

ABBREVIATED PSYCHOLOGICAL MEASURES I

Ivan N. Mensh, Washington University Medical School, and
William A. Hunt, Northwestern University

INTRODUCTION

The present military emergency once again has introduced pressing problems of manpower mobilization, among them the selection of personnel. World War II first high-lighted the use of abbreviated psychological tests in military selection procedures, although test abbreviation dates back to the pre-World War I period (57) and to Doll's pioneer work (17) 35 years ago. Indeed, the psychological "test," first devised by Galton, was introduced as "an experimental method of measurement . . . characterized by its brevity," but not until World War II was there any large scale development and application (65) of abbreviated psychological measures. In the past decade, one of the authors and his co-workers (40) recognizing the utility of abbreviated techniques for many situations, particularly during periods of rapid mobilization when inadequate numbers of trained personnel are available for screening military recruits, have developed and evaluated many series of brief tests.

Beginning with Doll's report in 1917 and in the more than 20 years prior to 1940, there had been published (65) but 36 articles on abbreviated techniques of psychological measurement. During the recent war years and in the immediate post-war period about 150 articles in this area appeared, and at present there are available nearly double that number of reports. These have increased our fund of information about abbreviated tests, and also have helped to sharpen our focus on the problems arising from their use. The need for brief or abbreviated measures is seen (65, 112) in many situations, civilian and military, ranging from application in the neuropsychiatric screening of military recruits to use in brief-contact clinics and in the screening practices of medical centers where heavy case loads and few personnel demand rapid survey methods for neurological, psychosomatic, and other forms of neuropsychiatric illness. Schools, courts, business and industry, penal and mental institutions, public opinion polls, all have recognized the need for rapid devices of psychological measurement. Doll (17) early pointed out the importance of economy of time in psychological study; Hunt et al (45) have emphasized the economic factors of "mechanics, manpower, and time;" and Bobbitt, Wechsler, and others (65) also have reported the need for abbreviated methods of psychological measurement.

Officers of the Navy also have been long concerned with the need for abbreviated psychological tests, as noted in Louttit's (57) historical review of psychological examining in the Navy. A symposium on intelligence tests reported in the U. S. N. Medical Bulletin of 1915 (47, 83, 85, 94) summarized adaptations of the Binet scale which had been tried out since 1912 by several Navy medical officers. They reported the advantages and limitations of the modified tests, and one of the writers, G. E. Thomas, presented requirements for tests which read like a 1952 study rather than one nearly 40 years old. After a year and a half of the adaptation and trial of the Binet scale at Portsmouth naval prison, Thomas wrote:

There has been much discussion by psychologists outside the Navy and by some of the medical officers in the service, of the value of the Binet system as a means to determine the mentality of the recruit . . . If a mental test is to be applied in the Navy it should be devised for the recruiting officer and it should answer the following requirements: 1. It should be fair in its requirements, and a definite minimum passing mark established. 2. It should be sufficiently varied to make evident the intelligence, education, and training. 3. It should be so devised that but slight, if any, variations are possible in the results of the different examiners. 4. It should not consume much time. (94)

The qualities of "cutting score," range, objectivity, and economy of time are described here, and to them Jenkins, another of the symposium participants, added (47) the requirement: "It can be applied by any intelligent person after a little training."

RATIONALE AND PROBLEMS

In 1946, Hunt and Stevenson (44) summarized important considerations underlying the rationale of abbreviated tests in noting that " . . . changes in efficiency must be evaluated in the light of the

1 This study is part of a larger project subsidized by the Office of Naval Research and conducted at Northwestern University under ONR contract 70NR-45011, NR 154 091. Thanks is due Edna B. Hunt for assistance with the bibliography. The opinions expressed are those of the individual authors and do not represent the opinions or policy of the Naval Service.

demands of each separate screening situation, and an increase in brevity is often worth the slight decrease in test efficiency that it entails. In military testing, the value of any test cannot be determined by fixed and absolute standards. Value is a relative matter determined by the economic factors involved." In the development of new brief methods of psychological examination or in the abbreviation of previous, longer forms, not only must the criteria specific to short forms be satisfied but also those for the usual-length test (65). For example, reliability and validity are just as real problems in abbreviated tests as in longer ones. Moreover, because of fewer items and shorter forms, reliability may decrease as pointed out by Symonds (71), Cattell (10), Wallen (100), and others. However, there are studies such as that of Brokaw (6) in which the reliability of a battery of 6 tests for classifying Air Force personnel for technical training changed only from .95 to .90 when abbreviated by 50 per cent, and the validity of the battery changed only from .57 to .56 after abbreviation.

The general problem of test criteria recently has been reexamined and reported (3) by the American Psychological Association's Committee on Test Standards. Standards of professional judgment in selecting and interpreting tests are presented, stressing the need for "sufficient information about a test so that users will know what reliance can be safely placed on it." Particularly relevant is the Committee's statement that "somewhat different standards should be stressed for different types of tests and not all types of information are equally crucial." This professional group judgment coincides with Hunt and Stevenson's earlier comment on the rationale of abbreviated tests (*vide supra*). Further, the Committee's definition of the scope of the standards presented in their report — "The present standards apply to tests which are distributed for use as a basis for practical judgments rather than solely for research" — applies directly to situations obtaining in military and naval screening procedures where practical judgments must be made continually on the acceptance, acceptance with qualification, or rejection of recruits. In the committee report, standards are given in terms of desired level of information about interpretation of tests (purposes and applications for which the test is recommended, professional qualifications required to administer and interpret the test, data to be taken into account other than test scores), validity (type of validity — predictive, status, content, congruent — and statistical analysis; validation groups comparable to samples for whom test is designed; criterion adequacy), reliability (coefficients of internal consistency, equivalence, and stability), administration and scoring, and scales and norms (percentiles and standard scores, appropriateness of norms, definition of normative samples).

These standards apply equally well for abbreviated tests. Already noted is the problem of reliability based on internal consistency when a parent test is shortened. In a statistical sense, the method of abbreviation operates to lower reliability, but reliability coefficients of equivalence and of stability may be computed as alternate forms of an abbreviated test are developed and applied, either by repeated samplings, by use of the test-retest situation, or in cross-validation studies with samples other than the normative groups on which the test was originally evaluated. The other criteria of test use — interpretation, validity, scales and norms — have the same significance for the abbreviated test as do they for the parent test. In general then, standards of test construction and use are equally the province of parent and of abbreviated psychological measures.

Against this background of test development and use there appear several principal problems in abbreviated testing. Mensh (65) has reviewed these in summarizing studies of the effects of practice, whether termed "experiential factor," "warm-up," transfer or "functional transfer," fatigue, or work decrement; effects of "filler material" or "dead wood" items in long tests; effects of contextual changes; rapport and motivation in abbreviated testing; examiner differences; and order of item difficulty and of administration of the abbreviated test within a battery of tests. Practice effects have long been the concern of psychologists and have been shown to be a function of various factors; "dead wood" items argue for shorter tests but Hunt, Conrad and others have cautioned that a priori guesses about efficiency must be replaced by actual trial of items; contextual, set, and motivational factors have been studied by Conrad (12), Cronbach (16), Horst (36), McCall (26), Mensh (64), and Sears (84) among others; and specific, restricted goals in abbreviated testing have been suggested by Brigham and by Doll (7, 17) early in the history of mental testing, and more recently by Cotzin and Gallagher (14, 15), Hunt and Stevenson (44), Kent (49, 50, 51), Terman and Oden (93), Vernon (98), Wonderlic and Hovland (110), and Zubin (112). Typical of the caution of these investigators is this comment: "... a highly serviceable measure ... Its success with defectives should not be assumed for other clinical groups without further investigation." (15)

The method of specific goals and successive testing proved its efficiency in the military and naval services where thousands of men were screened by brief examinations (25, 65, 112) designed for the specific purpose of discriminating a defined sample (principally two groups — mental deficiency, or personality disturbance serious enough to interfere with adjustment to the armed services) of the population under test.

Most of them were relatively rough. They stood up well in terms of the number of the desired

population identified (mental defectives, the emotionally unstable, etc.) but falsely identified many normal individuals as undesirable. The pick-up or correct identification rates of these tests ranged roughly from 60 to 90 per cent of the population to be identified. The false positive rates, or number of desirables incorrectly identified as undesirable, however, ranged roughly from 5 to 25 per cent of those tested. The final decision as to the use of the test always depended not only upon the efficiency of performance in these terms; but upon such economic factors as time and manpower required, whether or not a better technique were available, and the purpose for which the test were used.

Thus a military screening unit that had neither time nor personnel available for carefully interviewing and examining all the incoming recruits might use as a rough preliminary sieve a test which had a pick-up rate of 85% and a false positive rate of 25% and hold for further examination all men "failing" the test. In such a case, out of 1,000 men to be examined the test might select 300 for further examination. This group of 300 (containing 85% of the unsuitables it was desired to identify) would then be subjected to a psychiatric interview and further testing when desirable in order to separate the undesirables from the false positives. Such a procedure would result in a time and manpower saving of 60 to 70 per cent and still maintain an acceptable screening performance.

The consideration of such economic factors and the acceptance of limited and specific goals for test performance, however, operate against other aspects of test efficiency; and extreme caution must be used in both the clinical and practical inferences drawn from such testing. This caution is present in Doll's earlier insights. After indicating the advantages of abbreviated scales of intelligence, he objectively balanced them against limitations:

... It may be advisable to emphasize some of the limitations of the brief scale as well as its advantages. Equivalence in mental age rating must not be misconstrued as meaning complete psychological or clinical equivalence. Neither may one forget that a mental age rating does not in itself alone furnish a sufficient means of mental diagnosis or determinations of feeble-mindedness. The more complete measuring scales of intelligence furnish a much greater variety of standard situations in which the subject may be caused to display his mental abilities to the trained observer. Moreover, the results of the more extended examination are more satisfactory by reason of the more elaborate consideration of more phases of the subject's intelligence and rule out the possibility of invalidation due to exceptional circumstances of environment or education. (17)

Brigham referred (7) to Binet's explanation of the reason for having a series of test to measure intelligence, rather than a single test. This argues against abbreviated testing but must be considered against the background of the hypothesis suggested by Brigham and successfully put to test by Doll — that an efficient, brief scale could be developed from a longer one by using those items and tests which discriminate against some sample of the population, in Doll's case, the mentally defective. Binet's reasoning is consistent with the reasoning of a number of experimenters with abbreviated testing, e.g., Wonderlic and Hovland, whose brief form of the Otis Self-Administering Test (110) included a number of items distributed uniformly over the range of difficulty. In only one instance have single-item tests been devised (34) and these were for a specific purpose with a defined population sample.

The inherent nature of abbreviated tests places a limitation on the level of reliability of the measures. Among factors affecting reliability is the significant one of number of items. In general, tests are more reliable if the number of items is large (91), and Cattell has framed the question specifically: "Is it possible to cut down a test much below one hour and still get a measure of sufficient consistency (reliability) — not to mention validity — to be used as a basis for decisions affecting the individual's whole career?" (10). Lorge, too, has decried (56) the "tendency to use short tests with out adequate consideration of reliability or of consistency . . ." Yet Doll's study of the Binet-Simon Scale showed that the item intercorrelations were so high that more than half (3 of 5 at each age level) of the tests could be omitted without affecting reliability of the mental ages obtained (17). Also, Lawshe and Mayer (54) found that brief tests of 20, 40, 60, 80, and 100 items could be selected from among 300 items with reliability as high or higher than the long form. And, with respect to validity, in a study of 800 Army inductees and 625 Army prisoners Altus (2) concluded that "the validity of a test is not entirely a function of its length . . . it is possible by careful item selection to reduce a test to as few as 13 questions and still retain a fairly good approximate measure of verbal intelligence . . ." However, Altus cautioned that such approximation should be used only where there is a time premium permitting only one or two minutes.

Other limitations have been recognized by Hunt and Stevenson in their statement of "three common arguments against the use of shorter forms. First, they do not offer the fineness of discriminatory measure that the long tests do. Second, they do not offer the same richness of diagnostic possibilities, i.e., in the analysis of scatter. Third, their use demands more clinical background and skill on the part of the examiner. There is truth in all these arguments, but they are not as conclusive as they

seem" (44). We already have seen that brief tests can be efficient when designed for a specific discriminatory function, and that screening procedures imply that selection is a sifting process (44, 112). Also, recent studies of Hunt and his co-workers (37-43, 52, 63, 73) have demonstrated the clinical diagnostic possibilities of abbreviated tests.

In a comparison of various Wechsler-Bellevue abbreviations, Patterson examined critically the limitations of short tests of intelligence and personality. He recognized the necessity for developing and using brief tests, used and evaluated various measures himself, but pointed out certain dangers and was concerned with "undue emphasis in clinical psychology" on the trend toward shorter tests.

... Indeed, the need is for more good comprehensive tests, rather than shorter forms.

For clinical use, it would appear that an hour or more is not too exorbitant a time for determining the patterning of intellectual functioning, for example. Less than this amount of time decreases the reliability of the sample of the subject's behavior, and reduces the aspects of functioning that can be observed and tested. As a result, a short test not only gives an incomplete picture of the subject's abilities, but often an unreliable picture. Moreover, the limitation of testing to one or a very few functions or aspects of behavior prohibits the comparison of the subject's functioning in various areas . . . (72)

In summary then, economy in both subject and examiner time, in equipment, and in personnel dominates the motivation behind abbreviated measures. Economy in these areas, however, does not permit the criteria of good test construction to be overlooked. Thus the vital problems of reliability and validity are as central to abbreviated techniques as to longer forms of psychological tests. Other problems also must be considered — practice, "warm-up," transfer, "experiential" effects in brief testing as well as in standard-length tests; how can "dead wood" and "filler material" be best localized and eliminated to produce efficient brief measures; the specific goals and functions of abbreviated techniques; the role of examiner differences in the use of short tests; and the significance of set, motivational, and contextual factors which may change as a function of test abbreviation. The limitations of abbreviated testing are reflected in these many factors. As Hunt, Conrad, and others have pointed out, only experimentation can hold the answer to these problems. Some of the answers now are available through recent experimental studies. These show the promise of brief psychological measures which have served a useful function in meeting the need for psychological evaluation.

AVAILABLE ABBREVIATED PSYCHOLOGICAL TECHNIQUES

The experimentation within the recent war and post-war periods has produced a number of abbreviated psychological techniques, some of which are sub-test selections, e.g., vocabulary and other measures (17, 20, 40, 95, 96) from the parent test; others are item selections, as from the Minnesota Multiphasic Personality Test (25, 92); still others are inspection methods as Munroe's technique (68-70) with the Rorschach test; screening devices of which the Saslow symptom index (22, 82) is a sample; and specially devised techniques such as the Kent E-G-Y series (49-51).

Abbreviated psychological measures span the entire range of test materials and methods. There are brief tests for adjustment (1, 2, 76), alcoholic addiction (62), anxiety (25, 32, 92, 105), aphasia (27, 46), controlled association (61), feeling and attitude (35), food aversions (101, 102), memory function (21, 89, 104), mental deficiency (40, 48, 90), myokinetic and autokinetic response (85, 87, 99), neuroticism (19), optimism-pessimism (11), psychiatric prognosis (59), psychosomatic disturbance (67, 106, 107, 109), public opinion (79-81), reaction time (77), time appreciation (9), visual-motor function (4, 5, 6, 23, 24, 55, 58, 111), and vocabulary (13, 95, 96), among others. Samples studied range from childhood through old age, and from "normal" throughout the spectrum of behavior pathology.

Intelligence measures. The extensive use of the Wechsler-Bellevue Intelligence Scale has served as stimulus for use of this test as a parent form from which many abbreviated tests have been selected. Recently there has been a review of research with the W-B Test for the years 1945-50 by Rabin and Guertin (75) in which shorter forms are discussed. Prior to this review are those of Rabin in 1945 (74), and Watson in 1946 (103). In these three reviews nearly 200 studies are summarized, of which about one-fifth are with abbreviated forms. There also have been about 40 studies which report performance of abbreviated forms of the Stanford-Binet Intelligence Scale. Together, these two tests have served as parent forms in nearly 90 studies, more than a third of the reports on test abbreviation published in the psychological literature to date. A third test, the Kent Oral Emergency Test (49-51), different from the W-B and S-B tests in that it was devised as a brief test, has stimulated about 30 studies, serving either as criterion or as experimental test.

A comprehensive study of the clinical usefulness of abbreviated intelligence tests (37-43) has been carried out by the authors and their colleagues. Shortly after the close of World War II, a brief test battery was developed, consisting of the Comprehension and Similarities sub-tests from the Wechs-

ler-Bellevue Intelligence Scale, Form I, and Thorndike's 15-word vocabulary scale (95) taken from the Stanford-Binet vocabulary test. Known as the CVS Individual Intelligence Scale, it was selected for extensive investigation both because of its correlation with external criteria of intelligence and for its diagnostic potentiality. With a sample of 1,649 Naval recruits (40) a correlation of .80 was obtained between CVS and the Navy General Classification Test (GCT). Reliability has been largely inferred on the basis of consistent validity in the testing of separate samples, but a retesting of 116 mental defectives (40) after an interval of one year gave a reliability coefficient of .81. In view of the limited range of intelligence in the sample this can be considered satisfactory. A series of studies of the CVS Scale (40) with large numbers of naval recruits, and with repeated samples of clinical populations has demonstrated the clinical usefulness of this brief battery. The psychological literature now includes CVS data on samples of normal (38, 40, 52), mentally defective (39, 52), brain-damaged (39, 63), and psychotic (39, 52) subjects. The CVS Scale represents a brief verbal scale which can be memorized by the clinician, does not involve test equipment other than a record form of a single page, correlates significantly with external criteria of intelligence, and has potentialities as a rough diagnostic screen for indicating possible psychological disturbance. For situations where non-verbal material is indicated, there are several brief, individual intelligence scales which have been evaluated (38, 39) by Hunt and French. All combine vocabulary with nonverbal materials, correlate significantly (.69-.83) with both GCT and CVS, and have differentiated clinical samples of schizophrenics and mental defectives from normals. The goal of satisfactory diagnostic differentiation demands cross-validation with further samples and a more extended list of clinical disorders.

Personality inventories. The history of the personality inventory as a rapid screening method illustrates how the basic pattern of this technique, laid down 35 years ago, has remained unchanged. Zubin reports a personal communication from Woodworth in which there is the history of the first screening device to be used by the military (112). Woodworth had been appointed by the American Psychological Association in April of 1917 to chair a Committee on Emotional Fitness for Warfare.

Woodworth and Poffenberger worked assiduously on this problem at Columbia and after trying out various tests "hit upon the idea of assembling minor neurotic symptoms, as found by psychiatrists in the case histories of individuals who later developed neuroses or psychoses, and tallying up the score of positive answers . . . intended as a screening device with primary use of the quantitative score, but also with attention to certain 'starred questions' which the psychiatrists . . . believed would be of significance quite apart from the total score."

A comprehensive review by Ellis and Conrad (18) of the military applications of personality inventories, many of them brief methods, has yielded a number of conclusions about factors responsible for the favorable results in military practice and the disappointing findings in civilian practice. After examining studies of military personnel by inventories making use of a psychiatric criterion (prognosis or diagnosis of neuropsychiatric unfitness for military duty), the authors concluded that certain factors appear to have played a part in the results obtained. These factors were criterion contamination and overlap, use of extreme or atypical groups, differential motivation, inadequate statistical treatment of data, lenient evaluation of "false-positive" results and neglect of "false-negative" cases, sample heterogeneity, lower intelligence or greater naivete of military subjects with "less distortion" of responses than among civilians, specialized redign and validation, and application "for screening only, and not for elaborate personality analysis."

In studies making use of a performance criterion, as success in a training course, prediction was much less effective than with a psychiatric criterion. Ellis and Conrad attribute the difference to prior elimination of abnormals in selection for training courses, lack of reliability or validity of the performance measures, differences in aptitude and previous training rather than differences in emotional adjustment, and shift of criterion from validation in terms of the psychiatric criterion in the original standardization, to validation in terms of performance measures. They state:

1. Personality questionnaires should be especially designed for the group to whom they are applied, and should be validated against dependable external criteria. Criterion-contamination should be guarded against; and criterion-overlap, if it occurs, should be taken into account in evaluating the findings.

2. Special attention should be given to persuading or inducing respondents to answer the inventory items as truthfully as they can.

3. Personality inventories may possibly be more effective when used with relatively uneducated and less intelligent groups, than with groups that are more sophisticated.

4. The users of personality inventories should realize that only limited and specialized demands may be made on the inventory technique; and that broad and incisive personality diagnosis is still the specialty of the trained clinician employing subtler and more compre-

hensive psychological techniques. (18)

At the 1947 Maryland Conference on Military Psychology, Wexler summarized the Navy's experience with psychiatric screening tests in the following terms:

Perhaps the best way to summarize our experience would be to suggest, as a potentially valuable research instrument, an inventory with several broadly diagnostic scales, having items in paired-choice form, with a "self-idealization" scale as a possible corrective for scores on the units directed toward the measurement of maladjustment, and finally, with a separate inquiry, perhaps biographical or attitudinal, into defensive or integrating elements which might serve to counterbalance and negate the total picture of disturbance. (108)

Projective techniques. During the past 15 years there has been great emphasis upon projective methods, as distinguished from structured intelligence and other diagnostic devices, and personality inventories whose data are treated in traditional psychometric fashion. It would be expected therefore, that demands for abbreviated psychological measures also would include the projective techniques. As with the other two principal types of psychological methods there are now available in the projective field both abbreviated forms of parent tests and brief tests specifically designed for rapid evaluation.

The most widely used method, the Rorschach technique, has been modified by group administration, multiple-choice selection of responses, decrease in the standard number of stimulus cards presented, and rapid methods of scoring and interpretation. During World War II, Hertz (33) streamlined the 8-10 hours of administration, scoring, and interpretation with the standard methods, to 50 minutes of administration time using summary sheets and check lists to speed interpretation. However, problems of reliability and validity of the briefer methods are as major (66) as they are in use of the standard techniques. These difficulties also are found in the Harrower-Erickson screening modification (29-31) in which multiple-choice responses are introduced rather than free association, thus sharply limiting the range of response. Zuckerman (113) suggested further modification for large-scale Rorschach testing, with three exposures for each of the ten stimulus slides — 20, 15, and 15-second exposures, respectively — and ten multiple-choice items per exposure to be responded to on IBM answer sheets and scored by stencil. Munroe (70) has reported an experiment with group administration, three minutes per card, and scoring and tabulation in a 20-minute period by means of her Inspection Rorschach Check List. This latter device (68-70) represents still another avenue for the abbreviation of tests, with concentration on shortening significantly the time required for scoring and tabulating Rorschach data. Munroe (71) supports the use of projective methods in group testing in her comment that "the projective method offers a complex specimen of spontaneous action even when administered to groups . . . where current individual methods are adapted to group use, the group tester for the first time can approach the problem of evaluation with something of the resourcefulness and knowledge available to the clinician working with similar individual methods."

Another principal projective technique, often used in conjunction with the Rorschach method, is the Thematic Apperception Test and this also has been modified in both administration and scoring in order to reduce the time factor. The use of slides and a reduced number of stimulus cards have been experimented with as methods of economy in administration. Harrison and Rotter (28) used 5 slides in 30-second exposures with 7½ minutes allowed for each response period; and Smith, Brown, and Throver (86) used 8 cards of the TAT series as an aid in history-taking, diagnosis, and treatment situations in the neuropsychiatric clinic of a general hospital.

In addition to the Rorschach and Thematic Apperception Test abbreviations there are a number of other projective techniques which require relatively brief periods of administration. These include among others Mira's myokinetic psychodiagnostics (87), Bender's visual motor gestalt test (4), the Geosign Test (76), the graphomotor projective technique (53), van Lennep's Four-Picture Test (97), Machover's Draw A Person Test (60), and word association and sentence completion (78) techniques.

SUMMARY

The history of abbreviated psychological measurement extends back over the past 40 years, beginning with the efforts of medical officers of the U.S. Navy to adapt the Binet Scale for measuring intelligence to selection of recruits. Criteria for such brief techniques were formulated at that time which still hold for present-day testing, covering the requirements of "cutting" scores, adequate range, objectivity, economy of time, and simplicity of administration and scoring. These pioneers in brief psychological measurement also were aware of the limitations of the methods. Continuing concern arising from experimental evidence has indicated caution in their use.

World War II gave the major impetus to abbreviated tests and the present emergency and manpower mobilization problems again have stimulated interest in the development and validation of

rapid, objective methods for neuropsychiatric screening. There now are available in the psychological and psychiatric literature about 300 reports on abbreviated or brief psychological tests. These cover the range of intelligence and other diagnostic measures, personality inventories, and projective techniques; and sample populations of normal, neurotic, psychotic, and brain-damaged individuals. These many studies have attempted to meet the demands for brief psychological methods by the military and naval services, hospitals, clinics, schools, and business and industry.

Advantages of abbreviated measures lie in their economy of time both in subject and examiner time, in elimination of "deadwood" and "filler" items, in equipment, and in trained personnel. These have been demonstrated in studies of verbal and nonverbal test materials where their diagnostic usefulness has been proven. The limitations of brief measures must be examined in terms of their specific goals, and the significance of set, motivational, and contextual factors which may change as a function of test abbreviation. In conclusion we may repeat our previous quotation: "... changes in efficiency must be evaluated in the light of the demands of each separate screening situation, and an increase in brevity is often worth the slight decrease in test efficiency that it entails. In military testing, the value of any test cannot be determined by fixed and absolute standards. Value is a relative matter determined by the economic factors involved." (40)

1. Aitkus, W. D. The adjustment of Army illiterates. *Psychol. Bull.*, 1945, 42, 461-476.
2. Aitkus, W. D. The validity of an abbreviated information test used in the Army. *J. consult. Psychol.*, 1948, 12, 270-275.
3. American Psychological Association Committee on Test Standards. Technical recommendations for psychological tests and diagnostic techniques: Preliminary proposal. *Amer. Psychol.*, 1952, 7, 461-475.
4. Bender, L. A Visual Motor Gestalt Test and its clinical use. New York: Amer. Orthopsychiat. Assoc., 1938.
5. Benton, A. L. A visual retention test for clinical use. *Arch. Neurol. Psychiat.*, 1945, 54, 212-216.
6. Benton, A. L. A multiple choice type of the visual retention test. *ibid.*, 1950, 64, 699-707.
7. Brigham, C. C. Two studies in mental tests. *Psychol. Monogr.*, 1917, 24, No. 1.
8. Brokaw, L. D. Comparative validities of "short" versus "long" tests. *J. appl. Psychol.*, 1951, 35, 325-330.
9. Buck, J. N. The Time Appreciation Test. *J. appl. Psychol.*, 1946, 388-398.
10. Cattell, R. B. The measurement of adult intelligence. *Psychol. Bull.*, 1943, 40, 153-193.
11. Charot, S. N. F., & Myers, C. R. An approach to the measurement of mental health. *Amer. J. Orthopsychiat.*, 1936, 6, 134-140.
12. Conrad, H. S. Characteristics and uses of item-analysis data. OGRD Report No. 4034, 1944.
13. Corsini, R. J. The immediate test. *J. clin. Psychol.*, 1951, 7, 127-130.
14. Cotzin, M., & Gallagher, J. J. Validity of short forms of the Wechsler-Bellevue Scale for mental defectives. *J. consult. Psychol.*, 1949, 13, 357-365.
15. Cotzin, M., & Gallagher, J. J. The Southbury Scale: A valid abbreviated Wechsler-Bellevue for mental defectives. *ibid.*, 1950, 14, 358-364.
16. Cronbach, L. J. Response sets and test validity. *Educ. psychol. Measmt.*, 1946, 6, 475-494.
17. Dell, E. A. A brief Binet-Simon scale. *Psychol. Clin.*, 1917, 11, 197-211, 254-261.
18. Ellis, A., & Conrad, H. S. The validity of personality inventories in military practice. *Psychol. Bull.*, 1948, 45, 365-426.
19. Eysenck, H. J. A comparative study of four screening tests for neurotics. *Psychol. Bull.*, 1945, 42, 659-662.
20. Finkelstein, M., Gerboth, R., & Westerhold, R. Standardization of a short form of the Wechsler Vocabulary subtest. *J. clin. Psychol.*, 1952, 8, 133-135.
21. Freeman, E., & Josey, W. E. Quantitative visual index to memory impairment. *Arch. Neurol. Psychiat.*, 1949, 62, 794-797.
22. Glaser, G., & Ulett, G. The Saslow screening test as a measure of anxiety-proneness. *J. clin. Psychol.*, 1952, 8, 279-283.
23. Graham, F. K., & Kendall, B. S. Performance of brain-damaged cases on a Memory-for-Designs Test. *J. abnorm. soc. Psychol.*, 1946, 41, 303-314.
24. Graham, F. K., & Kendall, B. S. Further standardization of the Memory-for-Designs Test on children and adults. *J. consult. Psychol.*, 1948, 12, 349-354.
25. Grant, H. A rapid personality evaluation: Based on the MMP! and Cornell Selectee Index. *Amer. J. Psychiat.*, 1946, 103, 33-41.
26. Guilford, J. P. *Psychometric methods*. New York: McGraw-Hill, 1936.
27. Halstead, W. C., & Wepman, J. M. The Halstead-Wepman aphasia screening test. *J. Sp. Hear. Dis.*, 1949, 14, 9-15.

28. Harrison, R., & Rotter, J. A note on the reliability of the Thematic Apperception Test. *J. abnorm. soc. Psychol.*, 1945, 40, 97-99.
29. Harrower-Erickson, M. R. A multiple choice test for screening purposes. *Psychosomat. Med.*, 1943, 5, 331-341.
30. Harrower-Erickson, M. R. Modification of the Rorschach method for large scale investigations. *Assoc. Res. Nerv. Ment. Dis.*, 1946, 25, 340-344.
31. Harrower-Erickson, M. R., & Steiner, M. Modification of the Rorschach method for use as a group test. *J. genet. Psychol.*, 1943, 62, 119-133.
32. Hartogs, R. The clinical investigation and differential measurement of anxiety. *Amer. J. Psychiat.*, 1950, 106, 929-934.
33. Hertz, M. R. Modification of the Rorschach Ink Blot Test for large scale application. *Amer. J. Orthopsychiat.*, 1943, 13, 191-211.
34. Hildreth, H. M. Single-item tests for psychometric screening. *J. appl. Psychol.*, 1945, 29, 262-267.
35. Hildreth, H. M. A battery of feeling and attitude scales for clinical use. *J. clin. Psychol.*, 1946, 2, 214-220.
36. Horst, P. Item analysis by the method of successive residuals. *J. exp. Educ.*, 1934, 2, 254-263.
37. Hunt, W. A., & French, E. G. A second fifteen-word vocabulary test for use with abbreviated intelligence scales. *J. consult. Psychol.*, 1949, 13, 124-126.
38. Hunt, W. A., & French, E. G. Some abbreviated individual intelligence scales containing non-verbal items. *J. consult. Psychol.*, 1949, 13, 119-123.
39. Hunt, W. A., & French, E. G. The Navy-Northwestern Matrices Test. *J. clin. Psychol.*, 1952, 8, 65-74.
40. Hunt, W. A., & French, E. G. The CVS abbreviated individual intelligence scale. *J. consult. Psychol.*, 1952, 16, 181-186.
41. Hunt, W. A., French, E. G., Klebanoff, S. G., Mensh, I. N., & Williams, M. The validity of some abbreviated individual intelligence scales. *J. consult. Psychol.*, 1948, 12, 48-52.
42. Hunt, W. A., French, E. G., Klebanoff, S. G., Mensh, I. N., & Williams, M. The clinical possibilities of an abbreviated individual intelligence test. *ibid.*, 171-173.
43. Hunt, W. A., French, E. G., Klebanoff, S. G., Mensh, I. N., & Williams, M. Further standardization of the CVS individual intelligence scale. *ibid.*, 333-359.
44. Hunt, W. A., & Stevenson, I. Psychological testing in military clinical psychology: I. intelligence testing. II. Personality testing. *Psychol. Rev.*, 1946, 53, 25-35, 107-115.
45. Hunt, W. A., Wittson, C. L., & Harris, H. I. The screen test in military selection. *Psychol. Rev.*, 1944, 51, 37-46.
46. Jacobson, J. R. A method of psychobiologic evaluation. *Amer. J. Psychiat.*, 1944, 101, 343-348.
47. Jenkins, H. E. Mental defectives at Naval Disciplinary Barracks, Port Royal, S.C. *U.S. Nav. Med. Bull.*, Wash., 1915, 9, 211-221.
48. Keller, M., Child, I. L., & Redlich, F. C. Preliminary test of intelligence. A brief test of adult intelligence designed for psychiatric examiners. *Amer. J. Psychiat.*, 1947, 103, 785-792.
49. Kent, G. H. Oral test for emergency use in clinics. *Ment. meas. Monogr.*, 1932, No. 9.
50. Kent, G. H. Emergency battery of one-minute tests. *J. Psychol.*, 1942, 13, 141-164.
51. Kent, G. H. Series of emergency scales. New York: Psychol. Corp., 1946.
52. Klehr, H. Clinical intuition and test scores as a basis for diagnosis. *J. consult. Psychol.*, 1949, 13, 34-38.
53. Kutash, S. B. Recent developments in the field of projective techniques. *Ror. Res. Exch.*, 1949, 13, 74-86.
54. Lawshe, C. H., Jr., and Mayer, J. S. Studies in item analysis: I. The effect of two methods of item validation on test reliability. *J. appl. Psychol.*, 1947, 31, 271-277.
55. Lord, E., & Wood, L. Diagnostic values in a visuo-motor test. *Amer. J. Orthopsychiat.*, 1942, 12, 414-428.
56. Lorge, I. Intellectual changes during maturity and old age. *Rev. educ. Res.*, 1944, 14, 438-445.
57. Louttit, C. M. Psychological examining in the United States Navy: An historical summary. *Psychol. Bull.*, 1942, 39, 227-239.
58. Louttit, C. M. The Mirror Tracing Test as a diagnostic aid for emotional instability. *Psychol. Rec.*, 1943, 5, 279-286.
59. Low, A. A., Singer, H. D., & McCorry, C. L. Life situation tests as aids in psychiatric prognosis. *Amer. J. Psychiat.*, 1939, 96, 147-164.
60. Machover, K. Personality projection in the drawing of the human figure. Springfield, Ill.: Chas. C. Thomas, 1948.

61. Malamud, D. I. Value of the Maller Controlled Association Test as a screening device. *J. Psychol.*, 1946, 21, 37-43.
62. Manson, M. P. A psychometric determination of alcoholic addiction. *Amer. J. Psychiat.*, 1949, 106, 199-205. Also, *J. clin. Psychol.*, 1949, 5, 77-93.
63. Matarazzo, J. D. A study of the diagnostic possibilities of the CVS with a group of organic cases. *J. clin. Psychol.*, 1950, 6, 337-343.
64. Mensh, I. N. The effects of test abbreviation upon responses to the individual items. *J. clin. Psychol.*, 1949, 5, 22-36.
65. Mensh, I. N. Brief psychological measures. *Nerv. Child*, 1949, 2, 349-357.
66. Mensh, I. N. Statistical techniques in present-day psychodiagnostics. *Psychol. Bull.*, 1950, 47, 475-492.
67. Moore, R. C. Psychiatric screening tests at a precommissioning center. *U.S. Nav. med. Bull.*, 1947, 47, 676-682.
68. Munroe, R. L. Inspection technique: A modification of the Rorschach method of personality diagnosis for large scale application. *Ror. Res. Exch.*, 1941, 5, 166-191.
69. Munroe, R. L. The inspection technique: A method of rapid evaluation of the Rorschach protocol. *Ror. Res. Exch.*, 1944, 8, 46-70.
70. Munroe, R. L. An experiment with a self-administering form of the Rorschach and group administration by examiners without Rorschach training. *Ror. Res. Exch.*, 1946, 10, 49-59.
71. Munroe, R. L. The use of projective methods in group testing. *J. consult. Psychol.*, 1948, 12, 8-15.
72. Patterson, C. H. A comparison of various "short forms" of the Wechsler-Bellevue Scale. *J. consult. Psychol.*, 1946, 10, 260-267.
73. Pollaczek, P. P. A study of malingering on the CVS abbreviated individual intelligence scale. *J. clin. Psychol.*, 1952, 8, 75-81.
74. Rabin, A. I. The use of the Wechsler-Bellevue Scales with normal and abnormal persons. *Psychol. Bull.*, 1945, 42, 410-422.
75. Rabin, A. I., & Guertin, W. H. Research with the Wechsler-Bellevue Test: 1945-1950. *Psychol. Bull.*, 1951, 48, 211-248.
76. Reichenberg-Hackett, W. The Geosign Test: A semistructured drawing situation utilized as a screening test for adjustment. *Amer. J. Orthopsychiat.*, 1950, 20, 578-594.
77. Rodnick, E. H., & Shakow, D. Set in the schizophrenic as measured by a composite reaction time index. *Amer. J. Psychiat.*, 1940, 97, 214-225.
78. Rotter, J. B. Word association and sentence completion methods. In Anderson, H. H., & Anderson, G. L. *An introduction to projective techniques*. New York: Prentice-Hall, 1951.
79. Sanford, F. H. The use of a projective device in attitude surveying. *Publ. Opin. Quart.*, 1950, 14, 697-709.
80. Sanford, F. H., & Older, H. J. A short Authoritarian-Equalitarian Scale. *Phila.: Inst. Res. Human Relations*, June, 1950.
81. Sanford, F. T., & Rosenstock, I. M. Projective techniques on the doorstep. *J. abnorm. soc. Psychol.*, 1952, 47, 3-16.
82. Saslow, G., Counts, R. M., & DuBois, P. H. Evaluation of a new psychiatric screening test. *Psychosomat. Med.*, 1951, 13, 242-253.
83. Schier, A. R. Review and possibilities of mental tests in the examination of applicants for enlistment. *U. S. Nav. med. Bull.*, Wash., 1915, 9, 222-226.
84. Sears, R. Motivational factors in aptitude testing. *Amer. J. Orthopsychiat.*, 1943, 13, 468-493.
85. Sexton, M. C. The autokinetic test: Its value in psychiatric diagnosis and prognosis. *Amer. J. Psychiat.*, 1945, 102, 399-402.
86. Sheehan, R. Service use of intelligence tests. *U. S. Nav. med. Bull.*, Wash., 1915, 9, 194-200.
87. Simon, J. L. The myokinetic psychodiagnostics of Dr. Emilio Mira. *Amer. J. Psychiat.*, 1943, 100, 334-341.
88. Smith, J. A., Brown, W. T., & Thrower, F. L. The use of a modified TAT in a neuropsychiatric clinic in a general hospital. *Amer. J. Psychiat.*, 1951, 107, 498-500.
89. Stone, C. P., Girdner, J., & Albrecht, R. An alternate form of the Wechsler Memory Scale. *J. Psychol.*, 1946, 22, 199-206.
90. Strauss, A. A., & Kephart, N. C. Behavior differences in mentally retarded children measured by a new behavior rating scale. *Amer. J. Psychiat.*, 1940, 96, 1117-1124.
91. Symonds, P. M. Factors influencing test reliability. *J. educ. Psychol.*, 1928, 19, 73-97.
92. Taylor, J. A. The relationship of anxiety to the conditioned eyelid response. *J. exp. Psychol.*, 1951, 41, 91-92.
93. Terman, L. M., & Oden, M. H. *The gifted child grows up*. Stanford: Stanford Univ. Press, 1947.

94. Thomas, G. E. The value of the mental test and its value to the service. *U. S. Nav. med. Bull.*, Wash., 1915, 9, 200-211.
95. Thorndike, R. L. Two screening tests of verbal intelligence. *J. appl. Psychol.*, 1942, 26, 128-135.
96. Thorndike, R. L., & Gallup, G. H. Verbal intelligence of the American adult. *J. gen. Psychol.*, 1944, 30, 75-85.
97. Van Lennep, D. J. The Four-Pictures Test. In Anderson, H. H., & Anderson, G. L. *An introduction to projective techniques*. New York: Prentice-Hall, 1951.
98. Vernon, P. E. Short tests of low-grade intelligence. *Occup. Psychol.*, Lond., 1941, 15, 107-111.
99. Voth, A. C. An experimental study of mental patients through the autokinetic phenomenon. *Amer. J. Psychiat.*, 1947, 103, 793-805.
100. Wallen, R. Some testing needs in military clinical psychology. *Psychol. Bull.*, 1944, 41, 539-542.
101. Wallen, R. Food aversions of normal and neurotic males. *J. abnorm. soc. Psychol.*, 1945, 40, 77-81.
102. Wallen, R. Food aversions in behavior disorders. *J. consult. Psychol.*, 1948, 12, 310-312.
103. Watson, R. I. The use of the Wechsler-Bellevue Scales: A supplement. *Psychol. Bull.*, 1946, 43, 51-68.
104. Wechsler, D. A standardized memory scale for clinical use. *J. Psychol.*, 1945, 19, 87-95.
105. Wechsler, D., & Hartogs, R. The clinical measurement of anxiety. *Psychiat. Quart.*, 1945, 19, 619.
106. Weider, A., Brodman, K., Mittelman, B., Wechsler, D., & Wolff, H. G. The Cornell Index: A method for quickly assaying personality and psychomatic disturbances, to be used as an adjunct to interview. *Psychosomat. Med.*, 1946, 8, 411-413.
107. Weider, A., Mittelman, B., Wechsler, D., & Wolff, H. G. The Cornell Service Index: A method for quick testing of selectees for the armed forces. *J. A. M. A.*, 1944, 124, 224-228.
108. Wexler, M., Owens, W. A., & Porter, R. B. Psychiatric screening of naval personnel. In Kelly, G. A. (ed.) *New methods in applied psychology*. College Park: Univ. Maryland, 1947. p. 60-66.
109. Wolff, H. G., Weider, A., Mittelman, B., & Wechsler, D. The Selective Index: A method for quick testing of selectees for the armed forces. *Trans. Amer. Neurol. Assoc.*, 1943, 43, 126-129.
110. Wonderlic, E. F., & Hovland, C. I. The personnel test: A restandardized abridgment of the Otis S-A Test for business and industrial use. *J. appl. Psychol.*, 1939, 23, 685-702.
111. Wood, L., & Shulman, E. The Ellis Visual Designs Test. *J. educ. Psychol.*, 1940, 31, 591-602.
112. Zubin, J. Recent advances in screening the emotionally maladjusted. *J. clin. Psychol.*, 1948, 4, 56-62.
113. Zuckerman, S. B. A research suggestion in large-scale Rorschach. *J. consult. Psychol.*, 1948, 12, 300-302.