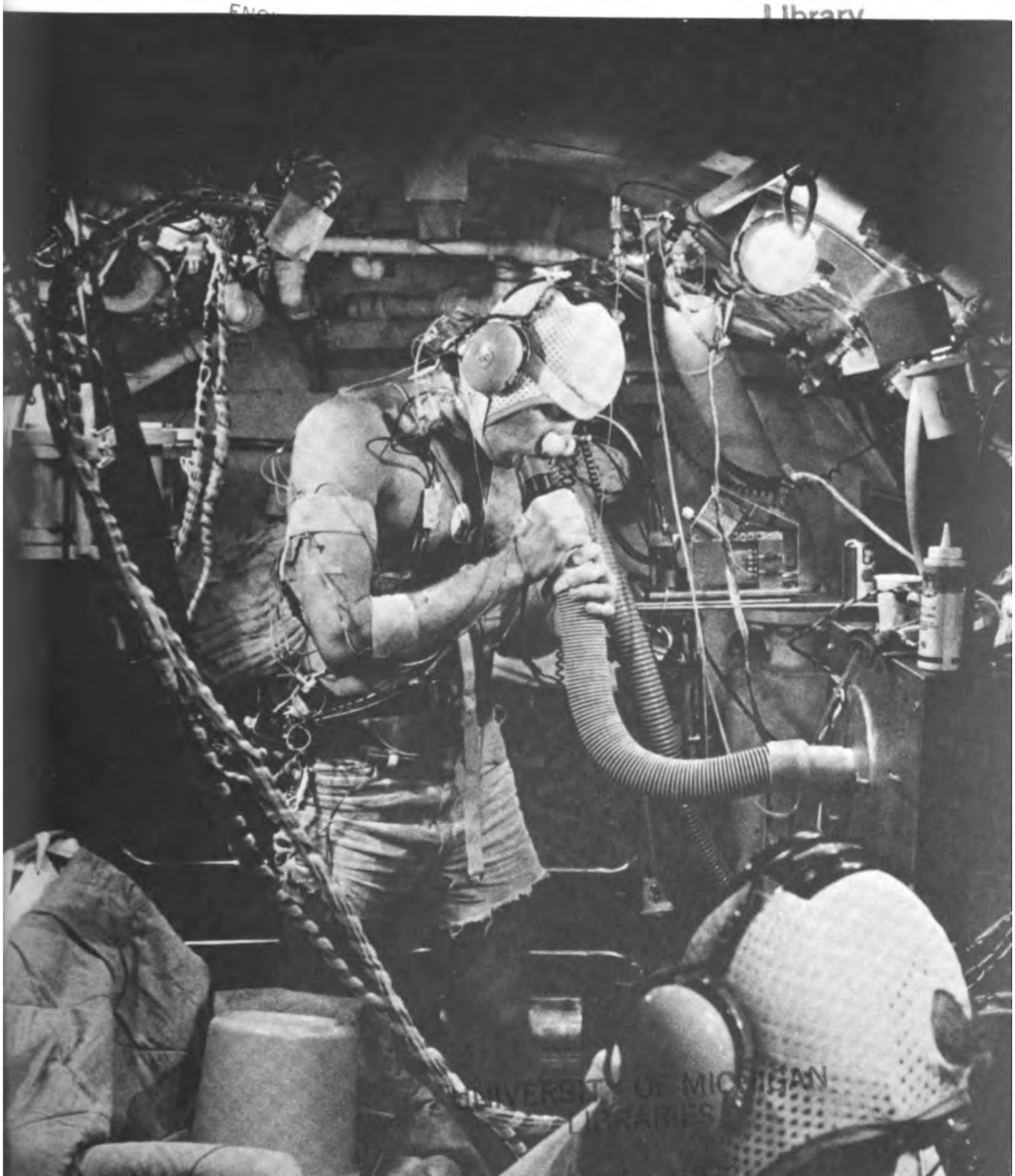
U.S. Office of Naval Research.

Engin

NAVAL RESEARCH

REVIEWS

NOVEMBER 1975



PROFILES IN SCIENCE

Dr. Paul M. Fye

Paul M. Fye is Director and President of Woods Hole Oceanographic Institution. A physical chemist, Dr. Fye worked in the areas of gas kinetics, photochemistry, and high explosives in his early career before turning to the ocean sciences. He was Chief of the Explosives Research Department and later Associate Director of Research at the U.S. Naval Ordnance Laboratory prior to joining Woods Hole in 1958.

He has served the Navy as a member of the POLARIS Steering Task Group, the POLARIS Ad Hoc Group for Long Range Research and Development, and as a current member of the Undersea Warfare Research and Development Planning Council.

Dr. Fye is an advisor to universities, companies, and government organizations, and serves on many of the panels that deal with current problems in ocean sciences and marine policy, including the State Department's *Law of the Sea* and *Ocean Affairs* advisory committees. He has been a recipient of the U.S. Navy Meritorious Award, Bureau of Ordnance Development Award, the Presidential Certification of Merit, and honorary degrees from five universities.

Computerized Adaptive Ability Measurement

David J. Weiss*
University of Minnesota

Introduction

The paper and pencil multiple-choice ability test developed during World War I has been widely used throughout the military services. During the last half century millions of men and women have been classified, assigned, trained and promoted in many careers and assignments based on scores from group-administered ability tests similar to the Army Alpha Examination used in World War I. Also during that half century, these tests have been constantly improved through a number of advances in the field of psychological measurement.

The group-administered ability test was a compromise which grew out of the necessity to classify large groups of men to mobilize our personnel for the first world war. Prior to that, ability tests were based on Alfred Binet's intelligence testing model, in which the test was individually administered to each testee by a trained psychologist. Under the psychologist's guidance, the testee was led through a series of pre-normed questions until the questions were too difficult for him or her. When the examiner was sure that the testee had reached a set of items that were too difficult for him, testing was terminated.

Binet's approach had three major characteristics: 1) Testing was begun for each individual with a set of questions at his estimated ability level, based on whatever prior information was available to the examiner 2) The difficulties of the test questions were adapted to the individual's ability level; testing was concentrated in the range of difficulty between the set of questions that were too easy for the individual (i.e., he answered them all correctly) and the set of questions that were too difficult (i.e., he answered them all incorrectly). 3) The number of questions administered to each individual was based on how long it took to identify the testee's ability level; testing was continued as long as necessary, resulting in shorter tests for some testees and longer tests for others.

These three characteristics formed the basis of an individualized or adaptive testing procedure. That is, the difficulties of the test items presented to each testee were adapted or tailored to his ability level based on information gained by observing the correctness or incorrectness of his responses to previously administered items.

*Dr. Weiss is Professor of Psychology at the University of Minnesota. His fields of research are psychometric methods and vocational psychological research.

As World War I changed the face of the globe and our social structures, it also changed the course of psychological measurement. The urgent need to classify millions of men and the lack of trained psychologists to administer individual tests resulted in the rejection of Binet's adaptive model and the development of the group-administered multiple-choice paper-and-pencil ability test. While Binet's approach has survived among clinical psychologists, the paper-and-pencil multiple-choice test is currently used more than 95% of the time for measuring abilities.

The group-administered multiple-choice paper-and-pencil test has grown in usage because it is, like Henry Ford's production line, efficient. Just as an auto maker can turn out thousands of new automobiles in a day, the conventional multiple-choice test can be administered to hundreds or thousands of people at one testing session. The efficiency of this type of test derives from the fact that its administration is highly standardized. All individuals start the test at the same item, proceed through the items in the same order, and end the test either at the same item or when the time limit is reached.

However, in achieving this high degree of standardization the conventional test loses the three major advantages of Binet's adaptive testing strategy: prior information is not used to determine a starting point for testing, items are not adapted to the testee's ability level, and all individuals answer the same items. This loss of adaptive properties in the conventional group-administered test leads to a loss of accuracy in test scores. It has been demonstrated in psychometric theory (e.g., Hick, 1951; Lord, 1970) that the most accurate measurement for a given individual is obtained when the difficulties of the test items are at or near the testee's ability level. In a conventional test, item difficulties are usually concentrated around the average ability level of the *group* for which the test is designed. Thus for individuals whose ability levels are near the group average, the conventional test will provide highly accurate test scores. However, for individuals of above-average or below-average ability levels, the test scores will be less accurate; and the further an individual's ability if from the group average, the less accurate his test score will be. This situation is illustrated graphically by the solid line in Figure 1.

In an adaptive test, however, item difficulties are selected specifically to be at or near the ability level of each individual tested. Thus, test scores will be highly accurate for individuals of all ability levels. This situation is shown by the dashed line in Figure 1. The almost horizontal nature of this line indicates a consistently high level of accuracy in adaptive test scores regardless of how high or low an individual's ability level is.

The major implication of a high and constant level of accuracy is that scores derived from adaptive tests are likely to be equally valid for

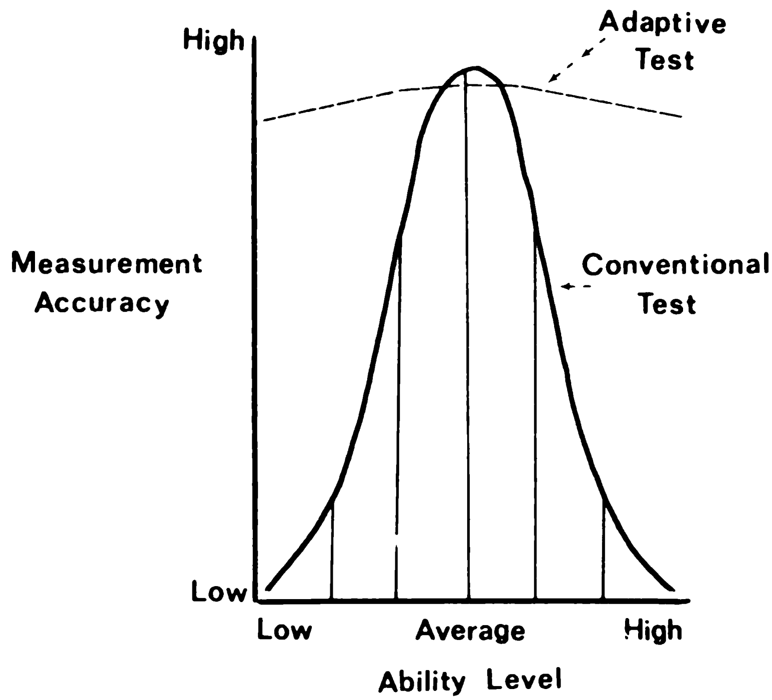


Figure 1 — Illustrative accuracy of conventional and adaptive tests as a function of ability level

all individuals, regardless of ability level. That is, more accurate or reliable scores more validly reflect the ability the test was intended to measure because there is less error in the test scores. And the more validly the test score reflects the true level of ability of an individual the more useful and accurate it will be for making other important predictions about him. In contrast, because scores obtained from conventional tests vary in accuracy, they will also differ in their validity and utility for predictive purposes.

Conventional tests, because their construction and administration is appropriate only to examinees of average ability, also tend to have adverse psychological effects on examinees. A low ability examinee, for whom the test items are much too difficult, may become frustrated and anxious and his test performance may deteriorate. A high ability examinee, for whom the items are much too easy, may become bored or be insufficiently motivated to perform to his fullest capacity. Both frustration and boredom may result in inappropriate responses to the test items, and the examinee's test score will not be an accurate representation of his ability level. In the Binet testing procedure, however, the process of adapting item difficulties to the testee's ability level and terminating the test when it becomes too difficult can reduce the occurrence of frustration and boredom and thus the extent to which the accuracy of test scores is affected by these adverse psychological reactions.

Thus, for both psychometric and psychological reasons, Binet's adaptive testing procedure yields test scores which are highly and uniformly accurate and, therefore, uniformly useful for predictive purposes. However, Binet's approach requires a trained psychologist to administer test items, evaluate their correctness, and choose the next test item to be administered. Modern computer technology now permits these same functions to be performed by interactive computer systems. The computerized adaptive test retains all the advantages of Binet's approach and adds to it a degree of efficiency even beyond that achieved by the paper-and-pencil multiple-choice test. A history of testing written twenty years from now may well show the same kind of re-direction of psychological measurement resulting from the application of computers to the testing process as resulted from the introduction of the paper-and-pencil test during World War I.

Computerized Testing

The research reported below was initiated in order to evaluate the relative merits of a variety of approaches to computerized ability testing. It has been supported since mid-1972 by a contract with the Personnel and Training Research Programs, Office of Naval Research.

The Testing Environment

The availability of on-line computer systems results in a new environment for ability testing. Rather than using a test booklet and an answer sheet, the testee is now tested at a cathode ray terminal (CRT). The CRT is either directly connected to a central computer or is connected by telephone lines, depending on the nature of the computer system and its physical proximity.

Test items are presented on the TV-like screen of the CRT. Following presentation of each item, the testee responds by typing an answer on the CRT keyboard. To insure that lack of familiarity with the CRT terminal does not interfere with the testing process, a series of instructions was developed to teach each testee how to use the CRT and its typewriter keyboard. This instructional sequence is administered by computer prior to the administration of tests to each individual. It introduces each special function key on the CRT, instructs the testee in how to record answers, and then gives him practice in using the CRT keyboard. If, after three tries at any particular instruction the testee still cannot execute it properly, the CRT calls a proctor who helps the student. Once the testee successfully completes this instructional sequence, some personal data is obtained, he is introduced to the kind of test he will take, and he is given several sample questions. Finally, when all the instructions are completed, actual testing is begun.

Our experience with this instructional sequence shows that most testees can learn to use the CRT equipment within five minutes. Several thousand college students and several hundred high school students have taken tests under this system, and the proctor has been required to explain use of the CRT's for only about 2% of the testees. These observations suggest that the man-machine interface problems in computerized testing will likely be minimal although certainly deserving of systematic research.

Our research experience has indicated that characteristics of the computer system itself may become an important component of the testing environment, particularly as it has an impact on the testee's psychological state during testing.

We originally began our on-line testing research using CRT's acoustically connected to a large time-shared computer system. This system was operated by the University of Minnesota to simultaneously serve a variety of time-shared users throughout the state. Our experience in using this system for over two years suggests that a large-scale multipurpose time-sharing system is not an ideal vehicle for computerized adaptive testing. One major problem encountered was that because access to the computer was by telephone lines the display speed of the CRT's was limited to 30 characters per second. This meant that our instructional screens, some of which contained about 2,000 characters of information, required up to a minute each to display. Because many students could read considerably faster than 30 characters per second, they appeared to be irritated at the slow speed of the display.

A second problem, which appears to be characteristic of large-scale time-shared systems, is that they frequently take a long time to respond to the testee (system response time). Modal system response time was usually between 5 and 10 seconds, but response times of 30 to 40 seconds were not unusual. During this period the testee had nothing to do but sit and watch the screen, waiting for the next question to appear. Frustration and negative feelings toward computerized testing was inevitable.

A third problem, which frustrated both the research staff and the testees, was frequent computer system "crashes" during which computer processing would stop for periods ranging from minutes to hours. These crashes were computer failures which were not the result of our computerized testing.

In an attempt to insure a standardized testing environment for future research in computerized testing we began early in the research to investigate alternative computer systems which would be useful for the research and which would ultimately serve as a prototype for operational adaptive testing systems. We were particularly concerned with developing a system which would minimize the extraneous psychological effects on the testee so that we could study the relatively pure effects of adaptive testing itself on such variables as anxiety, frustration and test-taking motivation.

The burgeoning field of minicomputers provided us with an answer to our problem. After carefully investigating a number of leading minicomputers, we took delivery of a Hewlett-Packard 9600E Real-Time system in the summer of 1974. This system provides us with CRT's that display at 960 characters per second; a screen is now fully displayed in two seconds. System response time is less than 1/2 second, and the testee no longer has to wait for the next question to appear. Now the testee, rather than the computer, controls the pace of testing. And computer system "crashes" have virtually ceased. A final advantage is that we can now accurately measure a testee's response latency—the amount of time he spends deciding on an answer to each question—as additional information for potential use in measuring ability or in making predictions.

Good testing practices dictate that tests be administered in a carefully standardized environment. Our experience has shown that computerized testing research and development will be facilitated by a testing environment in which the nature of the man-machine interface and the effects of characteristics of the computer system on each testee's scores are carefully controlled.

Adaptive Testing Strategies

The bulk of our research has been concerned with evaluating the utility and measurement effectiveness of a variety of strategies which have been proposed for adaptive testing. Many of these strategies have been suggested by other investigators, and several other strategies have been developed by our research staff. Each strategy represents a different approach to adapting the difficulty level of the items administered to the ability level of the testee. A testing strategy is defined by a set of rules specifying the procedures by which responses to previously-administered items are used to select the next item or items to be administered. Some of the strategies are relatively simple, mechanical approaches to the problem, while others are based on sophisticated mathematical and statistical convergence models borrowed from other fields of research. A comprehensive review of these strategies can be found in Weiss (1974).

One computer-administered adaptive testing strategy developed by our staff is an adaptation and extension of Binet's original testing strategy (Weiss, 1973). This strategy requires a pool of test items which is divided, or stratified, by difficulty levels. Each level or stratum of the item pool can be thought of as a conventional test in which items are of approximately the same difficulty level. For example, stratum 1 might consist of 20 very easy items with difficulties between p=.89 and p=.99 (i.e., 89% to 99% of the norming group answered these items correctly) and concentrated around p = .94. Stratum 9, at the other extreme of difficulty, might consist of 20 very difficult items with difficulties concentrated around p = .06 and ranging from p = .01 to p = .11. Between these two strata would be seven other strata, each consisting of about 20 items concentrated around difficulty levels ranging from .83 to .17, in steps of .11.

Given an item pool structured in this way, the procedure for moving an individual through the strata is adaptive—hence the name stratified-adaptive, or stradaptive test. Similar to Binet's strategy for individual adaptive testing, the stratum at which a testee begins testing is determined using prior information about the testee. If a testee's ability is expected to be relatively low—based, for example, on the individual's own estimate of his ability—testing might begin at stratum 2 or stratum 3, which consist of easy items. If the testee's ability is expected to be high, testing might begin at stratum 7 or 8 (the more difficult items). When there is no information on which to base an estimate of ability prior to testing, testing can begin with items of average difficulty.

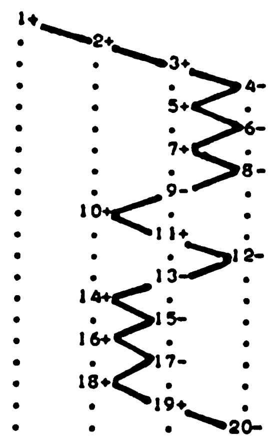
Figure 2 shows the record of an actual stradaptive test administration. The first item administered was the first item available at stratum 5, an item of average difficulty. That item was answered correctly (+). Consequently, the next item administered was a more difficult one—the first item at stratum 6. That item was also answered correctly, and the testee was branched to a more difficult item—the first item at stratum 7, which was also answered correctly. The fourth item administered to this testee was at stratum 8. Thus, in four items, the testee moved from an item of average difficulty (stratum 5) to a difficult item (stratum 8). Because the fourth item administered was too difficult for him, he answered it incorrectly (-). As a result, he was moved back down to stratum 7 for his fifth item, where he received the second item available at that stratum. From this point on, the testee generally alternated between items answered correctly and items answered incorrectly. Similar to Binet's test, the stradaptive test is designed to administer only items that are relevant to a testee's ability level.

While testing proceeds, the computer keeps track of the proportion of items that the testee answered correctly at each stratum. This information is used to determine when testing should be terminated. One termination rule is to cease testing when an individual's "ceiling stratum" has been identified. The ceiling stratum is the least difficult level at which an individual answers all items incorrectly, or answers at a level no higher than chance (when guessing is possible). Given a minimum of five items administered at a stratum and using a 5-alternative multiple-choice item testing can be terminated when the testee answeres 20% or fewer of

REPORT ØN STRADAPTIVE TEST

ID NUMBER: DATE TESTED: 73/07/12

STRATUM: (EASY) 1 2 3 4 5 6 7 8 9 (DIFFICULT)



PRØP.CØRR: 1.00 1.00 .56 0.00

TØTAL PRØPØRTIØN CØRRECT= .550

SCØRES ØN STRADAPTIVE TEST

Ability Level Scores

1. DIFFICULTY ØF MØST DIFFICULT ITEM CØRRECT= 1.49
2. DIFFICULTY ØF THE N+1 TH ITEM= 1.44
3. DIFFICULTY ØF HIGHEST NØN-CHANCE ITEM CØRRECT= 1.49
4. DIFFICULTY ØF HIGHEST STRATUM WITH A CØRRECT ANSWER= 1.33
5. DIFFICULTY ØF THE N+1 TH STRATUM= 1.33
6. DIFFICULTY ØF HIGHEST NØN-CHANCE STRATUM= 1.33
7. INTERPØLATED STRATUM DIFFICULTY= 1.37
8. MEAN DIFFICULTY ØF ALL CØRRECT ITEMS= .88
9. MEAN DIFFICULTY ØF CØRRECT ITEMS BETWEEN CEILING AND BASAL STRATA= 1.28
10. MEAN DIFFICULTY ØF ITEMS CØRRECT AT HIGHEST NØN-CHANCE STRATUM= 1.28

Consistency Scores

11. SD ØF ITEM DIFFICULTIES ENCØUNTERED= .59
12. SD ØF DIFFICULTIES ØF ITEMS ANSWERED CØRRECTLY= .46
13. SD ØF DIFFICULTIES ØF ITEMS ANSWERED CØRRECTLY BETWEEN CEILING AND BASAL STRATA= .18
14. DIFFERENCE IN DIFFICULTIES BETWEEN CEILING AND BASAL STRATA= 1.36
15. NUMBER ØF STRATA BETWEEN CEILING AND BASAL STRATA= 1

Figure 2 — Report of Stradaptive Test for a Consistent Testee

those items incorrectly. For the testee shown in Figure 2, no items in stratum 8 had been answered correctly after five had been administered, and testing was terminated. For that individual, stratum 8 was the ceiling stratum (no items correct) and stratum 6 was the basal stratum (all items correct). Stratum 7 provided almost optimal measurement of the testee's ability level, since he answered 56% of those items correctly. For this testee, ability level was determined using only 20 items.

Figure 3 shows the stradaptive test response record of a testee who required 41 items before his ability level was identified. This testee's response record began at stratum 8, based on a prior estimate of his ability level. It shows several wide oscillations between strata 4 and 9 before the ceiling stratum is finally determined at stratum 8 (only 20% of those items were answered correctly). In the strata between the ceiling stratum and stratum 4, which was the basal stratum, this testee answered between 54% and 67% of the items correctly. Of the total number of items administered to him he answered 49% correctly.

Figures 2 and 3 show a number of different scores for the stradaptive tests. In adaptive testing different items are administered to different individuals and, ideally, everyone would answer about 50% of the items correctly. Consequently, simple number correct or proportion correct scores are not appropriate, and new methods of scoring tests to estimate ability level have been developed and are under investigation in our research.

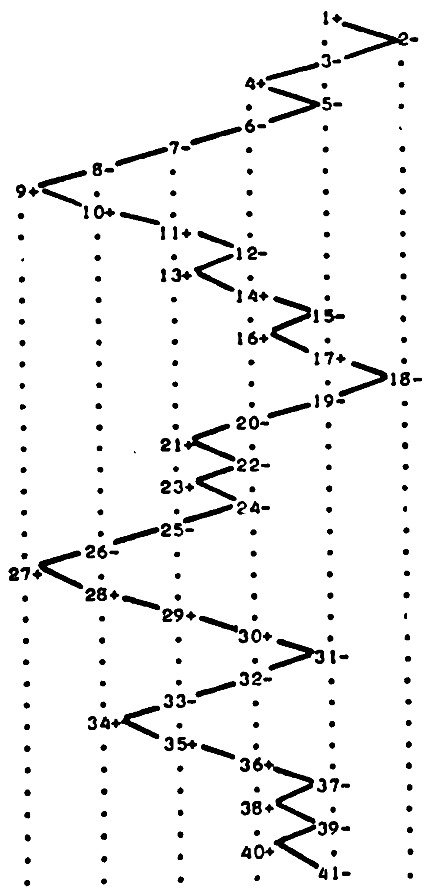
The stradaptive test, in addition to providing ability level scores, also provides what we have called "consistency scores." These scores reflect the range of difficulty of the items administered, and indicate how consistently a given individual interacts with a given pool of items. A comparison of Figures 2 and 3 shows that the test record in Figure 2 concentrates measurement in only a few strata, while the record in Figure 3 reflects an individual who is responding more inconsistently. We have hypothesized that these consistency indices should be related to the reliability of scores *for a given individual*, a concept not measurable with paper-and-pencil tests.

In addition to studying the stradaptive testing model, we are evaluating the relative merits of other adaptive testing strategies. Some of these models, such as stochastic process models, Bayesian estimation models, and maximum likelihood models, represent very sophisticated applications of modern mathematics and probability theory. Others, such as the two-stage, pyramidal, and flexilevel models are based more on the logic of the measurement procedures involved. The results of our comparative evaluation will permit us to limit our future research efforts to those strategies which hold the most promise for providing highly accurate measurement throughout the range of human abilities.

REPØRT ØN STRADAPTIVE TEST

ID NUMBER: DATE TESTED: 73/07/02

STRATUM: (EASY) 1 2 3 4 5 6 7 8 9 (DIFFICULT)



PRØP.CØRR: 1.00 .60 .67 .54 .20 0.00

TOTAL PROPORTION CORRECT= [illegible]

SCØRES ØN STRADAPTIVE TEST

Ability Level Scores

1. DIFFICULTY ØF MØST DIFFICULT ITEM CØRRECT= 1.89
2. DIFFICULTY ØF THE N+1 TH ITEM= 1.01
3. DIFFICULTY ØF HIGHEST NØN-CHANCE ITEM CØRRECT= 1.53
4. DIFFICULTY ØF HIGHEST STRATUM WITH A CØRRECT ANSWER= 2.01
5. DIFFICULTY ØF THE N+1 TH STRATUM= 1.33
6. DIFFICULTY ØF HIGHEST NØN-CHANCE STRATUM= 1.33
7. INTERPØLATED STRATUM DIFFICULTY= 1.36
8. MEAN DIFFICULTY ØF ALL CØRRECT ITEMS= .72
9. MEAN DIFFICULTY ØF CØRRECT ITEMS BETWEEN CEILING AND BASAL STRATA= .76
10. MEAN DIFFICULTY ØF ITEMS CØRRECT AT HIGHEST NØN-CHANCE STRATUM= 1.24

Consistency Scores

11. SD ØF ITEM DIFFICULTIES ENCØUNTERED= .86
12. SD ØF DIFFICULTIES ØF ITEMS ANSWERED CØRRECTLY= .74
13. SD ØF DIFFICULTIES ØF ITEMS ANSWERED CØRRECTLY BETWEEN CEILING AND BASAL STRATA= .50
14. DIFFERENCE IN DIFFICULTIES BETWEEN CEILING AND BASAL STRATA= 2.64
15. NUMBER ØF STRATA BETWEEN CEILING AND BASAL STRATA= 3

Research Approach

There are many questions to be answered before computerized adaptive testing can be most profitably applied in personnel selection, training, classification and promotion. Research is further complicated by the newness of the field. Since virtually no live-testing research had been done in computerized testing prior to 1973, we had to develop our own approach to evaluating the effectiveness of various testing strategies.

In addition to developing a research approach which would permit us to compare the relative effectiveness of about ten basic strategies of adaptive testing, we have had to take into account a very large number of within-strategy variations. Consequently, it was not feasible to systematically vary all these characteristics in live-testing studies (i.e., studies in which testees actually complete a computerized test).

To resolve this dilemma our research program uses a systematic combination of live-testing and computer simulation. First, we construct a particular adaptive test. The test is then administered on the computer to a group of testees in conjunction with either a conventional (non-adaptive) test and/or another type of adaptive test. These tests are usually administered again several weeks later to obtain estimates of the stability of test scores from the various strategies.

Given the data from the live-testing studies, and in conjunction with certain assumptions from modern test theory (e.g., Lord & Novick, 1968), we then construct a computer simulation model, using the same tests which were administered to the live subjects, which gives us results similar to those obtained from the live testing. This computer simulation model can then be used to rapidly assess the effects of varying a number of parameters associated with each testing strategy. Using the computer in this way, we "administer" a precisely specified hypothetical test to a "subject" in a second or so, score the test, and repeat the process for many thousands of "testees." We then change the internal parameters of the testing strategy and "administer" the modified test to another large sample of simulated testees.

Using this approach, we cannot only systematically vary the characteristics of the test, but we can also vary the characteristics of the "testees" themselves. In this way we can determine how a particular adaptive (or conventional) test might work with testees of high ability, low ability, or those whose ability is inappropriate for the test in question. Also, because simulation permits us to use very large groups of "testees" with known simulated ability, we can evaluate the results of the testing using additional criteria which are not available from live-testing studies.

After the simulation studies identify a subset of parametric variations within a strategy which appear to be optimal, it is then necessary to verify those results through live-testing. This is because there might be an interaction between characteristics of the branching strategy and the simulation model which require modifications in the model before the results are fully representative of how live testees behave. This process involves administering to live subjects the particular adaptive test which gave the best results in the simulation studies. The results of the live-testing and simulations are then compared. This empirical verification of the simulation model also serves as a check on the adequacy of the initial simulation model, which might have been based to some extent on the characteristics of a unique set of testees, or on their interaction with some particular adaptive test.

Some Recent Findings

Our research to date has been primarily concerned with evaluating the relative effectiveness and utility of the various adaptive testing strategies, and comparing ability estimates derived from adaptive tests with those from conventional tests. Live-testing studies have shown that scores derived from adaptive tests are more reliable over time than are scores from conventional tests. This higher reliability implies greater accuracy of measurement and greater utility in making longitudinal predictions. The simulation studies which supplement these live-testing studies show that the higher accuracy of measurement for the adaptive tests holds throughout the ability range for all strategies of adaptive testing studied, although some of the strategies are better than others in terms of maintaining a constant degree of accuracy throughout the ability range. Our studies have also shown that adaptive test scores reflect the distribution of ability in a group better than do conventional tests.

One particularly interesting set of data resulted from a live-testing, test-retest study using the stradaptive test strategy. As was indicated earlier, and illustrated in Figures 2 and 3, stradaptive tests yield "consistency" scores as well as ability level scores. These scores reflect the range of difficulty of the items encountered by an individual. One operationalization of the notion of consistency of response is the number of strata used by an individual before the termination crierion is met. Highly consistent individuals use few strata (e.g., Figure 2) and inconsistent testees encounter items at strata of widely differing difficulties (Figure 3).

We hypothesized that consistency should be related to reliability. Thus, the more consistent individuals should be more predictable over time, and the less consistent individuals less predictable. To study this hypothesis, we administered a stradaptive test twice to a group of individuals; an interval of about six weeks separated the two administrations. Using their first test results, we determined each testee's consistency score and sub-divided the total group into five subgroups ranging from very high consistency to very low consistency (N = 17 to N = 41 per group). Within each group we calculated the test-retest correlation (stability coefficient) for each of the ten methods of estimating ability with the stradaptive test.

The general pattern of results was in the predicted direction although it varied for different methods of estimating ability and consistency. The clearest pattern emerged for Ability Score 1 using consistency score 11 (see Figures 2 and 3). For the total group of subjects (N = 170) the test-retest reliability was .65. As consistency increased so did reliability. For the most consistent group, reliability was .94, while the intermediate consistency groups had reliabilities of .85, .84 and.77. For other methods of scoring ability on the stradaptive test, reliabilities for the most consistent group were as high as .98, indicating that for this subgroup their scores on the second testing were almost perfectly predictable from their scores on the first testing. On the other hand, individuals with very inconsistent responses at initial testing generally were least predictable on retest.

These results were, of course, obtained over a rather short time interval. If they can be shown to hold over longer time intervals, they have important practical implications, since they will permit us to identify individuals whose scores on a given test are more (or less) reliable. We will then be able to make selection, placement and classification decisions for specific individuals taking into account the reliability of *each person's* test scores.

Psychological Effects

In addition to investigating psychometric variables in adaptive testing, we are also systematically studying the effects of adaptive testing on certain psychological variables. Accuracy in measuring an individual's ability level can be reduced by the individual's behavior in the testing situation. Guessing, poor motivation, frustration, anxiety, boredom and other psychological variables can add error to an individual's test scores which will reduce the utility of those scores for practical decisions.

Because an adaptive test maintains the difficulty of the test items at or near the testee's ability level, we have hypothesized that it should reduce the adverse influence of these psychological variables and thus increase the accuracy of test scores. These benefits should be in addition to the purely psychometric increases in accuracy due to the adaptive testing procedure. Since different adaptive testing procedures are differentially successful in maintaining test items at the appropriate difficulty level, we are also studying whether these different strategies result in differential reductions in these extraneous variables.

This subset of studies includes a variety of designs which are currently being pursued. In one study, we are concerned with whether testees can accurately perceive the difficulties of test items presented to them. If they cannot distinguish easy from difficult items, then it is unlikely that items of different difficulty will produce these different psychological effects. A second study is concerned with directly comparing the effects of conventional and adaptive testing on test anxiety and test-taking motivation. This study is also concerned with the effects of "feed-back" on test scores and on the same psychological variables; that is, a subgroup of the testees receive information after each item as to whether or not it was answered correctly. In a third study, we are studying the effects of adaptive testing on guessing behavior. In this study, testees answer a multiple-choice test question by assigning probabilities to each of the five alternatives, within the constraint that the five probabilities assigned sum to 1.00. By an analysis of the relationship of the testee's probabilities to the known difficulties of the items, in conjunction with the testee's ability level, we can evaluate the likelihood that an individual was guessing on a test item. We can then compare the relative frequencies of guessing on conventional tests and on different adaptive strategies to determine whether adaptive testing does indeed reduce guessing behavior. Within this study we are also evaluating the more general utility of probabilistic responding as an important alternative to the conventional binary response mode. Combined with adaptive testing procedures, probabilistic responding may result in a very powerful approach to ability measurement.

Our concern with the psychological effects of adaptive testing is motivated partly by the problems raised in the use of ability tests with minority persons. We are concerned with increasing the accuracy of measurement for minority individuals and thereby increasing the fairness of testing for these groups. Consequently, we are striving to identify those psychological variables which are more likely to operate to reduce the accuracy of test scores in both minority and non-minority groups, and to eliminate them from the testing situation insofar as is possible.

Future Developments

Research in computerized testing to date has had three defining characteristics. First, it has assumed the existence of an unidimensional trait. That is, it is assumed that a given test item pool measures only one ability. A second characteristic of this research is that it is based entirely on between-item branching strategies. In other words, a testee's response to a given test item (or sequence of test items) is used to determine which item is to be administered next. Finally, the majority of adaptive testing research has used verbal test items, primarily because of the prohibitive expense of computerized display of pictorial materials.

Our future research in adaptive testing will concentrate on eliminating these three restrictive aspects of most current research. We will be concerned with developing branching models which do not require the restrictive assumption of unidimensionality. The unidimensionality assumption is unrealistic for several reasons. First, it is extremely difficult to obtain test item pools within any given ability domain (e.g., vocabulary or word knowledge) which are strictly unidimensional. This is due to the fact that dimensionaltiy is partly a function of the difficulty of the test items; very difficult test items will tend to define different factors of ability than will moderately difficult or easy items. Secondly, even though a test item pool can be shown to be reasonably unidimensional (e.g., McBride & Weiss, 1974), such apparent unidimensionality is defined on the basis of the *average* characteristics of a group, the group of testees on whom dimensionality is studied. However, unidimensionality of a set of items for a *group* is no guarantee that the set of items will be unidimensional for an *individual*. Elsewhere (Weiss, 1973, pp. 39-42) I have shown some preliminary results which imply that different individuals are differentially scalable on an apparently unidimensional set of items. Our future research in this area will be concerned with systematically replicating and extending these results. If the phenomenon is verified, we will develop models which take intra-individual dimensionality into account in the interactive adaptive testing process.

Personnel selection and placement is rarely successful when it is based on predictions made from measuring only one variable. Consequently, multiple ability batteries have been developed which are used to assess the capabilities of individuals on a variety of occupationally-relevant abilities. Even when these abilities are unidimensional, they are often correlated with each other. Using a strictly unidimensional branching model, adaptive testing would proceed by locating an individual's ability level on one variable, then proceed to another, and then to a third ability, and so on. But when abilities are correlated, this procedure wastes testing time since knowledge of status on one ability yields information relevant to an individual's status on other ability dimensions.

To minimize testing time there are two ways to proceed. First, we can begin by assessing an individual's position on one ability. Then, using our knowledge of the correlation of that ability with a second, we can generate an initial ability estimate for testing on the second ability. To estimate a score on the third ability, we could use the multiple correlation of the third ability with the first and second to reduce the range of initial ability estimates for the third ability, and so on through all abilities. A second approach would be to conceptualize the correlated set of abilities as a multidimensional space and develop adaptive testing procedures to systematically branch an individual through this space. Thus, in a minimum set of items we would locate the testee's position on all abilities simultaneously. This latter model would be advantageous in that it could also be designed to simultaneously handle the problem of intra-individual dimensionality.

Our future plans also include the development of computer-administered tests to measure abilities not now measurable with conventional paper-and-pencil tests. These would include tests of reasoning ability, based on intra-item branching techniques. In these tests, the typical test "item" might be a problem situation, and the individual would be given a choice of a variety of sorts of information which he can apply to the solution of the problem. His choice of a given type of information will lead him by a successive series of further choices down a particular path of reasoning used in the solution of the problem. Some paths might lead to inappropriate solutions and some to inefficient solutions. Thus this kind of test item could be thought of as a "conceptual maze," where the quality of an individual's reasoning is determined by the nature of his process in reaching the solution, including its efficiency and the amount of time spent in making each decision. Tests of this kind, administerable to large numbers of testees only under computer control, should have important uses in military personnel evaluation.

We also plan to develop tests ot short-term and longer-term memory in auditory, pictorial and verbal modalities. The precise control which the computer provides in the measurement process will enable the measurement of these new abilities. When appropriate computer terminal equip ment is available, we will also explore the utility of pictorial and spoken language tests to replace current printed verbal tests. Spoken language tests might supplement printed verbal tests for minority group members. Non-verbal tests might permit the adaptive measurement of abilities such as chart and map reading and visual scanning and interpretation to supplement currently available ability measures.

A Note on Costs

A discussion of computerized testing would not be complete without considering the economic feasibility of this mode of testing. The economic practicality can be viewed from two complementary vantage points: 1) the differences in actual costs of test administration, scoring and reporting, and 2) the savings resulting from the characteristics of the test scores.

A major characteristic of adaptive testing is that it can considerably reduce the amount of time a testee spends on taking tests. Some adaptive testing strategies can achieve with 30 items the validity and reliability of a 100-item conventional test. Thus, the average adaptive test need be only 10 minutes long compared to the 30 minutes required for a conventional test. Since the testee's time often represents both an individual and institutional investment, this two-thirds reduction in time should represent an important economic gain. If the testee's time is "free," i.e., no monetary savings are determinable from reducing testing time, there is still an economic gain possible since the time saved by the adaptive testing procedure can be used to measure the testee's abilities in other areas. In this way, adaptive testing could make accurate prediction of complex criteria possible without requiring inordinate amounts of testing time on the part of each testee.

Very little data is available on the actual cost of either adaptive testing or current paper-and-pencil procedures. In computing the costs of paper-and-pencil testing, a number of hidden costs must be kept in mind. These include costs of obtaining publisher's catalogues, ordering supplier, storing supplies, re-ordering, mailing answer sheets for scoring, keeping records of answer sheets submitted for scoring and the like. These costs are, of course, in addition to the costs of purchasing test booklets, answer sheets and scoring services (or hand-scoring keys and wages of scoring personal). To this must also be added costs associated with waiting—waiting for supplies to arrive, for testees to congregate for testing and for the test results to be scored. On the other hand, in computerized testing the test will be available at the push of a button or two. There will be no delays for score reports, no materials to order, no inventories to maintain. The test will be available when it is needed, and results (interpreted in a variety of ways) will be available instantly to assist in decision-making.

Waters (1975) gives data estimating the costs of a computerized stradaptive test at 87¢ per testee, excluding development costs and costs of terminal equipment. Our calculations, again excluding development costs but including terminal equipment, range from $1 to $1.25 per test. Consequently, in one hour of testing time, a testee could complete six tests at a total cost between $6 and $7.25. Given the decreases in the cost of computers that have occurred in the last five years, and projecting them a few more years into the future, the direct costs of computerized testing will, in the near future, be less than those of paper and pencil testing.

However, the real economic advantages of computerized testing will derive from the greater accuracy of test scores. Since adaptive testing results in scores which are both more precise and more accurate, adaptive test scores should result in fewer errors in selection, placement and classification. Since training one man for one assignment may cost many thousands of dollars, each inappropriate assignment based on unreliable or invalid test scores can waste those thousands of dollars. An overestimate of a person's ability can lead to an inappropriate personnel assignment which can result in a waste of many additional thousands of dollars if the individual is not competent in his assignment. Conversely, an individual whose test scores are lower than his actual ability is underutilized. In this case the potential productivity of an individual whose skills and abilities are underestimated by poor testing procedures is lost, again costing money to the employing organization. Finally, the underutilized individual might be a retention or morale problem In that case, the costs involved in training the individual are not fully recovered since his tenure (or performance) on the job will likely not be optimal.

Computerized adaptive testing does appear to be economically feasible. To the extent that it will give us tests which measure more accurately and span a wider range of occupationally-relevant abilities it should yield a substantial economic benefit for the military services. The search for the most practical and efficient way to implement adaptive testing procedures is just beginning. Our research results should provide the necessary data from which effective computerized adaptive testing systems can be developed and applied in the near future.

References

W. E. Hick, "Information Theory and Intelligence Tests." *British Journal of Psychology, Statistical Section*, 1951, **4**, 157-164.

F. M. Lord, "Some Test Theory for Tailored Testing." In W. H. Holtzman (ED.), *Computer-assisted instruction, testing, and guidance*, New York: Harper and Row, 1970.

F. M. Lord and M. R. Novick, *Statistical Theories of Mental Test Scores*. Reading, Mass.: Addison-Wesley, 1968.

J. R. McBride and D. J. Weiss, "A Word Knowledge Item Pool for Adaptive Ability Measurement." *Research Report 74-1*, Psychometric Methods Program, Department of Psychology, University of Minnesota, 1974. (AD 781894)

B. K. Waters, "An Empirical Investigation of Weiss' Stradaptive Testing Model." Paper presented at the Conference of Computerized Adaptive Testing, Washington, D.C., June 13-14, 1975.

D. J. Weiss, "The Stratified Adaptive Computerized Ability Test." *Research Report 73-3*, Psychometric Methods Program, Department of Psychology, University of Minnesota, 1973. (AD 768376)

D. J. Weiss, "Strategies of Adaptive Ability Measurement." *Research Report 74-5*, Psychometric Methods Program, Department of Psychology, University of Minnesota, 1974. (AD A004270).

Establishing Safe Limits for Divers*

The Tests

A four-man team recently completed the last of two simulated deep dives at the University of Pennsylvania to help scientists establish safe limitations for crews exploring for undersea oil, maintaining oil wells, or accomplishing salvage of sunken vessels. The first dive lasted 15 days and the second 21 days.

The quartet, all professional offshore oilfield divers who volunteered for the biomedical experiments, were subjected to increasing pressures until they reached the equivalent of 1600 feet of sea water in the hyperbaric (high-pressure) chamber complex of the University's Institute for Environmental Medicine.

The entire program entailed progressive study of effects induced by rapid exposure to 400, 800, 1200 and 1600 feet of sea water. Each step included a series of "excursions" from one depth to another. The program was climaxed by excursions from 1200 to 1600 feet to perform the actual practical work of maintenance on a well-head from an oil rig in the Institute's water-filled high-pressure chamber (Figure 1).

The team, which is backed up by two other volunteer divers, was monitored to determine the neurological effects of changes in hydrostatic pressure itself, as well as the interrelated influences of pressure and the rate of change of pressure. Their sensory, mental neural, neuromuscular and cardio-pulmonary responses were continuously measured throughout sequences of compression and decompression. The biomedical studies produced voluminous data which will require months of analysis and evaluation.

The research project is the fourth in a six-year series conducted by the Institute to investigate the basic limits of human tolerance and performance in undersea activity. It is under the overall super-vision of Dr. Christian J. Lambertsen, director of the Institute and an internationally noted authority in environmental sciences and operations.

Previous studies by the Institute concentrated on respiratory-pulmonary-exercise tolerance to 1200 feet of sea water with the volunteer teams breathing dense gases such as neon to simulate the equivalent of helium density at depths of 5000 feet. The results, according to Dr. Lambertsen, "indicated breathing during physical work to be practical to at least these depths," which would make possible exploration of the slopes beyond the continental shelves.

*Compiled from press releases written by Hal Helfrich of The University of Pennsylvania News Bureau

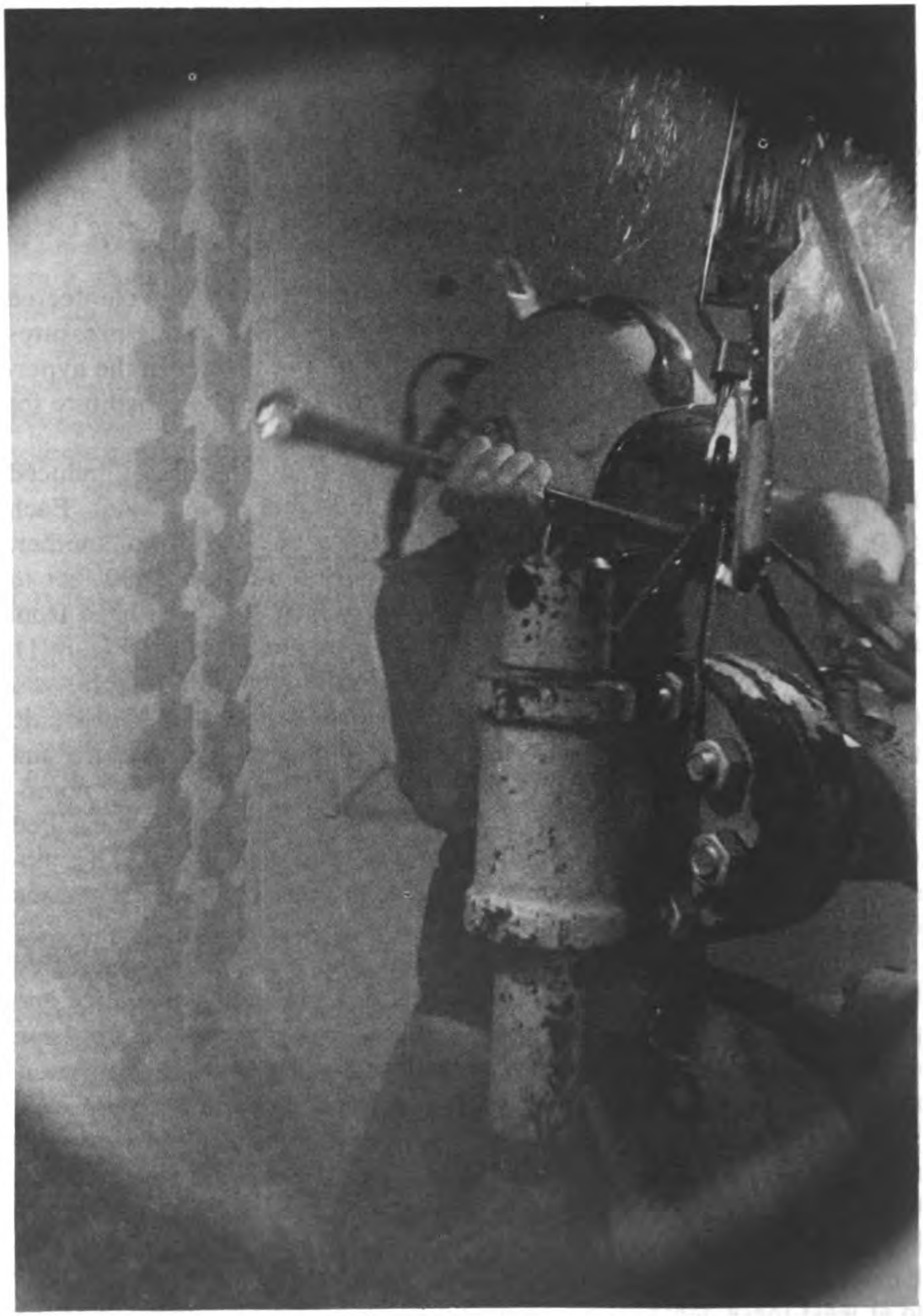


Figure 1 — Diver disassembling an actual oil well-head in water-filled "ocean simulator" chamber. This was the first practical working exposure to this depth (1600 fsw).

In a series of 1971 hyperbaric chamber tests, it took about 10 days to get to 1200 feet, doing a full range of studies each day along the way. During these tests a team of four University of Pennsylvania student volunteers were given three different inert gases - helium, nitrogen and neon-mixed with oxygen, in order to make direct comparisons of these three gases. These 1971 tests were the first studies of neon as a breathing gas at such great depths. The tests purposely covered the range of functions from mental capability to physical work capacity. The divers came through the stresses of pressure so well that they could carry out both mental and physical work essentially as well as 1200 feet deep as they could at the surface.

However, since the 1971 tests, helium-oxygen studies at 1600 feet by other laboratories in France and the United States have suggested factors involved in the absolute pressure of exposure or in the rate of compression may induce impaired neurophysiological functions such as tremors, decreased coordination and balance. In animal studies convulsions have resulted from extremely rapid compression. Distress in breathing also was reported in spite of the demonstration by the Institute's scientists that the lungs themselves could function to many thousands of feet of depth.

For these reasons the range of ambient and helium pressure between 1200 and 1600 feet of sea water required detailed investigation of nervous system function. This was to allow direct comparisons of any higher-pressure effects with the Institute's past findings which concluded that a normal work schedule at 1200 feet is practical; that it is theretically possible for a helium oxygen mixture to be breathed as deep as 5000 feet; and that with hydrogen in the breathing mixture even deeper capabilities should be possible (cover picture).

The current program's two major phases were separated by an active three weeks to allow time for data analyses, equipment maintenance, and additional training of the divers. In general, the first experiment sequence was made for the multiple purposes of establishing decompression rate schedules, neurological studies, and specific training for the deeper series. The second was intended to expand the physiological measurement studies to the maximum depth, determine optimum rates of compression to great depth, and then to study the divers' practical performance of specific skilled undersea work related to deep oil rig diving.

In the 1971 deep-pressure study the rate of compression was gradually reached by going to 100, 200 or 300 feet, then back to the surface, then to 700, 900 and 1200 feet to begin the long-multi-day, step-by-step approach.

This time it did not take so many days to reach a certain base depth. In the present tests the desired base depths were reached within an hour and then the divers quickly began excursions to greater pressures from each stable baseline of gas saturation. At each baseline (such as 800 or 1200 feet) exposure to increased pressure is long enough to saturate body tissues completely with the inert breathing gas (helium) used as an oxygen diluting agent in breathing.

The first dive of this year's experiment involved a rapid simulated descent to an 800-foot baseline depth, saturation, and then repeated excursions to 1200 feet of sea water. The second dive followed by saturation at 1200 feet to study repeated excursions to 1600 feet of sea water.

This approach of intermittent excursion as a means for safe attainment of undersea regions has not previously been accomplished to such depths.

The transient exposures, according to Dr. Lambertsen, provide means for determining the respiratory, mental, sensory, neurophysiological and physical competence of normal individuals not fully equilibrated with the high pressure of inert gas respired. By utilizing periodic rather than continuous exposure to pressures greater than 1200 feet of sea water, investigation can be made of the rate and degree of physiological adaptation to repeated compression. This should provide means for speedier recovery from initial functional disruptions which now limit man's ability to descend freely for useful work.

This method can also provide for studying the rate of alleviation of any compression-induced effects as the divers decompress to the stable saturation levels of 1200 feet of sea water and will also allow for determining the practicality of using other respirable inert gases, including neon and nitrogen, as drugs to overcome deleterious effects of compression with helium.

Year-long preparations for the project got under way in July, 1974. The complex of four 3-inch-thick steel chambers (two of them can be divided into separate compartments) were extensively reconditioned: valves and welds were replaced; carbon dioxide scrubbers were installed; and radiological, hydrostatic, air and helium tests were made. The many backup systems were tested and improved. A PDP-11 computer was moved into an adjoining room to augment the PDP-12 computer in the Institute's International Data Bank-containing vital information on divers, experiment procedures, decompression rate schedules, etc.-to provide rapid assistance if necessary.

A 2,400-pound, six-foot-high oil well-head similar to many used on the ocean bottom was placed in the wet chamber, and volunteers were chosen, given a wide range of medical tests and rehearsed for their roles in the experiments. Because the project is keyed to establishing the speed with which men can be gotten down to oil rigs and their ability to do maintenance work in such a high-pressure environment, the Institute required volunteer candidates to be:

"Of exceptional common sense and judgment; experienced oil field divers; interested and comptent in use of life support apparatus and environmental chamber control; technically proficient; physically sound, to age 35; clearly able to work well with others experienced in diving to 400 feet of sea water, preferably with saturation diving exposures; free of medical disorders."

After six volunteers were carefully selected, four of them were chosen to be the initial active test subjects. The other two were assigned standby status and duty as associate chamber operators with the more than 60 scientists and technicians taking part in the round-the-clock operations.

During the first experiment phase, concerned in part with the important problem of deep compression, to 800 and 1200-foot depths, two divers developed "bends" during decompression from 1200 feet, providing considerable new information for excursion operations.

"Bends," caused by an accumulation of bubbles in tissues or blood vessels, may be characterized by pain in the joints, by sensory disturbances, by dizziness or even paralyses. It can affect not only the limbs but spinal cord and inner ear mechanisms concerned with hearing and balance. It is fairly common in diving and tunnel decompression, and is a major target of the Institute's efforts to determine safe limitations for divers under laboratory conditions instead of in the open sea where the hazards would be much greater.

Both subjects with evidence of "bends" were returned immediately to pressures and breathing gases that eradicated their symptoms. They were then decompressed more slowly until they were able to proceed with the scheduled experiments under close medical monitoring to the end of the project's first phase.

The program represents the successful coordination of University, industiral and federal scientific interests. Through it the University established active and open collaboration of the offshore diving industry, the industrial gas industry, the National Aeronatics and Space Administration (NASA), the U.S. Navy, and other university researchers.

The core support is being provided by the Heart and Lung Institute of the National Institutes of Health, while major specific support for the underwater biomedical research is being provided by the Naval Medical Research and Development Command and the Office of Naval Research.

The Chamber System

The large environmental chamber system at the Institute for Environmental Medicine consists of six interrelated compartments: five dry compartments and a water-filled compartment (Figure 2).

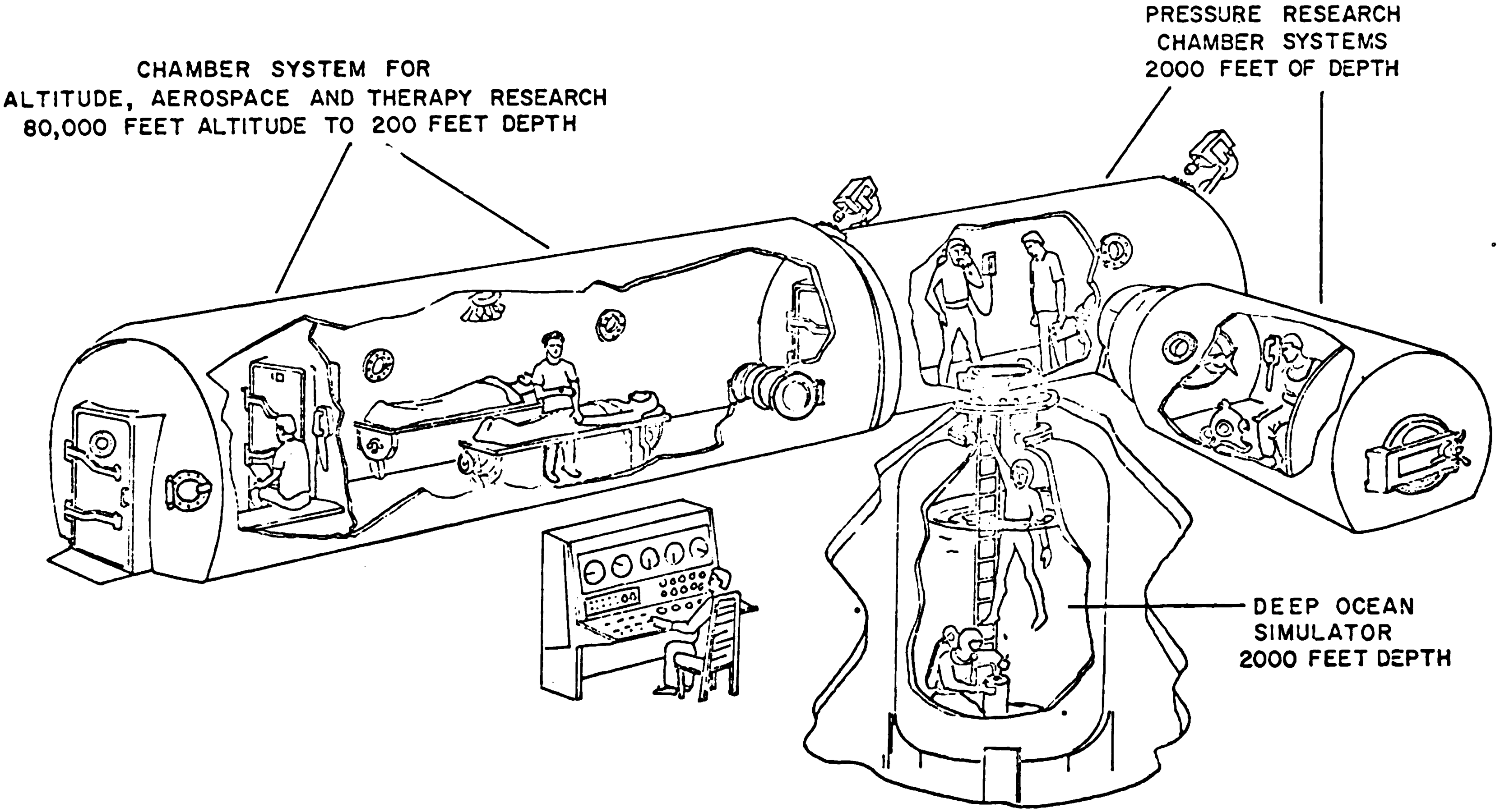


Figure 2 – The large environmental chamber complex at the University of Pennsylvania's Institute for Environmental Medicine is made up of four compartments, two of which can be subdivided. Five of the compartments are utilized for experimentation or clinical therapeutic hyperoxygenation in a dry environment capable of pressure ranges from approximately 80,000 feet altitude to 2000 feet of sea water. The sixth compartment is a "wet" or water-filled system.

The "wet" compartment constitutes a temperature-controlled diving compartment filled with water. It can be pressurized to simulate the deep, wet, cold, buoyant, and dense environment of the open ocean, or the characteristics of a mountain lake. It provides an intermediate step between conditions in the laboratory research compartments and those in natural open waters.

The two largest dry compartments – in a single, subdivided compartment – have a maximum high pressure capability of 200 feet of sea water (about seven times the atmospheric pressure at sea level). They are specifically designed for convenience in therapeutic hyper-oxygenation and for basic and clinical investigations of oxygen effects in normal subjects and patients. Space and utilities are provided for up to six treatment or experiment locations.

The three remaining dry compartments (chamber 3 is also subdivided) provide for acute or chronic continuous long-term exposures to unusual gaseous atmospheres of therapeutic or environmental significance. These compartments cover the full range of chamber system ambient pressures capability, from approximately 80,000 feet altitude to 2000 feet of sea water.

During this project, chambers 2 and 3 were used as living quarters and as laboratories for most of the physiological monitoring of the test subjects as they are exposed quickly to 800 and 1200 feet of sea water for saturation periods and excursions to depths down to 1600 feet of sea water.

The Institute also has a number of small high-pressure chambers intended for animal exposures to unusual gaseous environments. These chambers simulate a range of pressures equivalent to a maximum of approximately 5000 feet of sea water.

Centering around the large chamber systems are the support laboratories and offices that form an integral part of the Institute's environmental research and therapeutic activities. These include units for sea-level biochemical, analytical, computational and radioisotope work, for animal and human experimentation and recovery, for mechanical and electronic shop activities, and for administrative support (Figure 3).

Besides a PDP-11 computer set in a room adjoining the chambers for emergency use, the PDP-12 computer of the International Data Bank is located nearby to serve as a continuing system for evaluation, accumulation, storage and ready retrieval of information pertaining to inert gas exchange, decompression, and the development and treatment of decompression sickness. The Medical School and University Computer Centers can also supplement these two computers for additional information as needed.

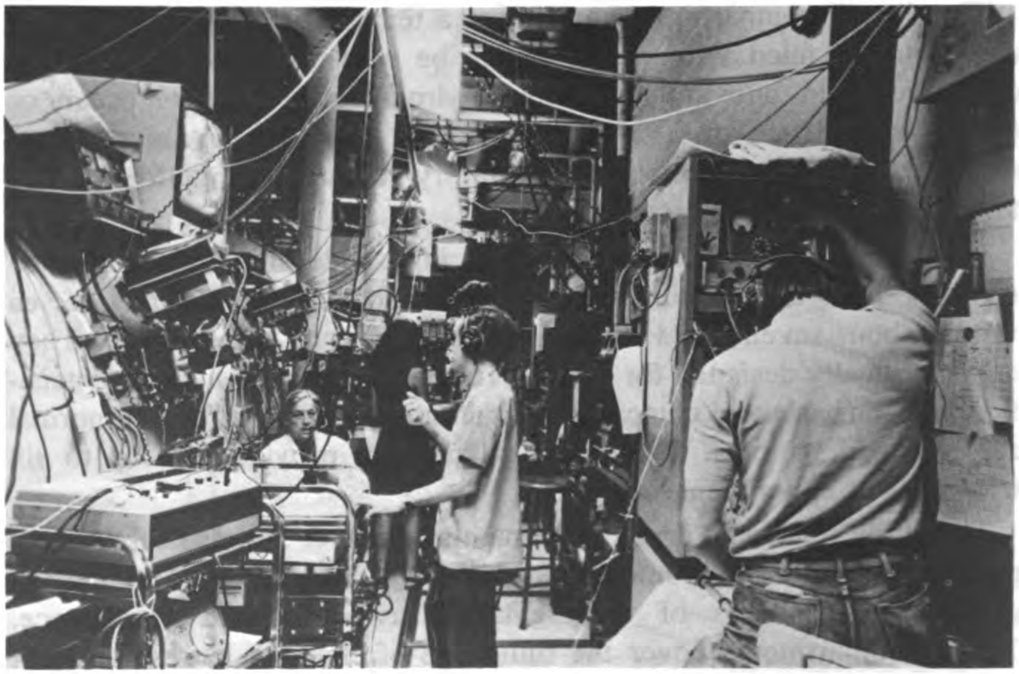


Figure 3 — An instrument station outside the experiment chamber.

Because helium rapidly removes heat from the body, temperature in the chambers is maintained at a level close to body temperature at a value which the test subjects deem comfortable. The helium also so modifies voice characteristics that the subjects would be unintelligible to each other and to outside investigators without the aid of electronic voice reconditioners (helium voice unscramblers). Redundant equipment is provided throughout the system for safety. It includes two backup power generators; two electric-powered communications modes and a sound-powered communication setup; a supply of gases set at specific depths with masks on line for emergency use; modes for sending in and taking out equipment, food and other necessaries; an expulsion system for waste liquids; and both interior and exterior automatic fire-quenching devices.

A basic, 24-hour operational crew outside the chambers always consists of a medical doctor, a primary chamber operator and several assistants responsible for maintenance of gas pressures and composition at the proper levels, continuous sampling of and analysis of gas mixtures, monitoring of carbon dioxide levels, and communications. These control and safety functions are closely integrated with the multiple and detailed measurement activities of the investigator teams performing the scientific observations.

Research Notes

Admiral Geiger is the New Chief of Naval Research

Rear Admiral Robert K. Geiger, USN, has assumed command of the Office of Naval Research (ONR), relieveing Rear Admiral Merton D. Van Orden, USN, who retired after serving as Chief of Naval Research since June 1972. The new Chief of Naval Research comes to ONR from the Naval Electronic Systems Command.

As Chief of Naval Research, Adm. Geiger will head an organization composed of the ONR Headquarters in Arlington, Virginia and several branch offices around the country and overseas. Also under ONR are the Naval Research Laboratory, Washington, D.C.; the Naval Arctic Research Laboratory, Barrow, Alaska; and the Naval Biomedical Research Laboratory, Oakland, California. The total number of people working for ONR is approximately 4000.

A native of St. Joseph, Missouri, Adm. Geiger graduated from the U.S. Naval Academy in June 1947. His first duty following the graduation was aboard the USS BAIROKO (CVE-115). In 1949 he entered flight training at Pensacola, Florida, obtaining his naval aviator wings in 1950.

During the years 1953-55 he attended the Naval Postgraduate School at Monterey, California receiving a B.S. in Ordnance Engineering. He then attended Massachusetts Institute of Technology graduating in 1956 with an M.S. in Aeronautical Engineering.

Since then he has served in several increasingly responsible assignments. These include duty with the Air Force Directorate of Special Projects at El Segundo, California, as Assistant Deputy Director and later as Deputy Director for Advanced Plans for three and a half years. During this period he was awarded the Air Force Commendation Medal and at the end of his assignment he was awarded the Legion of Merit.

In 1969 he reported to the Office of the Secretary of the Air Force as Deputy Director for Programs, Office of Space Systems. Just before leaving his assignment he was awarded a second Legion of Merit for "...exemplary managerial skills and technical competence in the development of programs vital to the Nation."

His last assignment before coming to ONR was that of Project manager, Naval Electronic Systems Command; concurrently he was Director of the Space and Command Support Division of the Office of the Chief of Naval Operations. He became Rear Admiral in July 1974.

In addition to the Legion of Merit citations and Air Force Commendation Medal, Adm. Geiger has the American Campaign Medal, the World War II Victory Medal, the Navy Occupational Service Medal and the National Service Medal with bronze star.

Toward Better Mooring Techniques

A three-legged undersea moored cable installation, consisting of more than 10 miles of cables and covering nearly a square mile of ocean territory, has completed its first year of successful operation; a milestone in the Navy's Ocean Engineering Program.

This unique large-scale, three-dimensional experimental model, SEACON II, is being used to gather data for evaluating cable array design techniques and to improve the state-of-the-art of moor installations and hardware.

Installed in 2,900 feet of water in the Santa Monica Basin off the Southern California coast by the Navy's Civil Engineering Laboratory (CEL), Port Hueneme, Calif., the combination of wire rope and sophisticated instrumentation continues to maintain nearly complete mechanical and electrical integrity.

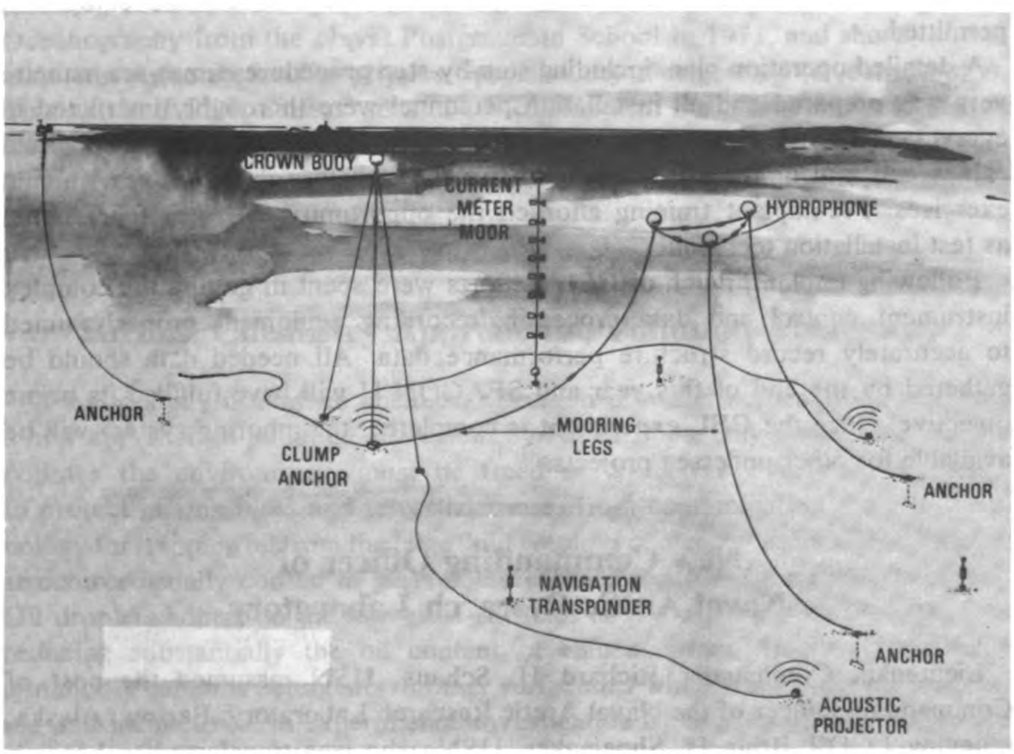
Development of newer and better mooring techniques and hardware is necessary as man goes deeper into the ocean with his exploration and construction. He will come to rely upon power sources, equipment and electronic devices that will be attached to subsurface platforms similar to SEACON II. Such a subsurface moor, nearly independent of the turbulent sea surface, can be an economical means for providing a stable platform in the ocean. Also, another obvious problem, that of vandalism, is pratically eliminated.

SEACON II marks the first time a large three-dimensional moor has been fabricated for the expressed purpose of obtaining in-situ design information. It has been gathering valuable engineering data on structure deflections and cable tensions due to ocean currents.

Since 1960, many analytical models have been developed to analyze the behavior of cable structures. These models attempt to predict the tensions in cables and the geometry of moorings affected by changing ocean currents. Before SEACON II, little experimental data existed to evaluate the accuracy of the models and to determine the most reliable methods for designing undersea cable structures.

The SEACON II structure consists of triangular-shaped cable system tethered 500 feet below the ocean surface by three 4,000-foot long mooring legs. A 5-1/2 foot diameter steel buoy supports each corner of the triangle which is 1,000 feet on a side. The cables which form the structure are "torque balanced", designed to ease the job of deploying cables from a ship at sea. Some cables contain electrical conductors to permit transmission of power and electrical signals between distant points on the structure.

Deep water explosive anchors, developed under a separate experimental program at CEL, secure two of the mooring legs to the seafloor. Experience being gained with the anchors during SEACON II is increasing confidence in their long-term anchoring capability. Engineers also are developing increased skills in anchor installation during a quasi-operational situation. Three anchors were installed successfully during a single work day at sea (a third explosive embedment anchor is being used to secure a single point construction moor at the site). An 8-ton dead weight clump anchor is used to moor the third leg of the structure. The clump anchor also doubles as a container for a radioisotope power generator (RPG) which constantly supplies 10 watts of power to operate instrumentation.



SEACON II

An acoustic positioning system is the main source of data acquisition on the structure's response to changing ocean currents at the construction site. The system consists of acoustic projectors (sound source) accurately positioned on the seafloor and hydrophones (listening stations) at key locations on the structure. Signals are sent between the projectors and hydrophones every half hour to monitor the changing shape of the structure. The system is designed to detect a position move as minor as one foot. Sensors for measuring cable tensions and water depths at key locations complete the structure response measurement system. All sensors are controlled and data recorded on magnetic tape by an electronics package situated in a buoy 60 feet below the surface. The package is accessible to divers who can service it easily. CEL's Diving Locker personnel, aboard their converted LCM-8 work boat, provide support service.

A current measurement system detects the magnitude and direction of ocean currents which move the structure about. A string of 10 current meters near the structure measures the varying current profile. A neutral float interacting with a local seafloor navigation system was developed to calibrate the current meters in place. The float is weighted to become neutrally buoyant at any desired depth. Once at a given depth, it is allowed to float through the site while its position is continuously monitored. The velocity and direction of the neutral float is then determined and compared with data obtained from the moored meters.

A distinguishing feature of the SEACON II operation was the eight-phase installation plan which provided needed flexibility to insure a successful implant. Each phase was designed to be completed in one day or less and equipment be protected from bad weather at the end of each phase. However, phases were able to be combined for a more efficient operation when weather and equipment permitted.

A detailed operation plan, including step-by-step procedures for at-sea maneuvers was prepared and all installation personnel were thoroughly instructed in every aspect. CEL riggers who manned the 135-foot implant barge used actual cables and real or realistic components of the structure during at-sea training exercises. Pre-implant training afforded the opportunity to instruct personnel as test installation techniques.

Following implant, much of CEL's efforts were spent in getting the complex instrument control and data processing/recording equipment properly tuned to accurately record structure performance data. All needed data should be gathered by the end of this year and SEACON II will have fulfilled its prime objective. Once the CEL experiment is completed, the mooring system will be available for other undersea projects.

New Commanding Officer of Naval Arctic Research Laboratory

Lieutenant Commander Richard H. Schaus, USN, assumed the post of Commanding Officer of the Naval Arctic Research Laboratory, Barrow, Alaska, relieving LCDR Brian H. Shoemaker, USN, who was transferred to become Commanding Officer of the Light Helicopter ASW Squadron based at Imperial Beach, California. LCDR Schaus, 35, a native of California, took command recently in ceremonies at Barrow. He was previously the Oceanographic Advisor of the U.S. Naval Mission to Colombia, where he was responsible for the U.S. Oceanographic Advisory Assistance Program performing advisory, liaison, research and graduate level teaching functions with government, naval, and academic communites. During this assignment he developed and published in Spanish a numerical circulation model and an estuarine flushing model which he applied to the study of natural and artifically induced estuarine regimes, with particular emphasis on pollution considerations. He also developed and published analytical reference texts in Spanish on the subjects of Dynamical Oceanography and Introductory Wave Theory and Coastal Processes.

As the Commanding Officer, LCDR Schaus will supervise the activities of the Naval Arctic Research Program including the following: provision of expert guidance to and coordination of field and laboratory research tasks assigned by the Chief of Naval Research; logistic support for all Arctic Research Programs; indoctrination and training of administrative and scientific personnel for living and working in arctic areas; development of techniques, methods and procedures for exploitation of arctic research; liaison between ONR personnel and scientists in the arctic; and advisory functions of Chief of Naval Research on current programs and requirements for future research.

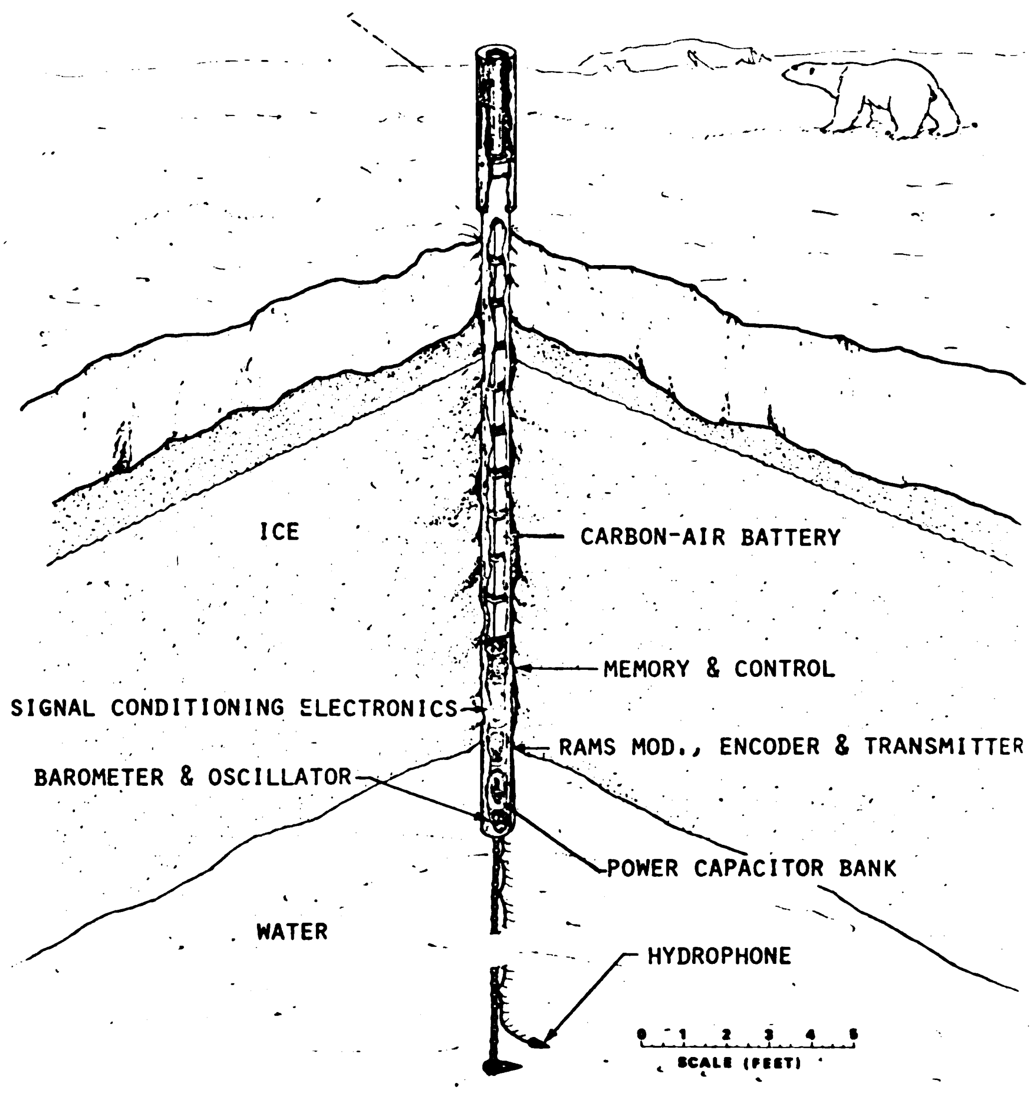
LCDR Schaus, who received his BS degree in Natural Resources from the University of Michigan in 1962, has been closely associated with antisubmarine warfare during the early part of his career, and with Oceanography during the past six years. He has served tours of duty aboard USS HANSON (DD832), USS SHIELDS (DD596), USS DALE (DLG19) and with Commander, Cruiser Destroyer Flotilla SEVEN. He received his MS degree in Oceanography from the Naval Postgraduate School in 1971, and shortly thereafter was selected for Geophysics Special Duty Designation (Oceanography). This will be LCDR Schaus' second association with the Naval Arctic Research Laboratory, having worked there in 1971 while developing a numerical model of the hydro-thermodynamics of an arctic lead during thermal closure, the subject of his masters thesis.

Surface Chemistry Approach to Pollution Abatement

A basic study of surface chemical effects in oil-water separation has opened a new approach to pollution abatement by Naval ships. Oily waste water, which pollutes the environment, must be freed of oil before overboard discharge to protect marine biota and terrestial shores from contamination. Current technology for trapping oil from the bilge fluid employs coalescer-filters whose internal structures usually consist of glass fibers secured together by a resinous binder. Oil droplets adhere to the fiberglass particles as the water passes through, thus reducing substantially the oil content of effluent water. However, the bilge commonly contains detergents or other surfactants which can render the coalescing unit ineffective. An experimental investigation by S. Kaufman of the Surface Chemistry Branch of the Naval Research Laboratory has shown that oil drops in equilibrium with water adhere more tenaciously to the low-energy surfaces of *polymers* than to the high-energy surfaces of glass or the resinous binder holding the glass particles together. The oil composition, the kind of surfactants present, and the surface constitution of the solid used for oil entrapment, all affect the separation. The measurable parameters which determine the adhesion of oil to the solid are the contact angle of the oil on the solid surface, under water, and the interfacial tension between the oil and water. Pollution abatement can be achieved through control of these parameters, and the use of polymers in place of fiberglass.

Buoys Collect Noise Data in Arctic Ocean

An ONR contractor recently completed the development and installation of the data buoys for the collection of data on environmental acoustics in an ice-covered ocean. Mr. B. M. Buck, president of Polar Research Laboratory, Inc., has concluded the design/development and installation phases of his ONR program and is ready to start the data analysis phase on the most comprehensive collection of ambient noise data ever undertaken in the ice-covered Arctic Ocean.



U. S. GOVERNMENT PRINTING OFFICE : 1975 O 596-274

The array of ten buoys lies on approximately a 400-km radius circle centered on the main manned camp of the Arctic Ice Dynamics Joint Experiment. This array covers the major portion of the Beaufort Sea and is considered to be generally representative of the dynamical behavior of sea ice cover at large scales. Ambient noise is dominated by the sounds produced by the ridging, shearing, ripping, buckling, etc. motions of the ice and is correlated with the intensity of these atmospherically-driven phenomena.

Each buoy represents a major advance in the cost effective technology of measuring environmental sound for long time periods. The buoy records and transmits the acoustics in convenient third octave bands—ready for analysis and correlation with the ice dynamics data of AIDJEX. Also recorded are data on atmospheric temperature and pressure. These parameters are related to the intensity of ice dynamics and are useful in forecasting ice behavior, and ultimately, the environmental noise. All data are stored in the buoy and transmitted daily to the NIMBUS-F satellite for readout in Fairbanks and Goddard. The buoy is powered to last two years.

Thus, a new era is born in the satellite collection of ocean acoustics data and in the long-term low-cost technology of unmanned data collection.

NAVAL RESEARCH REVIEWS publishes highlights of research conducted by Navy laboratories and contractors and describes important naval experimental facilities. Manuscripts submitted for publication, correspondence concerning prospective articles, and changes of address, should be directed to Code 730, Office of Naval Research, Arlington, Va. 22217. Requests for subscriptions should be directed to the Superintendent of Documents, U.S. Government Printing Office, Washington, D.C., 20402. Subscription price: $6.50 per year in the U.S. and Canada; $8.15 per year, foreign; $0.55 per individual copy. The issuance of this periodical approved in accordance with Dept. of the Navy publications and printing regulations.

NAVSO P-510

IN THIS ISSUE

VOL. XXVIII, NO. 11

Cover Caption

Diver in background, performing measurement of flow resistance helium in the lung at 1600-foot "depth." See page 19.

DEPARTMENT OF THE NAVY
OFFICE OF NAVAL RESEARCH
ARLINGTON, VA. 22217

OFFICIAL BUSINESS

PENALTY FOR PRIVATE USE, $300.

POSTAGE AND FEES PAID
DEPARTMENT OF THE NAVY
DoD-316

U.S. MAIL