OFFICE OF NAVAL RESEARCH

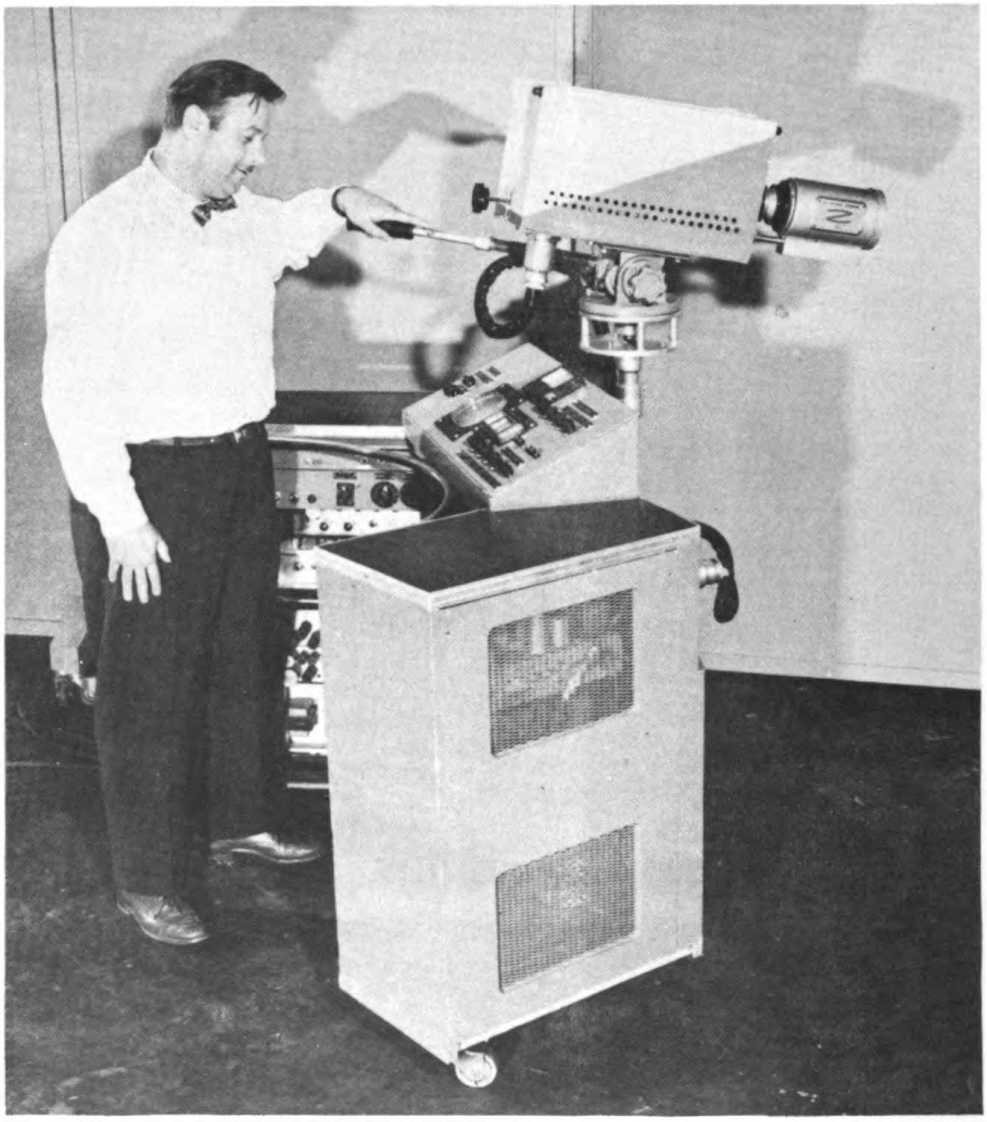
JULY 1956

RESEARCH REVIEWS

NAVEXOS P-510

DEPARTMENT OF THE NAVY

WASHINGTON, D.C.



Master console and camera (image orthicon)
for Instructional TV System

Television Research and the Study of Man

Ken Thomas
Television Requirements Coordinator
U. S. Naval Training Device Center

In October 1955, a group of doctors in Chicago saw the first American motion pictures taken of the inside of the living human eye. It is notable that this film was taken with a television camera manned by photographers of the Naval Training Device Center (formerly the Special Devices Center)—a field activity of the Office of Naval Research.

The kinescope recording these doctors saw was the "pick-up" of a television camera, supplemented by the recorded voice of a research scientist who described in detail the pictorial information recorded by the camera. The experience was a memorable one for the small group as they saw and heard about the human eye—for they were able to study on an intimate basis such features as the retina, the end of the optic nerve in magnification, and a blood vessel pulsating with the patient's hearbeat (Figure 1). In their enthusiasm, these doctors predicted that the experimental technique so vividly demonstrated might readily lead to an improved method of studying the blood circulation and many other physiological processes in the intact human being.

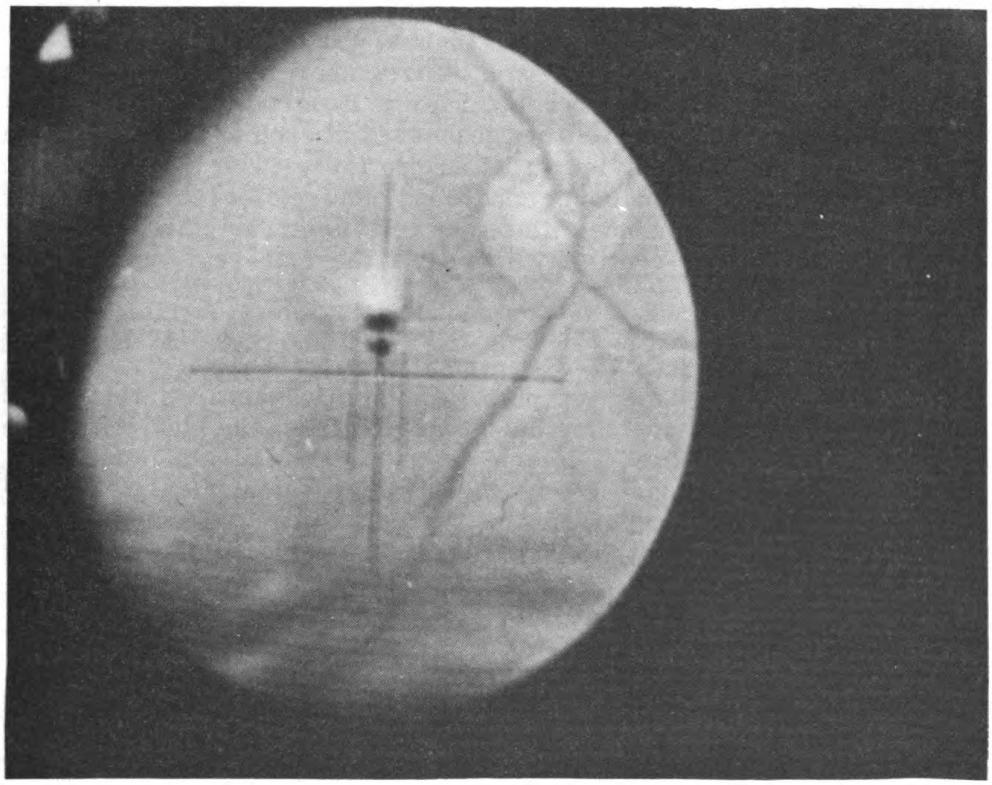


Figure 1 - Intimate view of the human eye as revealed by a kinescope record. The above is a photo made from the kinescope of an eye operation televised in Chicago.

The motion picture of the inside of the human eye is a highlight of a joint project on television and the study of man that began back in 1952. The Navy's Training Device Center and the National Institutes of Health joined forces on a program that began with a study of the communication complex in the new NIH Clinical Center at Bethesda, Maryland. The result of this study was a report entitled, "An Analysis of Television for the National Institutes of Health." The basic recommendations of the inclusive survey were that closed-circuit television be considered as a "system," and that the principle of flexibility in a proven originating system developed by the Navy be applied to the distribution system for the Clinical Center.

INSTRUCTIONAL TELEVISION SYSTEM

The Navy had developed a TV originating system, known as the Instructional Television System, as a research tool for testing television in a range of teaching and training situations. This I-TV-S consists of three portable consoles and cameras employing both image orthicon and vidicon picture tubes. Each console powers its own detachable lighting fixtures, and it may be operated as a multiple-camera or single-camera system. In routine operation it requires only one cameraman-technician per console.

Experiments were conducted at the NIH Clinical Center in the next step of the project to demonstrate different types of television and learn which of its many uses could aid in improving the volume and accuracy of communication. Materials from the experiments and earlier phases of the project were developed into a 16-mm sound film entitled "Electronic Image Processing." The film report, approximately 40 minutes in length, is listed as ONR Technical Film Report Number ONR 2-55 (unclassified). Kinescope recorded sequences of each experiment shown are presented in sufficient length to enable the viewer to evaluate the effectiveness of television under conditions required for the experiment. Motion picture film sequences illustrate television equipment arrangements and provide for continuity in telling the four-year story of the project.

In general, the film illustrates in actual and simulated situations the various roles of the I-TV-S as a medical research tool. The system is first shown playing an active part in scientific research relating to the visibility problems in using the ultracentrafuge; the pictures on the monitor and the off-screen comments are recorded simultaneously. Television is demonstrated as an unobstrusive communication instrument in formal educational activities where natural lighting is necessary, and where it is desirable to have the camera beyond the boundaries of audiences and speakers. As the unobserved observer, both in institutional and remote situations, television plays an important part in recording the unprejudiced behavior of subjects. Observations and recordings by television in the field of microscopy are tested. Color sequences show several fields of instruction in which color plays a vital part.

This film report itself was an experiment to learn whether it had advantages over a written report in communicating audio-visual project data to the field. The reaction of reviewers would indicate that the experiment proved successful.

The work done with the National Institutes of Health serves to prove the application of the I-TV-S to medical research and its adaptability to meet the requirements of a specific scientific situation. Its present and potential use as an important tool in medical training and communication is becoming more evident as greater applications are demonstrated and more and better equipment is evolved. As a research instrument the I-TV-S introduces flexibility to closed-circuit television; it is the TV that goes to the users—not the other way around.

Flexibility in a distribution system has given instructional closed-circuit TV the means to deliver signals to natural receiving locations in the using institutions. Such a system is under development now at the clinical center of the NIH. This system will distribute black-and-white and color TV signals in required bandwidths. It will distribute other than TV signals, such as, facsimile signals and electrocardiograms and electroencephalograms. Furthermore, it will distribute two or more signals from natural points in the institution to portable equipment in natural receiving locations.

The intense lighting associated with the TV studio facility is undesirable, and usually unnecessary, particularly in locations occupied by instructors. Natural and near-natural lighting, a condition favoring instructor acceptance of television, has been used successfully for both live and recorded television.

Teaching habits of the instructor have been primary considerations in developing methods for instructional closed-circuit television. In most situations the instructor is given the opportunity to manage his own TV presentations. His notes, in the form of an outline, are used for reference purposes by him and his one or two assisting cameraman-technicians. TV receivers are so placed that he may see pictures in rehearsal. Communication between him and his technicians is, insofar as possible, in the language of the instructor. His control extends through his actual presentation because he can see what he is presenting. Simplified methods and equipment have put the instructor—not television—in the dominant position. The methods of presentation aid greatly in instructor training, and the projection of a kinescope recording of his practice work permits him to concentrate on his own presentation.

Television equipment requirements should present, at the most, only minor problems to equipment manufacturers. However, in some fields such as medical research where requirements can be severe, or unique, there is need for more equipment research. Possible requirements for fixed-position equipment or a studio-type facility, may be derived from an analysis of the total communication problem in the planning institution.

USE OF I-TV-S SYSTEM

There are many inherent advantages in the I-TV-S (Figure 2) for the instructional situation over standard, commercial broadcast systems. It is a self-contained system, more compact and rugged in construction, and low in cost and personnel requirements. It is completely portable, ready to plug in and operate ashore or aboard ship. The cost is considerably less than that of commercial systems and the number of

men required for operation and maintenance is less than one-third. Training and operation are also greatly simplified for an instructional closed-circuit TV.

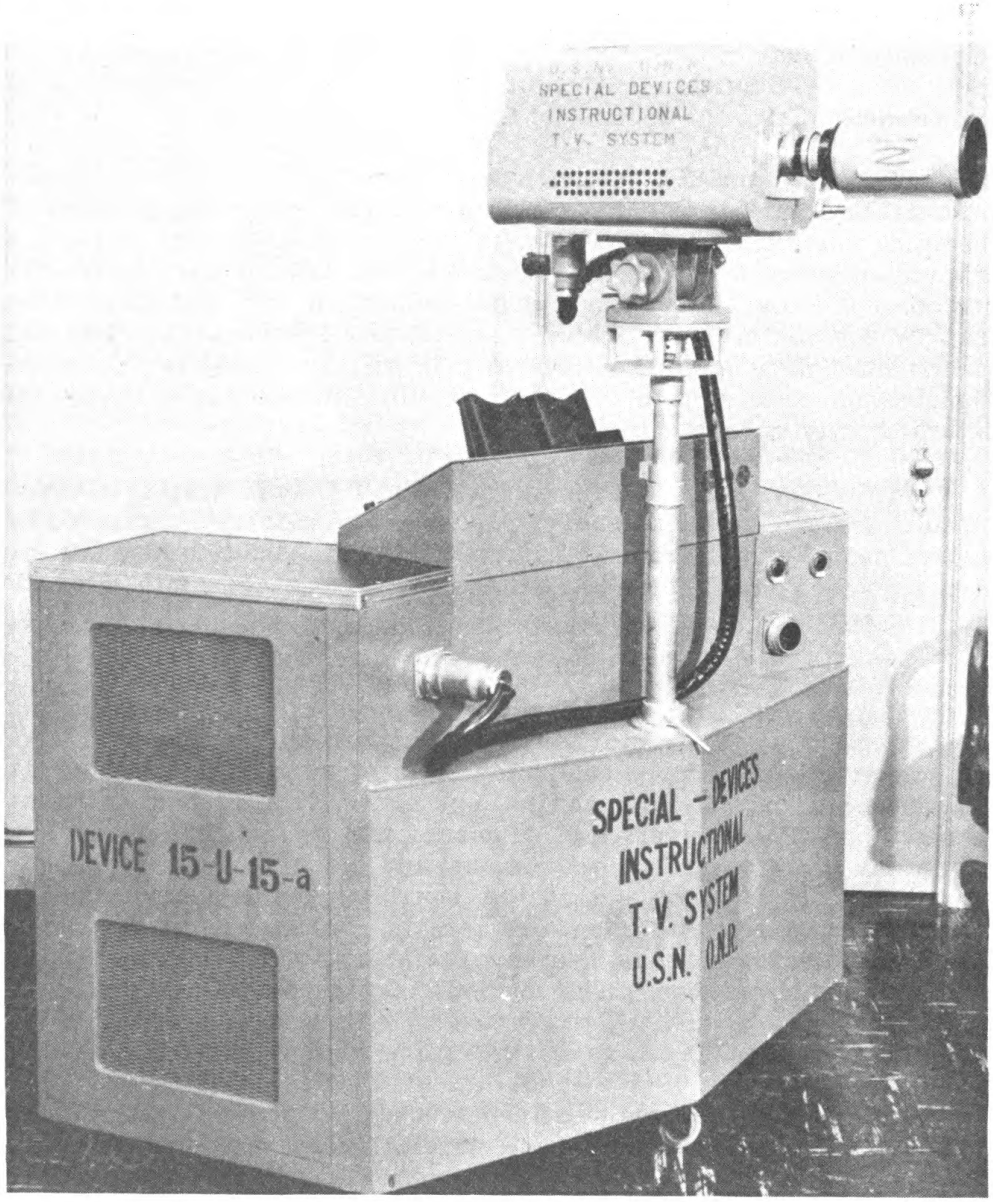


Figure 2 - The I-TV master console and camera.

In addition, the I-TV-S will distribute closed-circuit signals from nearby or remote activities, transmit internally-originated signals for closed-circuit distribution to those activities, and provide for the reception and distribution of broadcast television programs within the organization.

When the development of flexible TV equipment began at the Naval Training Device Center back in 1952, it became apparent that the extent of operating usefulness of instructional closed-circuit television would be influenced by the known types and numbers of real applications. Since

then an analytical approach to aid in the search for real applications has been applied, perhaps most consistently in medical TV research. Because results indicate an encouraging increase in the knowledge of real applications, a brief discussion of the approach may be useful.

As a first step, certain elements in conventional teaching situations having adverse effects on either the quality or quantity of communication are identified as limitations in the visibility, accessibility or availability of teaching materials, or limitations in the availability of qualified instructors. Then consecutively, corrective characteristics of I-CC-TV are considered.

To effect improvements in visibility, image-size amplification, usually part of the conventional TV demonstration, has proven useful, is fairly well known and its values are appreciated. Light amplification can improve visibility. For example, with TV and light amplification a nighttime scene may be presented in useful detail on a TV receiver. More types of information may be gained with contrast amplification. With television, a microscope and contrast amplification to test the effect on visibility of a radio audiograph specimen, finer particles were seen more easily and in greater number than through the microscope.

POSSIBILITIES OF SYSTEM

To correct for a limitation in accessibility, pictures and sounds of teaching materials, inaccessible because of space limitations or of the dangerous nature of the materials, may be distributed by television to convenient receiving locations. A shortage of instructors or teaching materials may be compensated for by live or recorded television. It may be noted that the instructor's recording of his own presentation may, when he is not available, be his own expert substitute.

Certainly, there are obstacles to the widespread use of instructional closed-circuit television. Like other media when they were new, this system was confronted by such deterrents as inertia, incomplete research findings, and lack of funds and personnel. Unlike other media, it has met obstacles peculiar to it, and one of them is "live" color. However, the general opinion is that color is essential only when color itself is teaching material, as in certain medical subjects.

Another obstacle to widespread use is the confusion in objectives of broadcast educational TV and I-CC-TV. A difference in objectives, which may be learned from a comparison of elements in their two basic and unlike positions, indicates in turn differences in requirements for equipment, methods, personnel, funds, and housing. It should help each, if they were considered to be separate in concept and be freed to go their own separate ways in the search for real applications.

Recently, American viewers have been yielding without apparent resistance to a new-found power of television, sensed whenever it turns away from entertainment and deals in reality. This public thirst for reality is an unmistakable sign of great potential for instructional closed-circuit television, a field in communications where reality is a cardinal quality. And this raises a question: "Will the millions of Americans who must be trained in times ahead accept training that does not recognize their preferred communication habit?" Instructional closed-circuit television has arrived with proof of effectiveness and with the TV-communication habit well established on the national scene; the new system is ready for consideration.

Algal Growth Unlimited

Dr. Robert W. Krauss
Department of Botany, University of Maryland
College Park, Maryland

Now it is well established that at least as much photosynthesis takes place in the ocean as on land. It may be difficult to realize that the trackless ocean wastes are in reality vast cultures of microscopic algae producing a tremendous tonnage of food daily. Reliable estimates indicate that 200 billion tons of carbon are "fixed," or converted, from gaseous CO_2 in the atmosphere to organic compounds such as sugar, protein, and fat by the algae. This process provides the energy source for the food chain which supports the fishes and ultimately supplies a major resource for human food.

The biochemical system by which such a great tonnage of carbon is assimilated is under continual study. There is no doubt that the algae are able to convert at least 20 percent of the visible light which they absorb into chemical energy. As a result some can double in numbers as much as 12 times every day for an indefinite period. This means that one cell which begins growing at noon today could produce over 1000 cells by noon tomorrow if optimum growing conditions could be maintained.

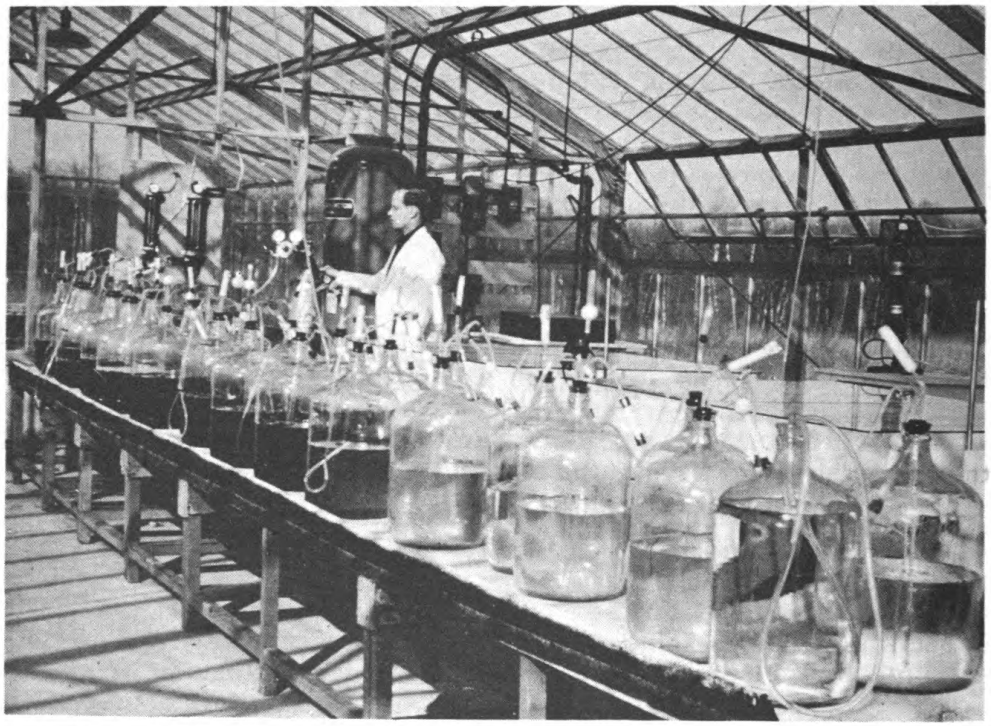


Figure 1 - Growing of algae in greenhouse-type laboratory.

An immediate question arises. Why are not the oceans filled with algae, or what limits the algae from growing as rapidly as they are able? It is quickly obvious that in replying to such a question one need only point to the low levels of the important nutrients such as nitrogen and phosphorus found in both oceans and fresh waters during most of the year. Other factors such as temperature, carbon dioxide, and light may be far from the best for the growth of algae.

Careful study of any given oceanic area can establish which of these factors may be limiting growth at any given time. Such variable factors are controlled by season, currents, and offshore drainage, and usually some one of these factors is inadequate to support more than a small, dilute population of algae. However, it has occurred to some workers that if algae could be supplied with an optimum environment, they might be cultured in large quantities and provide a much needed supplement to world food and energy reserves. At present, studies are under way in the Department of Botany at the University of Maryland to explore the feasibility of such mass cultures and to examine the factors which limit growth when the primary objective is continuous maximum yield.

A number of algae have been studied, but primary attention has been given to Scenedesmus and Chlorella. These single-celled organisms grown in large shallow vats (Figure 1) yield a maximum continuous harvest when at concentrations of near 167,000,000 cells per cubic inch of medium. When grown under continuous artificial illumination in 80-gallon vats, 0.3 to 0.4 lb of algae can be harvested daily.

Repeated analyses (Figure 2) of the algae for the elements absorbed from the nutrient solution have made possible good estimates of the rates of removal for each of the nutrients. This in turn has made it possible to prevent the medium from becoming depleted by replenishing daily of the elements absorbed by the algae. This may be done with considerable accuracy because oversupply would soon raise to damaging levels some of the essential but toxic elements such as copper, manganese, and cobalt. Daily harvests have been obtained over a period of three months from cultures receiving only recycled medium plus the supplement of nutrients removed by the algae.

The elements which have been found necessary in the replacement medium are: potassium, magnesium, calcium, nitrogen, phosphorus, sulphur, iron, manganese, cobalt, copper, and zinc. All are constituents required for higher plant growth but in rather different ratios. The utilization of these inorganic elements in supporting a system of organic metabolism results in a plant body which normally has the highest known protein content of any plant--between 50 and 60 percent. It is this high-protein level, which has been shown to have an equally high nutritive quality, that makes its use as food a distinct possibility. Of considerable interest is the pathway by which this protein is formed from the nitrate in the nutrient solution.

Studies have shown that the algae are strongly oriented toward protein synthesis. Experiments with cultures grown under conditions of nutrient deficiency maintained their protein content for long periods.

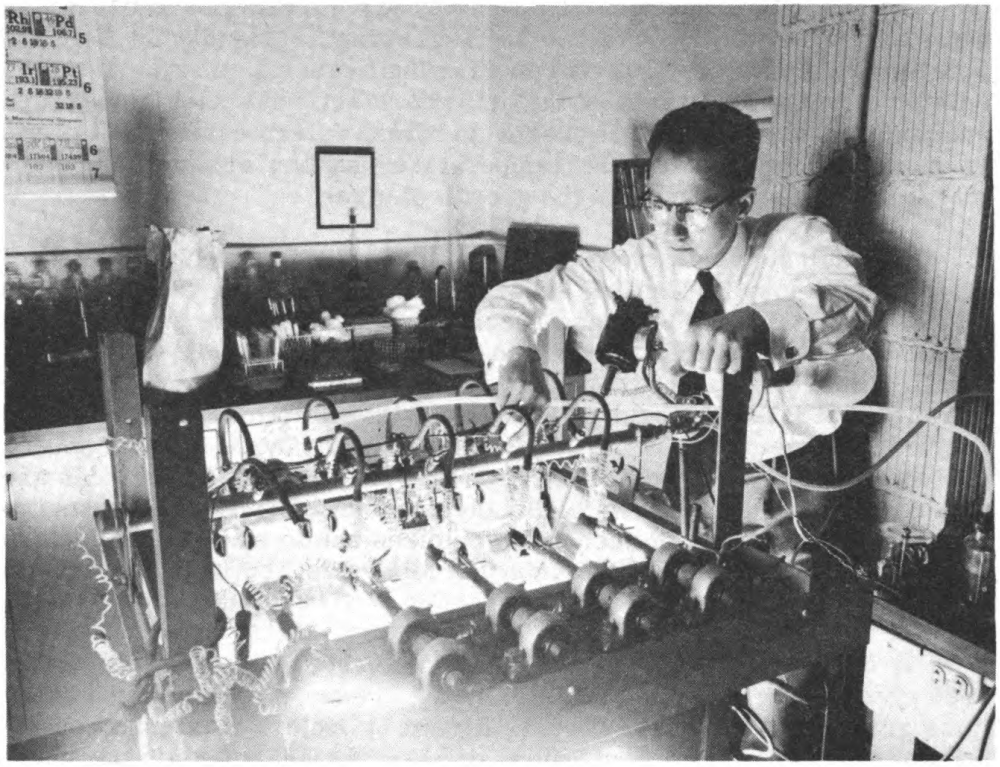


Figure 2 - Culture apparatus used at the University of Maryland for the determination of rates of ion up-take by algae.

Only nitrogen was required to maintain the protein once it was formed. In none of a series of experiments could the algae be forced to divert significant amounts of nitrogen into compounds other than protein--an encouraging observation when food value is considered.

These and other studies with the algae are supplying information which is bringing closer the day when they may be more usefully employed for our benefit. At the same time we are increasing our understanding of the organisms which already play so critical a role in the life of the oceans.

New Information Retrieval System

At a meeting of the Special Librarians Association held recently in Pittsburgh, Pa., Dr. Mortimer Taube of Documentation, Inc. presented a paper on a storage and retrieval system developed under a contract with the Office of Naval Research. The broad field of the research conducted by Dr. Taube is "New Methods for Processing Information." More than 1000 special librarians who attended the meeting saw the operation of EDIAC, an electronic device for information association and content which instantaneously retrieves information previously stored in it.

Improving Personality Evaluation Procedures

Erwin K. Taylor
Personnel Research and Development Corporation
and
Roy Hastman
Psychological Research Services, Western Reserve University

Every supervisor, whether of military or civilian personnel, must at one time or another either formally or informally evaluate the performance of those who serve under him. Periodic formal evaluations, like the fitness report, first became popular during and after World War I. Since then, formal evaluation procedures have had a stormy history with numerous critics and few champions. While no evaluation system is without its faults, the critics of formal evaluation usually fail to realize several things:

1. That the evaluation of his subordinates for one purpose or another is an inherent part of almost every supervisor's job;
2. that if formal periodic systematic evaluations are not made, informal unstandardized ones will be;
3. that weaknesses of formal evaluations are not created by formalization; they exist in an unknown state and probably to a greater degree when administrative action is taken in the absence of a formal system.

Informal evaluations may conceal inequalities and injustices, but they do not eliminate them.

The most popular type of formal system involves the use of so-called "graphic rating scales." While many varieties of graphic rating scales exist, they have in common the inherent characteristic of representing each trait, attribute, or behavior to be evaluated—by a line on which a mark is made by the rater to indicate the ratee's standing on the trait under consideration. Scales may be (and probably have been) prepared for almost every conceivable attribute of the human being. Some deal with characteristic traits such as "honesty," "loyalty," and "ambition." Others deal with such attributes of performance as "quality," "quantity," and "accuracy," while still others concern such work habits as "punctuality," "attendance," and the like.

Points on the scale may be indicated by a series of numbers ranging anywhere from five to one-hundred points, by a series of adjectives such as "poor," "fair," "good," and "excellent," or by any one of a number of other means. A considerable amount of rating research has been conducted during the period between the two World Wars and since. From these studies, we have learned a great deal about the construction of rating scales and about their faults and weaknesses. We now know that generally raters tend to be overly lenient in evaluating their subordinates. One study of fitness reports, for example, showed 70 percent

It is important to separate the amount of work a person does from how well he does it. In this rating we are interested only in how much each worker does. It does not matter how good or bad it is. When making this rating, evaluate the overall volume of completed assignments and disregard the quality and accuracy of the work or whether corrections, additions, or reprocessing may be necessary to meet minimum standards.

Frequently is rather unproductive; fails to produce acceptable amount of production.	Periodically falls below the average amount of production maintained by his co-workers.	Turns out an average amount of work.	One of the faster workers who always produces his share and occasionally exceeds the average quota of the group.	A pace setter when it comes to the amount of work which is completed; his production is always near the top of the list when compared with others.

Trait rating scale from the graphic rating form.

of naval officers to have been rated as belonging in the upper 10 percent of the navy officer population.

We know, too, that raters fail to differentiate among the characteristics of the subordinate. They tend to form a general impression, good or bad, and then assign high or low ratings to all traits on the basis of the overall impression. This is the well-known "halo" effect.

To reduce these major faults of graphic ratings, Stevens and Wonderlic in 1934 suggested that supervisors instead of rating a single subordinate on all of the traits of the scale—before proceeding to the next individual—reverse the procedure, and evaluate all other subordinates on the first of several traits before rating any of them on the second. When conclusive research by the Department of the Army in the closing days of World War II demonstrated that conventional rater training accomplished little in overcoming the two most serious flaws of conventional rating, Taylor and Manson suggested the use of a trained personnel technician to assist the supervisor in the evaluation of his subordinates. While the results of Stevens and Wonderlic as well as those of Taylor and Manson provided evidence that their respective suggestions had merit, controlled experiments designed conclusively to test their superiority over the traditional rating procedures were needed. This study was undertaken to determine under controlled conditions the way in which format and administration affected the results of a rating system.

The subjects used in the study were the civilian clerical personnel of the Navy Finance Office in Cleveland, Ohio. Eight scales were prepared to cover the job-related activities considered important by both the project technicians and members of the personnel administration of the Navy Finance Center. The chart shown here illustrates one of these scales. The best available techniques of scale construction were employed. Trait names were avoided and operational definitions substituted in their stead. Behavior descriptions were employed to designate the various levels of each scale rather than misinterpretable numbers or adjectives.

The scales were assembled into two different rating forms. One of these was the traditional graphic rating on which a supervisor rated one of his subordinates on all eight traits before rating his next subordinate on any of them. In the other forms the same eight scales were assembled in such a manner as to require the supervisor to rate all of his employees on the first scale before rating any of them on the second.

The traditional graphic was administered by sending the necessary number of forms to each of the supervisors involved along with an especially prepared rating manual, having them accomplish their evaluation at their convenience and return them through inter-office mail to the personnel department.

Another group of supervisors followed the same administrative procedure, but used the Stevens-Wonderlic form rather than the traditional graphic. A third group of supervisors, also using the Stevens-Wonderlic form, met in small groups with a trained personnel technician and accomplished their evaluations under his guidance during the

course of the meeting. In the fourth group, the personnel technician visited personally with each rater. During the course of these rating conferences, the personnel technician informed the rater of the typical pitfalls in graphic ratings. The technician remained with the supervisor while the latter made his evaluations and provided guidance and assistance throughout the process.

The number of people rated in each group varied from a low of 145 to a high of 193. There were 22 raters in each of the groups except for Group I, in which there were sixteen. For each group of ratees, the second-level supervisor rated the same subordinates, using the same technique as was used by the immediate supervisor in each case. For the first two groups, traditional graphic and traditional Stevens-Wonderlic, there were ten second-level supervisors in each. In the last two groups, the group and individually administered Stevens-Wonderlic, there were eleven second-level raters per group.

The individuals rated represented three job families: (1) Clerical, (2) auditing-accounting, and (3) tabulating machine operations.

Statistical analyses were conducted for the purpose of comparing the four rating techniques. In the interest of brevity, results only with respect to the traditional graphic and the administered Stevens-Wonderlic are presented below.

TRAIT INTERDEPENDENCE

The tendency of raters to evaluate subordinates on the basis of generalized impressions rather than to be able to distinguish degree of relationship of each trait rating with every other. Such relationships were measured by means of the product-moment correlation coefficient which ranges from minus one to plus one, with zero indicating a complete lack of relationship. The use of eight scales yielded 28 intercorrelations within each group. The higher the intercorrelation between trait ratings, the more the rates can be considered to have been judged on the basis of a general impression rather than rating each trait independently.

For this group as a whole, the intercorrelations averaged nine points higher for the traditional graphic than for the administered Stevens-Wonderlic. From this we may conclude that the use of a trained technician to assist the rater in making his evaluations enables the supervisor better to distinguish between the separate aspects of an individual's performance than is the case with the traditional graphic rating.

LENIENCY

The tendency of raters to be tough or lenient in rendering their evaluations can be measured by the magnitude of the average rating given on each trait by the raters in each of the experimental groups. In six of the eight traits in the study, the mean given by the Stevens-Wonderlic group was lower than that given by the graphic group. These six differences range from a low of .02 to a high of .62, with an average for all eight scales of approximately .21.

Another factor which may have affected the relationship was the magnitude of the means themselves. In all cases the average rating given was under 6.00 on a ten-point scale. This in itself is quite unusual in graphic ratings. Usually, mean graphic ratings are considerably closer to the top of rating scales. Three possibilities suggest themselves as explanations of this phenomenon:

- Considerably more care was taken in the construction of these scales than is usually the case. It may be that the nature of the scales functioned to reduce the leniency effect.

- This was known by the raters to be a research study. It has been demonstrated elsewhere that in traditional graphic ratings, the leniency effect is less evident in an experimental than in an operational situation. This phenomenon may have been operating in this situation and may have had a differential effect on the two types of ratings under consideration.

- It may be that the supervisory staff of the Navy Finance Center has been better trained by that organization's personnel department than most raters and hence less apt to be overly lenient.

DISPERSION

The function of any performance evaluation is to discriminate among workers with respect to their productivity. A rating system which places everybody or almost everybody into the same category (as does the present Federal Civil Service rating system) is practically useless. Figures 1 and 2 compare the distributions of ratings rendered within the two systems for the scales dealing with the amount and acceptability, respectively, of work produced. Note the extent with respect to both scales to which the traditional graphic concentrates individuals in a single-rating category.

Dispersion of variables is customarily measured by a statistic known as the "Standard Deviation." This statistic was larger in six of the eight scales for the administered Stevens-Wonderlic than for the graphic ratings. Again none of the differences is startling, but the trend indicates a slight superiority in favor of the administered Stevens-Wonderlic.

CONCLUSIONS

The marked superiority claimed for the Stevens-Wonderlic form of rating was not substantiated by this study. Except perhaps for the lower intercorrelations, the gains achieved by the Stevens-Wonderlic format, even when individually administered, do not appear adequate to justify their added cost.

It should be noted in this connection that the results obtained from the ratings in this study have been rather unusual in several respects. First, the magnitude of the intercorrelations in all four situations was considerably less than is generally found in merit-rating research. Secondly, the means of the ratings were generally near or below the

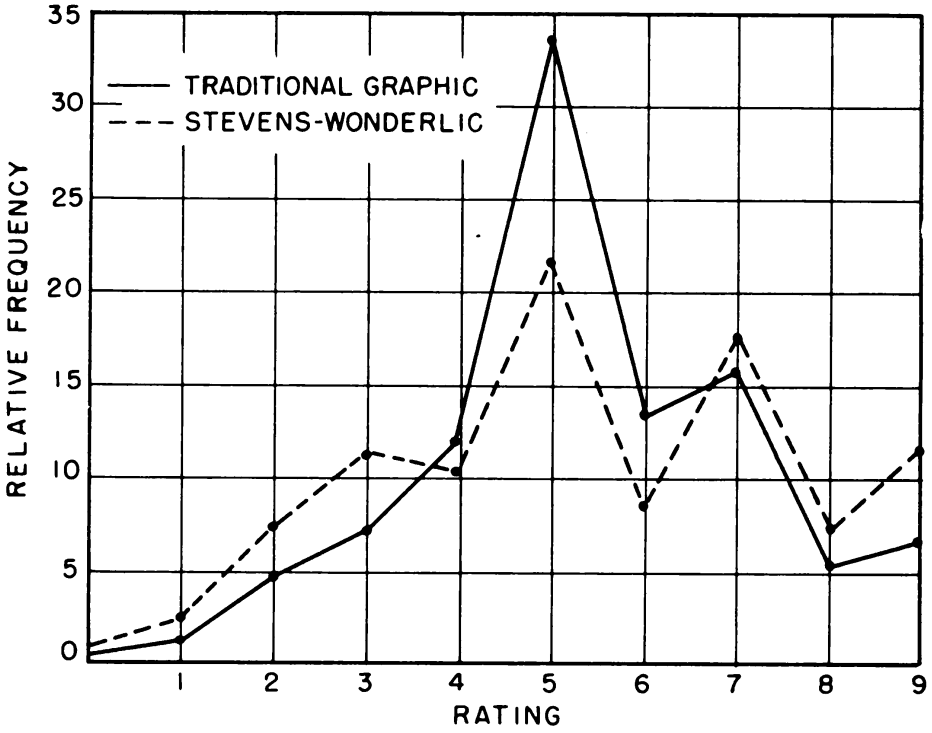


Figure 1 - Comparative relative frequency curves between administrations for "quality of work."

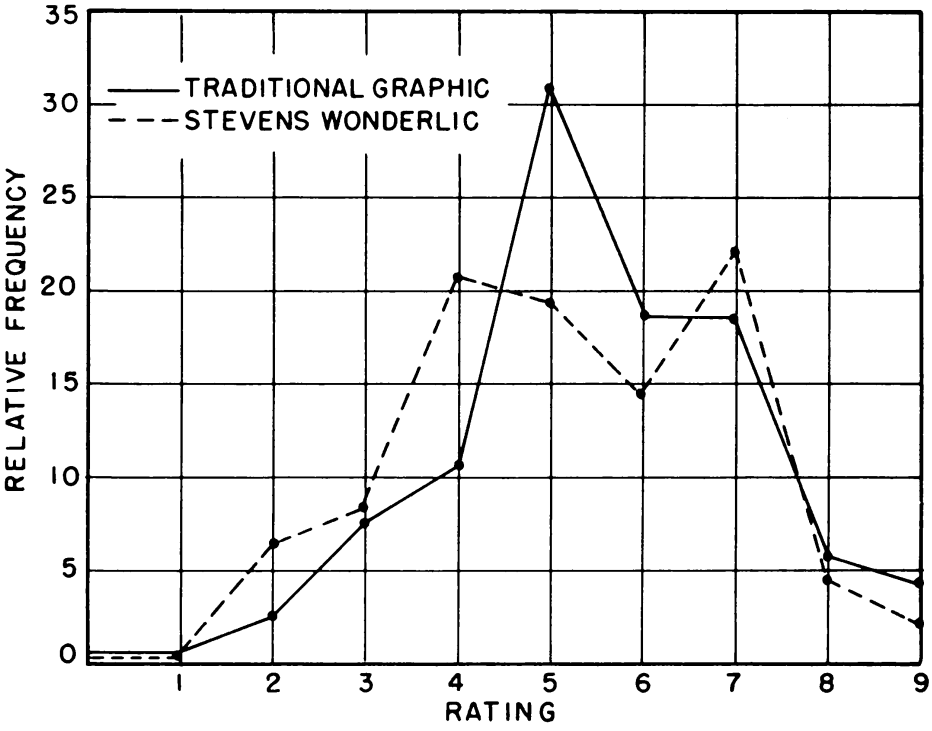


Figure 2 - Comparative relative frequency curves between administrations for "acceptability of work."

midpoints of the several scales. Thirdly, the entire range of the scales was used rather than merely the upper portion, as is commonly the case. Fourthly, statistical measures of skewness (lumping the cases at one end or the other rather than symmetrically about the center) were slight in all cases. Ordinarily in graphic scales there is a lopsided distribution, with most cases falling in the upper scoring categories. For all four situations, the curves were practically symmetrical.

Two conclusions can be drawn from this study. One is that most of the undesirable features of graphic ratings can be substantially reduced. The second is that format and nature of administration make only a small contribution to these improvements.

We entertain the hypothesis that much of the improvement may be attributed to the nature of the several scales and to the care with which they were constructed. Possible contributing factors may be the training previously afforded the supervisors by the Finance Office's personnel department.

To test this hypothesis, it would be necessary to repeat this experiment. This time we would hold the format and administration constant and vary the manner of scale construction.

New Name for ONR Field Activity

The Special Devices Center of the Office of Naval Research has received a new name. Effective 30 April 1956, this naval activity became the U. S. Naval Training Device Center. The Center remains under the military command of the Commandant, THIRD Naval District and under the management and technical control of the Office of Naval Research, Washington, D. C.

The need for a more realistic title to conform with the Center's mission and operation prompted the name change by the Navy. "Frequent inquiries reaching the Center concerning research not pertaining to training or training devices indicated widespread misconceptions of the purpose of the Center both within the Navy and the other services," according to CAPT C. H. S. Murphy, Commanding Officer and Director of the Center.

The main purpose of the U. S. Naval Training Device Center is to provide for development and production of training equipment to improve the effectiveness of military training in the Navy and throughout the Armed Forces. Through training aids, devices, and training systems developed by the Training Device Center, Navy personnel are trained more quickly, more economically, more safely, and more effectively. By this means the Training Device Center contributes significantly to the operational readiness of the Fleet to carry out its mission in time of war.

The training aids and devices vary in design and in use from the somewhat simple to the highly complex. Some are designed for mass training, some for individual training, and others for team training. But they are all designed to serve one definite purpose—to assure more effective training.

A Better Military Rating System Through Peer Ratings

E. P. Hollander
Assistant Professor of Psychology
Carnegie Institute of Technology

Psychological measurement may be defined simply as the evaluation of some aspect of an individual's behavior, or potential for behavior. Such things as intelligence, personality, interests, and attitudes are typical of those which fall within this definition. More usually, the psychologist measures these through a sample of behavior obtained from some kind of specialized test.

In long-range terms, one of the major practical goals of psychological measurement is the prediction of future behavior. When we say, for example, that a person has a high intelligence test score, we are implying something about his probable performance in a future situation. This is essentially the principle underlying the use of entrance examinations for admission to various schools. It is presumed—normally with some valid basis in evidence—that a person with a sub-standard score ought not to be admitted; he has such a low probability of successful performance in the school that his admission would be wasteful.

TEST VALIDITY

When a test proves out by actually predicting what it ought to predict, this is important evidence that it is "valid." The literal "validity" of a test may be determined by relating it to relevant performances, like school grades, which we call "criteria." This matter of validity is quite critical to the psychologist, and also to the institutions which he serves. However good a test may appear to be, it is useless for prediction until it can be shown scientifically to have significant validity for the purposes intended. There is no escaping this issue, even if you think a test is good at face value.

One other matter is that of "reliability." This concerns the consistency of measurement, either over time or within the test itself. It would never do to rely on a test score which could fluctuate tremendously from one time to another. Thus, we must determine how consistently a test measures what it is supposed to measure, in addition to its validity.

Still another feature of this evaluation problem is the matter of multiple prediction. It is often the case that no single measure will yield the kind of over-all prediction that is required. This results in the use of a so-called "test-battery" which measures several aspects of the person which are related to the criterion. A good illustration of this is to be found in the battery of tests the Navy uses to select aviation trainees; no one test is relied on to do the job of predicting probable success in Naval Air Training. Instead four major tests are used—each of which has already been shown to have validity. These measure intelligence, personality, mechanical aptitude, and spatial visualization ability. The scores achieved on each are then weighted and pooled to give a general

score indicating the candidate's probability of success in the training course; aside from the physical examination, the score serves as a basis for deciding whether a candidate should be admitted to the program. Here we have a case of multiple prediction.

Even with the use of a test battery of the kind described, prediction is never perfect. Considering, however, the variety of factors which go to make up performance in some future situation the results are often surprisingly good. Still, the question which we must ask always is whether our prediction could be better. Can we add something to our battery in order to increase the accuracy of our evaluation and ultimate prediction? All of this boils down to the fact that psychologists are continually looking for new or additional ways of predicting future performance.

PREDICTION OF PERFORMANCE

Before we proceed to talk about one of these new ways, it is important to realize the complexity of the criterion problem itself. To this point, we have been dealing with a school performance criterion as if it existed in a vacuum. Actually, we have a hidden assumption about this criterion which sometimes escapes us; that is, we think of performance in training as having a relationship to performance, on the job, after training. And this is partly so. The evidence we have on this matter tells us that while there is no complete correspondence between performance of one kind and performance of the other, there is a positive relationship at some level worth heeding. Other factors come into play, however.

This leads us to the simple notion that a training program itself may serve as a basis for predicting the more distant criterion of performance after training. In other words, a training program may be as much a selection device as the tests used to admit people to training. This is not so unreasonable when you reflect on the fact that the selection tests are never perfect and that some people in training ought not to be there. We may simply extend our period of selection into the training cycle. This would mean that we would try to identify poor performers as soon as possible.

Rather than talk in general terms, let us consider a concrete situation in which we have carried on some studies, namely, Officer Candidate School (OCS). The problem here is pretty much as already stated. We use a battery of tests to select trainees for this school; we know that these tests are valid in predicting potential success in this school; and, we also know that some small percentage of those people admitted may not make the grade either as officer candidates, or as officers in the fleet. Our immediate question now is how to identify such individuals.

The administrator who is daily confronted with this question answers it by using aptitude test scores, course grades, and instructor ratings, of one kind or another. A group of administrators will often meet to go over the records of deficient trainees and take time to interview them personally. All of this information will then be pulled together and a prediction will be made, based upon multiple measures. The problem for the psychologist is not to make this decision himself,

but to provide the best set of measures for the use of responsible personnel. It is here that we come to grips with the business of finding new devices to improve the accuracy of prediction.

One of the things that our selection is apt to miss, which evaluation in training needs to pick up, is the way in which a man behaves toward, and is seen by, those around him. Certainly, we can agree that this has something to do with performance as a naval officer. One of the few ways we can get at this is by having men rated on personal characteristics or potentials by their superiors and by their peers. These latter ratings, referred to variously as "peer ratings," or "buddy ratings," represent one new way of evaluation. We have reason to believe that when a substantial number of a man's peers say, for example, "This man shows leadership qualities," this is worthy of note. Furthermore, there is evidence to show that this is a great deal more than just a sign of popularity.

OCS PROGRAM EXPERIENCE

Since the time of World War II, an increasing body of evidence has accumulated from OCS programs, indicating that the score obtained from some form of peer rating is a significant predictor of later performance in the field. At the Marine OCS, peer-rating scores have been found to be the best single predictor of success in OCS and combat performance. At Pensacola, we have found that peer ratings at Pre-Flight are a significant index of success in flight training. There are many more such studies which substantiate the same point. The reasons to account for this are basically the following: Peer ratings are based upon (1) close, personal contact, and (2) pooled judgment by many raters.

The studies completed to date demonstrate that peers are capable of making ratings which will add to the multiple prediction of these performance criteria. Actually, this makes a lot of sense. If you ask men in training to rate each other on leadership ability, and you pool these ratings, you are probably getting a picture of the man which is somewhat different from that obtainable from grades, or intelligence tests, or even superiors' ratings. These things are valid all right, but they require supplementary information. Peer ratings provide precisely this.

Still another aspect of this is the question of reliability—or the consistency of measure. In the case of peer ratings, this may be determined by calculating two scores for each individual, one based on the odd-numbered raters in the group, the other on the even-numbered raters in the group. This procedure simply assures randomness. We find that these scores are highly related, that is, that one half of the group rates each individual about the same as does the other half. We are entitled to say, therefore, that peer ratings are reliable measures.

Peer ratings of one kind or another are being widely used in many military training programs of the OCS variety. But several questions still remain regarding their most effective use. For one thing, we are interested in just how early in training we can administer a peer-rating form to give us substantially the same validity as we get from later ratings. It is apparent that an early, valid rating has much use to the

administrator deciding on the disposition of a borderline case. For another thing, we are interested in the kind of form which will give us the highest validity for the two kinds of criteria we have already considered.

At the Newport OCS, we have recently been conducting studies to deal with these problems. The training sequence is of sixteen weeks' duration and a peer rating is routinely administered about the middle of this period. This particular form calls for ratings of success in the future as an officer. Our study has used this form, and several others as well, administered at various stages of the program.

Right now we have data on the final OCS standing of our study population and, also, we know who failed from training. Therefore, for the prediction of total performance in OCS—the in-training criterion—we can calculate the validity of different forms administered at various times.

We find that certain forms administered very early in training have acceptably high reliability and significant validity in predicting this criterion. For example, one of our forms, requiring ratings on probability of success in OCS, yielded significant validity even where the groups to which it was administered had been together for only four or five days.

APPLYING PEER RATINGS

As a rule, our results indicate that forms may be used at the third week of training, or even earlier, to yield validity which is comparable to that obtained later in training. Furthermore, we have found that specialized forms may be used to good advantage in tapping certain specific features of the individual, like leadership qualities and motivation.

Since we do not yet have data on the fleet performance of our groups after graduation, it is not possible to determine the ultimate validity of a form indicating future performance as an officer. Insofar as we can offer information, however, it seems that this form could pay off in giving an early picture of the man in his probable relationships to others when placed in a command position.

One objection to peer ratings, which we have specifically studied, is the matter of their relationship to friendship. Some people continue to contend that these ratings are loaded with a friendship factor. From our work, we find that this is simply not so. Fewer than half of a man's acknowledged friends are rated "high" by him on the various forms. The correspondence, therefore, of high ratings to friendship is nowhere near one-to-one.

In addition to this issue, we often are asked whether group members might not resent rating each other. The answer is that they might, but they need not. This depends largely upon the way in which the task is presented. At Newport, we have found that excellent cooperation may be secured if: First, the instructions stress the importance of officers knowing how to rate others; second, the point is firmly made that ratings

are to be pooled so that no one individual's rating of another is critical in the over-all picture; and, third, members of the group are seated sufficiently far apart—for example, in columns—so that they cannot see each other's papers.

There are other aspects to this technique which space does not permit us to consider adequately. Two such aspects concern the size of the group, and the way in which ratings are scored. Regarding the first point, there is evidence to indicate that reliable and valid ratings may be obtained from training groups as small as fifteen and as large as thirty-five. As to the second point, the data suggest that several scoring procedures may be used—simple summation of "high" and "low" ratings, for one—without seriously altering the status that emerges for any given man.

The successful use of peer-rating procedures in the field encourages us to believe that they have decided utility for evaluation. But, obviously, peer ratings are not the answer to all evaluation problems. They represent simply one new departure which can assist us in the prediction of future behavior. Other measures are still required. Viewed practically, however, it is safe to say that peer ratings can make a significant contribution toward the early identification and elimination of potential failures and misfits.

Award to Retiring NRL Scientist

Mr. Sverre Gulbrandsen, Associate Superintendent, Chemistry Division, of the Naval Research Laboratory, was recently presented the Meritorious Civilian Service Award for "significant scientific guidance, inspirational leadership, and untiring devotion." CAPT P. H. Horn, Director of the Laboratory, made the presentation at a luncheon held in recognition of Dr. Gulbrandsen's retirement on 31 May, culminating 15 years of Federal service.

Fourth Annual Human Engineering Conference

The Office of Naval Research is sponsoring a conference on Human Engineering to be held at New York University on 17 and 18 September. This is the fourth such conference to be held and will be attended by representatives from military organizations, industry, consulting groups, and university laboratories. Joint sessions on human engineering are being arranged with the Instrument Society of America for 19 September. Papers to be presented during these joint meetings will cover areas of human engineering in systems design and in the design of control panels. This is the first such joint meeting with an engineering society.

Bacteria in Marine Sediments

Richard Y. Morita* and Claude E. ZoBell
Scripps Institution of Oceanography
University of California

Because of the multiple effects of bacterial activity on the chemical composition and other properties of marine sediments, and because marine bacteria cause changes in both organic and inorganic matter, the Navy is interested in learning more about the occurrence, kinds, and physiological activities of bacteria in the sea (Research Reviews, May 1955 - "Microbes with an Appetite for Iron").

Until recently only the uppermost strata of sediments from nearshore or relatively shallow water have been available for observation; very few observations have been made on the microbial population of marine sediments from any depth. Deep-sea expeditions, notably the 1950 Mid-Pacific Expedition of the Scripps Institution of Oceanography and the 1950-1952 Danish GALATHEA Deep-Sea Expedition, have now made it possible to demonstrate the occurrence of bacteria in some of the deepest known parts of the ocean.

The development of improved coring devices has provided the marine microbiologists (Figure 1) with a means of collecting sediment

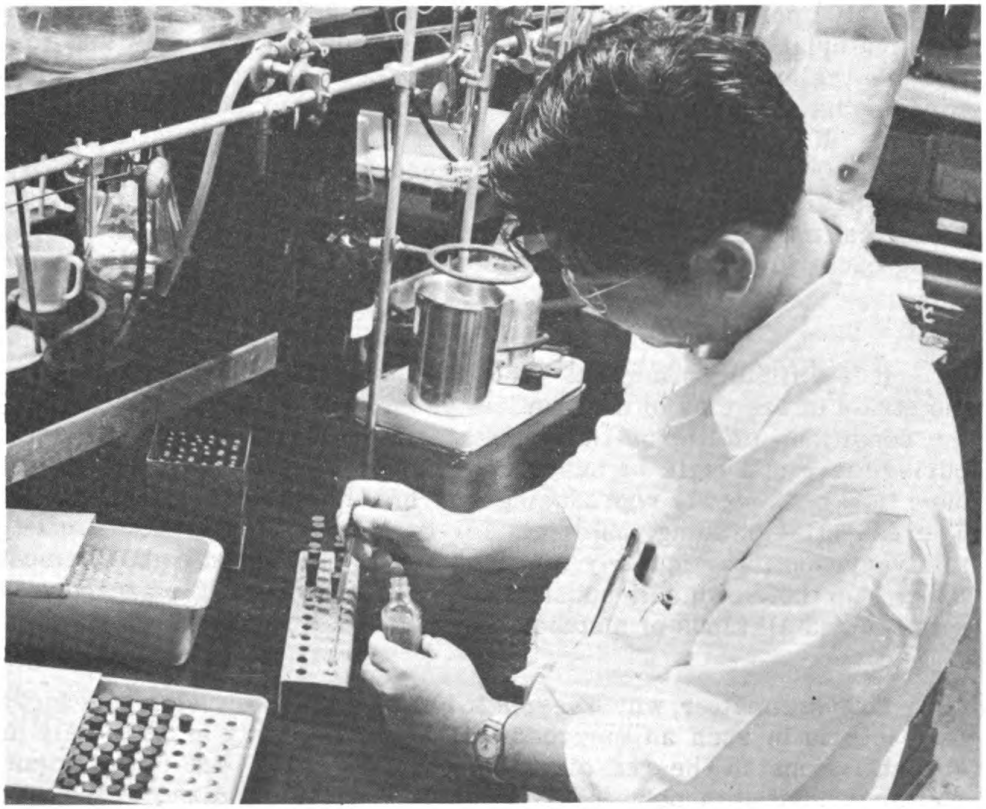


Figure 1 - Dr. Richard Y. Morita preparing bacteriological samples for high-pressure studies.

*Present address: Department of Biology, University of Houston, Houston, Texas.

samples several meters below the sea floor. It appears from geologic markers that the cores brought up from great depths had bottom-most strata that were deposited tens of thousands of years ago. Moreover, these marine sediments so snatched from the ocean bed in small samples contained living bacteria.

In the Channel Island region off the coast of Southern California a core of green mud 752 cm in length was obtained in 1950 with a modified Kullenberg piston-type coring device. Living bacteria were found throughout the entire length of the core. In the recently deposited sedimentary material at the surface of the core, there were from 10^6 to 10^7 bacteria per gram of mud. The bacterial population decreased with core depth or with increasing age of the sediment, but even at the bottom of the core, 752 cm below the mud-water interface, there were between 10^2 and 10^3 bacteria per gram of mud.

On the Mid-Pacific Expedition to the Marshall Islands in 1950, participating microbiologists analyzed 19 cores of pelagic sediments taken at water depths ranging from 1700 to 6000 meters. These pelagic sediments, consisting of red clay or globigerina ooze, were found to contain much less organic matter and far fewer bacteria than terrigenous sediments. The most recently deposited pelagic sediments occurring at the mud-water interface had bacterial populations of the order of 10 to 10^4 per gram. With increasing core depth where the sedimentary material was older, the bacterial population was less and less. In some the lower limits of the biosphere seemed to have been reached at core depths of four to seven meters where no bacteria could be demonstrated. It is noteworthy, however, that bacterial populations of the order of 10^2 to 10^3 per gram of mud were found in several samples of pelagic sediment from core depths of four to five meters. According to geological markers, some of these deeper pelagic sediments in which living bacteria were found were deposited more than a million years ago.

It is difficult to account for the occurrence of living bacteria in the strata of red clay of such antiquity, because we know so little about the conditions of life on the deep-sea floor. Have the bacteria been buried there in a state of suspended animation for a million years or have they been slowly reproducing and metabolizing through the ages? The essential building materials for bacterial reproduction are certainly present, particularly since bacteria may recycle the elements making up their own cell substance (protoplasm), but the reproductive process and all kinds of metabolic activities require energy.

Organic matter, which seems to be the most likely source of energy for bacteria in such an environment, is present only in extremely low concentrations in the red clay. The preponderant part of the organic matter deposited in pelagic areas appears to be oxidized at the mud-water interface. The small quantity of organic matter which persists in pelagic sediments must be highly refractory to bacterial attack. This is indicated by its very persistence as well as by its chemical composition.

Conceivably organic matter in minute amounts could be supplied or replenished in marine sediments by the activities of one type of bacteria in the presence of free oxygen or hydrogen. This possibility has provoked much speculation regarding the liberation of oxygen and hydrogen by electrical currents, radioactivity, or other mechanisms, but no thorough investigations have been made as yet.

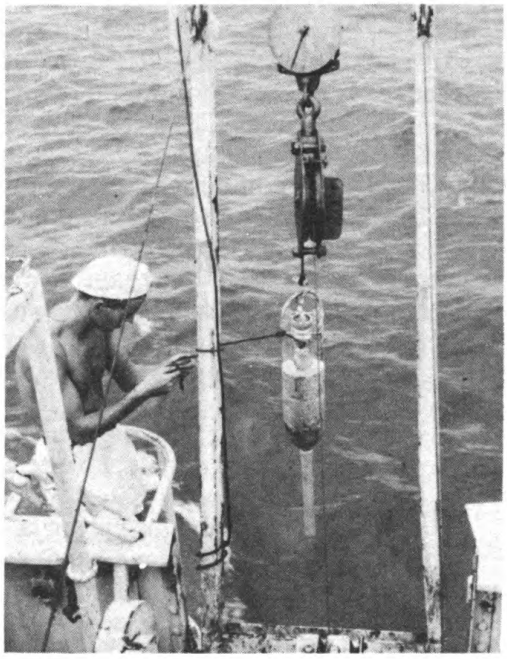


Figure 2 - Gravity coring device used aboard R.D.N. GALATHEA.

During the GALATHEA Deep-Sea Expedition (Figure 2) bacteria in abundance (10^3 to 10^6 per gram) were found in marine sediments at several stations where the water depth exceeded 7000 meters. Nine sediment samples were taken from depths exceeding 10,000 meters in the Philippine Trench and in the Kermadec-Tonga Trench. A bacterial population of 10^5 per gram of mud was demonstrated in bottom deposits from the greatest depth sounded, 10,280 meters. A good many of these deep-sea bacteria were barophilic, meaning that they grew preferentially or exclusively at high hydrostatic pressures characteristic of the deep sea. The pressure-depth gradient in the sea is approximately 0.1 atmosphere per meter, so at a depth of 10,280 meters the hydrostatic pressure is approximately one thousand atmospheres or about 15,000 psi.

Bacteria from these great oceanic depths grew in nutrient medium when incubated in a refrigerator at 3° to 5° C. This was not surprising, because the temperature in the deep sea is generally around 3° C, and most marine bacteria are active at refrigerator temperatures. It was surprising, however, to find that all of the deep-sea bacteria tested for their temperature tolerance grew much more rapidly at 15° to 20° C than at lower temperatures. Higher temperatures were not tolerated. Many of the deep-sea bacteria were injured by even short exposures to temperatures exceeding 25° C. Their heat-sensitivity causes us to wonder how many of the bacteria in the deep-sea sediment samples failed to survive subjection to the higher temperatures characteristic of surface water and the ship's laboratory. Unfortunately, there were no provisions for keeping them cool while the samples were being hauled to the surface and examined in the laboratory, where the temperature ranged from 30° to 40° C.

Some of the bacteria found in sediments taken from great oceanic depths grew readily in a nutrient medium incubated at surface conditions as normal as one atmosphere. Whether such bacteria represent surface-dwelling species which have settled to the sea floor, and whether they are physiologically active at both atmospheric and deep-sea pressures, have

yet to be determined. Investigations currently in progress indicate that hydrostatic pressures characteristic of the deep sea have pronounced effects on the reproduction, metabolic processes, and viability of bacteria.

That bacteria are physiologically active in deep-sea sediments is indicated by the chemical composition of profile series of mud cores. In certain of these the sulfate content decreases with core depth, ostensibly because of the activities of sulfate-reducing bacteria shown to be present. A gradual decrease with core depth in the organic content of certain sediment samples is believed to reflect the activity of bacteria which oxidize organic matter. This belief is substantiated by the relatively abundant ammonia-nitrogen content of such cores (Figure 2), the ammonia having probably been liberated from proteinaceous compounds undergoing bacterial decomposition. Bacteria which thus liberate ammonia have been demonstrated in deep-sea sediments.

The greatest concentration of bacteria in the sea, of course, occurs at the mud-water interface where there may be millions per gram of sedimentary material. We are still trying to learn how rapidly these bacteria reproduce and respire on the deep-sea floor. Our investigations show that if the bacteria reproduce only once a year, such a bacterial population would consume something like 0.6 milligrams of dissolved oxygen per liter. More exact or quantitative information on the rate of reproduction and other physiological activities would contribute appreciably to our understanding of processes and conditions in the sea. So far we have barely scratched the surface in surveying the significance of bacteria as geochemical and biological agents in the deep ocean areas.

Figure 2 - A core of sediment lifted from the floor of the ocean is being slipped out of the Kullenberg piston-type coring device.

On the Naval Research Reserve

Research Pays off for Company 11-4

What to do about training programs? This letter points out a way -

"Concerned because unavoidable attrition had reduced their active membership to such a small number that deactivation was being considered, the members of Research Company 11-4, then of Long Beach, California, were called by the commanding officer to a special business meeting to consider the problem.

"The first question raised was, 'In the light of present international conditions is the Research Reserve essential to our national safety?' An emphatic and a unanimous 'YES' being the answer to this basic question, the next question was, 'What can be done about increasing our membership?'

Guest speaker at training meeting of Company 11-4—Dr. Carl D. Anderson (center) with LCDR Philip B. Wilson, Personnel Officer, and LCDR M. S. Allen, Commanding Officer of Company 11-4.

Company 11-4 hears about smog control in Los Angeles County.

Brig. Gen. Robert C. Oliver talks about U. S. Air Force in Far East.

"There was much discussion about the difficulty of having outstanding speakers appear before seven or eight officers and pretty general agreement that we would have to get more members before we could get better speakers, although there were no suggestions at first as to how this might be done.

"Finally the old Navy spirit that has so often been expressed in times of stress began to come out once again. The proposal was made to offer and to publicize widely a curriculum of such outstanding character that enough visitors from the general public would attend, and at the same time prospective members would be impressed with the worth of the company and join up to become members. The decision was reached to move from San Pedro on the outskirts of the Long Beach district to Compton Naval Training Center, with the expectation of attracting many new members from surrounding communities. Meeting topics were to bear on matters readily identified with current problems of outstanding interest and importance.

"The curriculum adopted covered such topics as transistors, environmental testing techniques for military aircraft, Air-polution, Industrial Nuclear Power, the Application of Reenforced Plastics to Naval Aircraft and Guided Missile Structures, with such speakers as Mr. Smith Griswold and his staff who operate the Air Polution office for Los Angeles County; Mr. A. C. Werden Jr., Director of Nuclear Power for the Southern California Edison Company; and Dr. Carl D. Anderson, Nobel Laureate and Professor of Atomic Physics at the California Institute of Technology.

"Our membership has more than quadrupled and with guests there have been as many as 100 at some of the meetings. Future meetings have outstanding speakers on Flying Saucers, a series on the International Geophysical Year, including the earth satellite program, Naval Technical Intelligence, and many others.

"Scientifically trained officers within driving distance of Compton interested in affiliating with Research Company 11-4, which meets on the second and fourth Tuesday of each month, are invited to communicate with the commanding officer at the Naval Reserve Training Center, 600 No. Alameda Ave., Compton, Calif.

M. S. Allen, LCDR, USNR
Commanding Officer"

TRAINING PROGRAM OF RESERVE RESEARCH COMPANY 11-4

During the months from January through June 1956, Research Company 11-4 had a schedule to include:

CAPT D. P. Tucker from the Pasadena Branch Office of ONR spoke on "Naval Research Reserve, Its Significance and Opportunity."

Mr. Jackie O'Brien of the Pacific Telephone Co. talked on "Transistors - Their Development and Applications."

Mr. Richard J. Gifford of the Rheem Mfg. Co. talked on "Applications of Environmental Techniques to Military Aircraft."

Mr. Smith Griswold and Staff employed by the County of Los Angeles discussed "Smog Control Research in Los Angeles County."

Mr. A. C. Werden, Jr., Southern California Edison Co., talked on "Problems and Progress in Industrial Nuclear Power."

Mr. William Brahm and associates from the Zenith Plastics Co. discussed "Applications of Reinforced Plastics to Naval Aircraft and Guided Missile Structures."

Dr. Carl D. Anderson, Nobel Laureate and Professor of Atomic Physics, California Institute of Technology talked on "Nuclear Particles and the Atom."

Mr. Ben Collins of Douglas Aircraft talked on "The Douglas A-4D."

LCDR M. S. Allen, CO of Company 11-4, talked on "Solar Radiation in Relation to Air-Conditioning Load."

Mr. J. B. Arthur of North American Aviation talked on "The Development and Testing of Fusion Welding in Aircraft."

Brig. Gen. Robert Oliver, USAF Ret., talked on "The U. S. Air Force and the Far Eastern Theater."

LT Robert M. Kripp and several members of Company 11-4 discussed "Individual Research Activities of Members of Company 11-4."

Dr. Joseph Rigney of the University of Southern California talked on "Research Methods Employed in an Analysis of C.I.C. Billets."

August Commanding Officers' Meeting

On 11-12 August 1956 the commanding officers of all the Naval Research Reserve companies will gather at the Naval Air Station, Minneapolis, Minn. for their second annual week-end conference at which they will be informed of the current policies and plans affecting the Naval Reserve in general and the Research Reserve in particular.

Naval Reserve Dental Co. W-1 Hear CAPT Ludwick

The dental research program of the Navy was described in a paper CAPT William E. Ludwick (DC) USN read in a recent meeting of the Naval Reserve Dental Co. W-1 held at the U. S. Naval Dental School, NNMC, Bethesda, Md. CAPT Ludwick is Head of the Medicine and Dentistry Branch of the Office of Naval Research. The presentation included an outline of research projects supported by the Navy in dental schools and other non-profit institutions. In addition, CAPT Ludwick reviewed the studies being conducted by naval dental officers in naval activities. Newer research findings were discussed with particular reference to progress in research on dental caries.

U. S. GOVERNMENT PRINTING OFFICE : 1956 O - 390670

NAVY SPOKESMEN ARE SAYING THIS --

"The Navy approach to research and development has always placed heavy emphasis on basic work. We feel that we have responsibility for helping to maintain a strong science in this country. After World War II, ONR had no hesitation in helping to rebuild university science. This wasn't altruism or charity; it was common sense. It was an investment. And the Navy has been repaid for its foresight many times over. The ships and aircraft and missiles and electronic gear that we have been developing these past few years are to a large extent the children of post-war research."

RADM Rawson Bennett, USN
Chief of Naval Research

"We have accepted the fact that we might be faced with an enemy submarine threat of a magnitude never before equalled in the history of navies. Now, how would we meet this grave threat?

"We have built submarines designed to combat other submarines in their own element. These we call 'killer' subs.

"We have developed long-range aircraft especially designed to control the seas and to deny the submarine the freedom of the surface. These aircraft have means of detecting a snorkeling or surfacing submarine and to attack it.

"We have developed new blimps for the long-range patrol of ocean surfaces.

"We have developed the helicopter into a promising antisubmarine weapon using what we call 'dunking' sonar, a listening device something like a doctor's stethoscope, which can be lowered into the sea at any point.

"We have developed new sound devices for improved surface ships which increase their detection potential.

"We have developed a family of homing weapons, homing torpedoes which follow, track, and destroy maneuvering submarines.

"We have developed new and insidious mines which will be used to block the passage of enemy submarines to sea.

"We have developed and have tested the first atomic submarine. The NAUTILUS has given us vital knowledge and training on the new ASW to be encountered with a nuclear submarine."

Hon. Charles S. Thomas
Secretary of the Navy

In This Issue

COVER PHOTO: Oceanographers aboard Scripps' research ship HORIZON preparing to take deep-mud samples from the bottom of the Pacific Ocean. The long pipes are coring devices. See article beginning page 21.

RESEARCH REVIEWS endeavors to report briefly highlights of technical progress in research by Navy laboratories and contractors and the development of important naval research facilities. Articles which meet these broad criteria and possess a good degree of readability are welcomed. Manuscripts and changes of address should be sent to Code 740, Office of Naval Research, Washington, D. C. . . . Research Reviews is listed for sale by the Superintendent of Documents, U. S. Government Printing Office, Washington, D. C. Subscription price: $1.50 per year in U. S. and Canada; $2.00 per year, foreign; $.15 per individual copy.

Approved by Bureau of the Budget, 28 February 1955