DOCUMENT RESUME

ED 174 669 TM 009 574

| | |
|---|---|
| TITLE | Proceedings of the Invitational Conference on Testing Problems. (20th, New York, New York, November 3, 1956). |
| INSTITUTION | Educational Testing Service, Princeton, N.J. |
| PUB DATE | 3 Nov 56 |
| NOTE | 118p. |
| EDRS PRICE | MF01/PC05 Plus Postage. |
| DESCRIPTORS | *Achievement Tests; Age Differences; Conceptual Schemes; *Curriculum Development; Educational Testing; *Educational Trends; Higher Education; Individual Characteristics; Intellectual Development; Intelligence; Intelligence Factors; *Intelligence Tests; Nontraditional Students; Personality Development; *Personality Tests; Personality Theories; Student Testing; Test Construction; *Testing Problems |

ABSTRACT

The conference theme was Testing--Then and Now, based upon the assumption that psychometrists should review their origins and look forward to their future. In the area of intelligence testing, Nancy Bayley discussed the research that re-assessed previous concepts of the nature of intellectual growth and decline. Thelma Thurstone discussed new evidence regarding the differences in the rates of achieving intellectual primes for the primary mental abilities. Gardner Murphy presented a new orientation to the field of personality appraisal, and Morris Krugman discussed the relationship between the newer concepts of personality and current methods of appraisal. The luncheon address was presented by Ben D. Wood. As for achievement testing, Ralph Taylor reviewed the approaches to curriculum then and now; John Dobbins reported on recent measurement of achievement in today's schools; and Louis Hacker discussed the need for appraising the ability and achievement of adults who are returning to college after normal working hours. (Author/BH)

# 1956



EDUCATIONAL TESTING SERVICE

Library of Congress Catalog Number: 47-11220

# INVITATIONAL CONFERENCE
# ON
# TESTING PROBLEMS

NOVEMBER 3, 1956

IRVING LORGE, *Chairman*

Changing Concepts of Intelligence

Changing Concepts of Personality

The Changing Curriculum and Testing

EDUCATIONAL TESTING SERVICE
PRINCETON, NEW JERSEY LOS ANGELES, CALIFORNIA

# FOREWORD

Most appropriately for a twentieth anniversary, the 1956 Invitational Conference on Testing Problems had as its theme "Testing—Then and Now." The speeches and discussions that follow are eloquent testimony to the great progress that has been made over the past few decades in such important areas as intelligence, achievement, and personality appraisal. One can hopefully dream that 20 years from now we will be able to look back to as great a measure of accomplishment as reported at this meeting.

For the success of the 1956 Conference we are sincerely indebted to Chairman Irving Lorge. His imagination and energy, coupled with his sense of humor, did much to make the Conference instructive and enjoyable. To him and to the speakers I would like to say, sincerely, "Thank you."

HENRY CHAUNCEY
*President*

# PREFACE

The 1956 Invitational Conference on Testing Problems, sponsored by Educational Testing Service, was held at the Hotel Roosevelt in New York City on November 3, 1956. This, the twentieth Conference of the series, was attended by more than 550 educators, psychologists, and personnel workers concerned with measurement and evaluation.

The Conference was organized around the theme "Testing—Then and Now" on the assumption that psychometry could afford a retrospective look to its origins and a forward look to its future.

The 1890's were the years of beginning: Binet and Simon's work on the measurement of intelligence, and Joseph Mayer Rice's rejection of the concept that the elementary curriculum must be accepted as an invariant "given." Indeed, for the United States, the 1890's saw the start of the expansion of educational opportunity for all the children with its inevitable consequent that the educational expectations for children would depend on the calibre of the child.

Intelligence testing in the United States was accelerated not only by Terman's Stanford Revision of the Binet Simon Intelligence Scale but also by the use of Army Examination Alpha during World War I and its subsequent utilization in civilian testing. Indeed, Jones and Conrad's generalizations about the curve of intelligence was based on the use of Army Alpha with adult civilians.

The testing movement expanded rapidly in the 1920's. The objectives of the curriculum were challenged by the psychometrists and their test results, as exemplified by Learned and Wood's reconsideration of the meaning of knowledge as well as the testing of the acquisition of general education whether obtained formally in school or informally from the candidate's experiences and transactions with his physical and social environment. Indeed, the Pennsylvania study indicated the need for a Cooperative Test Service which in due course became one of the roots of Educational Testing Service. The Conference was fortunate in having Ben D. Wood review the origins of the testing movement in light of his personal participation with it.

In the area of intelligence testing, Nancy Bayley integrated the research that reassessed the earlier concepts of the nature of intellectual growth and decline; Thelma Thurstone, too, brought new evidence about the difference in the rates of the achievement of intellectual primes for the primary mental abilities.

For personality appraisal, Gardner Murphy gave a fresh orientation to the field, and Morris Krugman showed the relation between the newer concepts of personality and current methods of appraisal.

In achievement testing, Ralph Tyler reviewed the approaches to curriculum then and now; John Dobbin reported on the newest measurements of achievement in today's schools, and Louis Hacker showed the need for appraising the ability and achievement of adults who are profiting from the opportunities to obtain a college education after work hours and after a hiatus in their educational progress.

I wish to express my appreciation to the speakers for their contributions, to the stimulating extension of ideas by the many able discussants, and to Educational Testing Service for its educational leadership in the sponsorship of Educational Conferences, then and now and in the years to come.

IRVING LORGE
*Chairman*

# CONTENTS

# GENERAL MEETING

## Changing Concepts of Intelligence



9

# A New Look at the Curve of Intelligence

NANCY BAYLEY

In the past 30 years the accepted form of the age curve of intelligence has become pretty well stabilized, as the result of a number of studies in which many people of different ages were tested. Miles (25) set the pattern which follows the course of increasing scores to the early 20's followed by a consistent but slower decline throughout the adult years. Subsequent studies have followed this pattern with minor variations. Notable among them are the studies of Jones and Conrad (19), of Wechsler (41), and of Foulds and Raven (11).

In recent years repeated tests on the same persons as they grow older have yielded scores that do not follow this pattern, but indicate that at least some intellectual abilities may continue to increase slowly to 50 years of age or older. These, among other findings on early development, such as the instability of infant scores, are forcing us to reconsider the whole subject of age changes in intellectual abilities. As a part of this reconsideration we need to review the methods by which intelligence can best be tested and evaluated, at different ages and levels of complexity. We must investigate the limitations of current tests and develop more adequate and discriminating ones.

### *The Complex Nature of Intellectual Abilities*

It is, of course, understood that the "curve of intelligence" is derived from performances on tests that measure samples of the whole range and variety of mental abilities. This means that all of the following discussion presupposes that the tests must be validated and re-validated, as occasion requires and permits, against useful outside criteria of intelligent behavior. That is, if we wish to know the nature of intellectual change over time, we must insure that we are using a valid measure of intelligence. It may be necessary, also, for us to re-evaluate our criteria of intelligence.

In formulating any adequate theory of the nature of intelligence it seems to me necessary to take into consideration the fact that the human organism, in all of its aspects, undergoes continual processes of change throughout its life span. These changes are more rapid at some periods of life than at others, and in some processes than in others. They

involve increments and decrements in size and waxing and waning of functions. They also involve developmental increases in complexity, and processes of maturation of both structure and function, followed eventually by retrogressions, declines and other manifestations of senescence.

In the field of intelligence these processes are complex and difficult to work with, and hence have been only partially mapped out. For one thing, the concept of intelligence is very general, and covers a variety of intellectual functionings. It is necessary, then, for an adequate appraisal of the course of intelligence from birth to old age, that we consider not only the course of general intelligence, but also age changes in mental organization, and that we try to identify and independently to measure the various intellectual functions. When this has been done, it will become possible to trace the developmental changes in each function or factor, and to see how each fits in to any general, over-all curve.

Of course, many investigators have been working on the problem of analyzing intelligence into its component parts. Early in the history of intelligence testing efforts were made both to define intelligence and to construct tests that would sample and score separately its different aspects. At first the various intellectual faculties to be tested were selected on an *a priori* basis. But with experience, and the help of statistics, we have found ways of isolating and measuring relatively discrete intellectual functions.

Considerable work has been done on what seems to me the logical approach to building new test batteries, i.e., through the use of such methods as factor analysis, cluster analysis, or analysis of variance. But very few investigators have actually applied the results of such analyses to construct factorially independent scales. A notable exception, of course, is the series of Thurstone tests of Primary Mental Abilities (39). Following from such a start as this, by successive constructions, analyses, and additions of test items and test areas, we can hope to tease apart and thus to identify, measure, and label the different components of intelligence. Something very like this process has been reported recently by Guilford (14) who presents an elaborate scheme for what he calls "The Structure of Intellect." He reports systematic attempts to detect and to organize into a conceptual frame the various factors of intellect. With this schema as a basis, new types of function are hypothesized or identified, and appropriate tests are devised to measure them. It seems to me that Guilford reports the kind of research we badly need if we are to differentiate and understand the various intellectual processes that make up "intelligence." But in his elaborate

analysis he does not take into consideration the further complications of age and maturational differences.

However, a number of investigators are interested in this developmental aspect of intelligence, and with the instruments available several studies have been made on age changes in intellectual ability and in mental organization (13). Much of this work has been done on tests of cross-sectional samples: that is, on tests administered once to persons of different ages. But there are very pertinent data now available from a number of longitudinal studies in which the same persons have been tested repeatedly as they grew older. None of these longitudinal studies covers the entire age span, but some start as young as one month of age and others continue through 50 years and even older. By splicing together some of these data it should be possible to construct longitudinal age curves of several different intellectual factors.

### *A Suggested Age-Curve of Intelligence*

Any curve of intelligence will be dependent upon the behavioral components included in the test scores on which it is constructed. The general curve which I presented last year in an article "On the Growth of Intelligence" (6), is based on total scores of fairly comprehensive tests. To the extent that these tests are broadly inclusive and representative of "general intelligence," this curve might be thought of as representing the usual course of age changes in general intellectual capacity.

Everyone agrees that intelligence grows throughout infancy and childhood, though with the accumulation of records we have had to move along (from a start at 13 years), the age at which growth was assumed to stop. As for the *form* of the childhood curve, minor differences from one investigator to another are probably related to differences in tests and sampling, but there is the further problem of constructing comparable units of growth. In the tentative curve I constructed, for the period one month to 25 years, I used the test scores of the Berkeley Growth Study. It was necessary to splice together the scores from a number of different tests by converting the scores into standard deviation units based on the mean and S.D. of the 16-year scale scores on the Weehsler-Bellevue Test.

This part of the curve shows rapid acceleration in the first year and again a moderate acceleration between 8 and 10 years. After about 10 years the rate slows down so that by 25 the increment is very gradual. Freeman and Flory's (12) curve, although based on units derived in a different way, is similar to the Berkeley Growth Study curve for the ages it covers, 8 to 17 years. According to the V.A.C.O. curve, there is

accelerated growth between about 10 and 12 years. This is later than the Berkeley Growth Study period of rapid growth. However, different tests were used, and when the subtests of the V.A.C.O. are considered, each has a different course of growth with different ages of greatest increment.*

By fitting on to the Berkeley Growth Study curve, the curves of either Owens' Iowa data, or of the scores made on the Concept Mastery test by the spouses of Terman's Gifted subjects, it was possible to construct a 50-year curve from just two longitudinal samples. Rather than choose between them, I put both adult studies onto the curve. Each one contributes something that the other lacks. The Alpha test probably samples a wider variety of abilities, but the Concept Mastery test offers more possibility of expansion into higher scores. Also, the Terman study subjects covered a sufficiently wide range of ages tested that it was possible to plot approximate scores by age at 5-year intervals from 20 to 50 years. Thus from the Terman material it appears that a relatively greater proportion of this adult growth occurs between 20 and 25 years, with the subsequent increments relatively smaller, and constant.

Of course, this curve can be viewed only as a tentative one based on limited tests and limited samples. In constructing more adequate curves we need to take into account many things. For example, different mental functions appear to have different rates of growth with different ages at maximal contribution to the total. In order to spell this out we need to have both more clearly defined and more adequately measured subtests or "factors."

*Changing Components of the Curve*

It seems to me probable that in the early part of the curve the independent factors tend to occur successively, with simultaneously developing and operating factors appearing only after some complexity of intellectual functioning has been achieved. For example, in the Berkeley Growth Study both the early age trends in standard deviations, and the patterns of correlations for consistency of scores, indicate age changes in the nature of abilities. These changes are also shown by Hofstaetter's (15) factor analysis of my table of correlations for consistency of scores from birth through 18 years (2). He obtained 3 distinct factors that operate successively. The first, which he named Sensory-Motor Alertness, is predominant for the first two years, with a very high loading at months 7 through 12. The second factor, "Persistence" is high from

*It is interesting to note that the childhood period of acceleration in both of these studies comes *before* the adolescent spurt of physical growth in boys. For the Berkeley study it occurs in both sexes at the same age and is over before physical acceleration starts.

2 to 4 years, while the third, "Manipulating Symbols" accounts for most of the variance after 4 years. The three factors are about equal in weight at two years.

This use of the total scores gives evidence for some age-specific factors and even indicates the beginnings of concurrent factors that operate over longer segments of growth. The total scores are made up from a variety of behaviors which could represent several different factors. Support for only general factors in infancy is lent by the study of Richards and Nelson (31) who factored the Gesell Scale scores for 80 infants at 6, 12 and 18 months. In this scale they included gross motor coordinations that are left out of the California First-year Mental Scale. The fact that they obtained just two factors, "Alertness" and "Motor Ability" would make it appear that I had already separated the factorially independent functions in the first year by putting the gross motor coordinations into a Scale of Motor Ability. Therefore, the "Sensory Motor Alertness" factor of Hofstaetter and the "Alertness" factor of Richards and Nelson appear to be practically identical. This factor may very well represent the lion's share of "intelligence" for the first 9 months. After this Hofstaetter's factors II and III appear to be operating concurrently for a while.

I shall not attempt to cover the work on factor analysis of intelligence tests. There are others here far more able to do this. But it may be of interest to consider a few isolated bits from longitudinal studies that it seems to me contribute to our understanding of this general problem.

Freeman and Flory's (12) longitudinal data show different slopes of increment for their four subtests and evidence for different ages at highest ability. Relative to the means and S.D.'s of the 17-year scores the Analogies score has shown most change during the 9-year interval, 8 to 17 years, with Opposites scores gaining second. The slopes of these curves indicate that they have approached close to their mature status. The slower-gaining Vocabulary and Completion tests appear to be still gaining at 17 years (4).

Some data based on retest scores on the Wechsler-Bellevue for my small Berkeley Growth Study sample may be relevant in this same connection. This test was given at four ages (16, 18, 21 and 25 years), the number for the first three testings ranging from 35 to 45. So far, 24 have been given the test at 25 years. For the full scale and for the Verbal and Performance halves, there is no indication that these young adults have reached their intellectual ceiling. But for this group, the subtests of the Wechsler-Bellevue are unequal both in difficulty and in the slo[illegible] of their age curves. Each appears to be following a different course of change with age. These differences show up most clearly when constant

subsamples are selected, so that the same cases are included at all ages used in the growth curves.

Other subsamples, selected to be homogeneous in certain respects, show interesting differences in scores and growth rates. For example, in a division by sexes boys and girls do about equally well on the Full Scale but the boys are better on the Verbal, and the girls on the Performance Scale.

The group was divided into higher and lower intelligence halves on the basis of their average scores at 16, 17 and 18 years. The highly intelligent group was found to do equally well on the Verbal and Performance parts of the test. The lower-scoring less "intelligent" half, however, does much less well on the Verbal than on the Performance Scale. Both highs and lows exhibit increasing scores but with some indication that the high group is approaching a ceiling. This ceiling is at least in part due to a lack of top in the Wechsler-Bellevue scale.

If we consider the 11 subtests separately, for constant sub-samples, the greater differences between the higher and the lower intelligence groups occur in the 6 verbal tests. Among the 5 Performance tests the two groups differ most in Picture Arrangement and Block Design (tests that require organization of spatial relations) and least in Picture Completion and Digit Symbol substitution (tests that require recognition of patterns). Those subtests in which the highs do not show continued increase in scores are tests in which they have already reached the upper limits of the scale. The lows show continued growth in most subtests but appear to have reached their top capacities in the Vocabulary and Block Design subtests, and possibly Picture Arrangement. We may have here some indications of differential growth according to level of ability.

These comparisons are on a very small sample, and on a test that is not completely adequate for such purposes. The data therefore are only suggestive, but they do seem to indicate some probable differences in ability-tied growth rates for different intellectual functions. However, credence may be lent to these differences by the fact that they are in general congruent with those found by H. E. Jones for the Adolescent Growth Study for a similar age range on the subtests of the Terman Group test (18).

To carry the picture beyond 25 years we may turn to the consideration of age changes in both cross-sectional and longitudinal studies.

The cross-sectional studies tend to agree in the finding that some abilities, such as information and word knowledge, are maintained with little or no loss to an advanced age, while other abilities such as arithmetic, analogies, and organization of spatial relations, decline with age

(after 20 or 30) at varying rates (17). Those functions that drop off most rapidly in the older subjects appear to differ with the tests used, and to some extent with the populations studied. The more recent the studies and the more complete the population sample included, the less do scores drop with advancing age. Other differences in age trends are often found to be related to general intellectual level of the subjects and to the amount of their education.

In the longitudinal studies, however, there is an invariable finding that the scores in at least some subtests are higher on the second testing. The earlier of these reports, such as the studies of Freeman and Flory (12), and of R. L. Thorndike (38), were usually on tests of young adults, mostly college students in their late teens and early twenties. But more recently repeat test scores have been secured on adults at later ages. These also show increased scores earned at the later testing after the subjects have grown older, with 12 to 30 years elapsed time between tests. I should like here to summarize some of these studies.

Owens (27) repeated the Army Alpha test on fifty-year old men who had taken this same test thirty-one years before as college freshmen, and found an increase in tested ability at the later age, with greatest increase in the Information test, and least in Arithmetic. Bayley and Oden (7) found increases on repeat tests after a 12-year interval, for gifted adults on Terman's difficult Concept Mastery test (35). When these scores are expressed in Standard Deviation units for each subtest at its initial administration, there was twice as much increase in the Synonym-Antonym (or word-knowledge) half as in the Analogies (or abstract relationships) half of the test. The increases occur in all age groups tested (20 to 50 years), though they are least in the Analogies test for the older ages. Nisbet (26) reports repeat tests after a lapse of about 23 years, on 141 teachers in Aberdeen, Scotland, who were first tested as post-graduate students when about 22 years old. On a timed verbal group test (Simplex Group Intelligence Scale) improvement occurred on all 14 subtests; the increase was significant in all but one. Greatest increases, expressed in S.D. of the first testing, were in Substitution, .71 $\sigma$, Vocabulary, .67 $\sigma$, and Number Series, .65 $\sigma$; they were least in Digit Memory, .16 $\sigma$, Verbal Rearrangement, .21 $\sigma$, and Analogies, .26 $\sigma$. H. E. Jones (18) has reported increased Terman Group Test scores between the 16½ year tests and retests at 33 years for 83 cases of the Berkeley Adolescent Growth Study. He found smaller gains in the lowest-scoring quartile of his population and in those subtests involving problem-solving. The greatest gains were in Vocabulary.

Thus, the studies cited, both the cross-sectional and the longitudinal, show similar age trends in changing organization. At the older ages the

subjects do relatively better in tests of information and word knowledge and less well in tests of reasoning and seeing relationships.

*Evaluating Intellectual Changes in Adults*

The difference between the two methods of study is found in comparisons of the actual scores. In the longitudinal studies the subjects nearly always do better at the later testing, when they are older, on most of the subtests, as well as on total scores. Although the longitudinal studies were made on different populations and using different tests, they agree in finding that different intellectual functions show varying amounts of increase, ranging from no change to as much as a standard deviation between test and retest. In no instance do these studies show the precipitous decline in scores after the age 25 to 30 that occurs in most cross-sectional studies of age differences.

Let us consider some of the things that could account for this difference.

(a) When subjects of different ages are tested at one time, there is the problem of selecting comparable samples. Apparent changes in relative scores could be artefacts of differential sample selection for subjects of different ages. In the older age groups there may be selective elimination of certain segments of the population, through deaths or through lack of cooperation. It is true that when a complete population is tested, e.g., a whole Vermont community (19) or an entire prison population (9), the curve of intelligence scores shows less decline with age. We might conclude that when willingness to be tested is a selective factor the brighter older persons are less cooperative about taking tests or else more adept at finding excuses. In any event, when the same persons are retested at successive ages, whatever the selective factors in sampling, we do have a constant sample.

(b) But with a constant sample there is the ever-present problem in psychological testing, the inescapable fact that a test is never the same for a person the second time he takes it. When comparing retests it is necessary to take into account possible practice effects and generally increased familiarity with the testing procedures. However, in many studies practice effects are of little import. When the elapsed time between tests has been as much as 12 to 30 years (as in the Owens, the Nisbet, and the Bayley and Oden studies), it seems obvious that there can be very little, if any, direct memory of problems and their solutions from the first to the second test. Also, when alternate test forms are used there may be very little increase after short intervals. I have, for example, some data on practice effects of the Concept Mastery test. There was practically no carry-over from one form to the other after an interval of only 1 or 2 weeks for the sample of 148 on which the 2

forms of the scale were equated. The difference was less than two points or about 1/15 of the S.D. of scores.

(c) Another important difference between the longitudinal and cross-sectional data is the *temporal* one. Perhaps this has not been given enough weight. In cross-sectional studies all tests are given at the same time, while in longitudinal studies not only do the same subjects grow older before their retests, but with the lapse of time they have all experienced the same general changes in the environment, and have been responding to similar changes in such things as world events and means of communication.

It is relevant here to point out that the scores earned on the Concept Mastery test by Terman's Gifted Study subjects who were tested only once increased just as much in the 12-year interval, 1939-40 to 1950-52, as did the scores of the twice-tested subjects. One hundred-twenty-nine of the men who missed the first administration of the test but took it in 1950-52 earned scores averaging half an S.D. above the mean for the men tested in 1939. A similar increase over the 1939 means was found for 98 Gifted Study women who were first tested in 1950-52. For each testing date the means of the once-tested and twice-tested groups are very similar. Thus, there appears to have been no selective factor differentiating the groups: whether or not a subject was tested in 1939 and/or 1951 was pretty much a matter of his geographical availability. Furthermore, for these bright people the average scores earned at the later date and older age were evidently not dependent on the experience of having taken the Concept Mastery test at the earlier date. In general, then, it seems to me that we should consider very seriously the possibility that at least some kinds of intelligence may very well continue to improve slowly from 20 to 50 years or older.

### *Environmental Determiners in Intellectual Growth*

So far I have touched only briefly on the whole array of potential influences on intellectual growth that we may classify as environmental. This is a very general term, including both emotional climate and opportunities for intellectual stimulation and for practice in intellectual activities such as reasoning and problem solving. No organism develops in complete absence of stimulation, and up to a certain stage in infancy the minimal requirements for life furnish adequate stimuli to afford normal development. Ordinarily the human infant's environment is much richer than these minimal conditions. The infant responds to those parts of his environment that are relevant for his degree of maturity (neural, sensory-motor, perceptual and organizational). Also, as he grows the actively healthy child will interact with his surroundings,

seeking stimulation to the extent of his capacities to utilize and cope with it. Those who are inherently gifted may well be the ones who continue to seek out and to find challenging intellectual problems and experiences in any normal life situation. However, there are probably here also inherent individual differences in this kind of active intellectual curiosity. Granting this, we may further explore the possibility that extreme environmental deprivation, or depressing emotional climate could restrict the growth of even the most intellectually alert, while optimal intellectual stimulation and emotional climate could enhance the development even of those with little inherent drive or capacity.

If we assume the relevance of these variables to the growth of intelligence, then we need to study the conditions of their effects on the course of mental development. We should inquire what kinds of emotional climate are optimal at what ages, what effects the attitudes of responsible adults such as parents and teachers have on intelligence, and whether certain attitudes are more important at some ages than at others. We should inquire further: What kinds of intellectual stimulation, or "environmental enrichment" are optimal for infants, preschoolers, children, adults? Is deprivation at certain stages crucial in determining whether development will be normal? That is, are there environmental "critical stages" for mental growth analogous to those embryological stages at which trauma can result in such deformities as cleft palate, ovarian agenesis, or possibly Mongolism and other conditions associated with feeblemindedness?

We have now some relevant information, offering tentative answers to some of these questions, but so far most of the information is based on casual observations or on studies that have not been sufficiently well designed and controlled to give us definitive answers.

A number of reports (8, 34) suggest the retarding effect on infants of life in institutions which offer little in either normal parental care or mental stimulation. But we do not yet know either the amount or generality of this effect, or on the other hand, what are the crucial aspects in regard to type of care, the nature of deprivation, or the ages at which the child experiences them. The fact that institutions do not invariably depress intelligence is brought out by Rheingold (29) who studied the effects of a significant caretaking mother-figure on 6-month old institutional babies. The six babies she cared for 40 hours a week for 8 weeks, compared with 6 control babies, showed no significant differences in IQ at the end of care.

There are studies that show significantly low IQ's for children in backward rural communities (33, 12). The studies of racial differences

in intelligence are plagued with the problems of environmental impoverishment in certain racial groups.

The many comparisons that show a correlation between socio-economic factors and IQ have made clear the fact of the relationship, but give little information about the specific ways in which such factors might effect mental growth.

Studies of the relation between intelligence scores and such personality factors as emotional tone, effort, and persistence, usually show a moderate positive correlation in young children. But so far we have little information on the effect of long-term emotional influences on the growth curve of intelligence. However, we do have reports of the differential effects of praise and blame on learning, and the depressive effect of anxiety on achievement in school (28). Ability to learn is often classed as one form of intelligence.

In the Berkeley Growth Study (5) as well as in other studies, such things as the children's emotional tone and generally optimal conditions for testing yield $r$'s with IQ that are in the neighborhood of about .30 (24, 16, 43). However, when we considered individual children, and for each child correlated his "Optimal" and "Attitude" scores with his intelligence ratings for series of 10 to 15 ages we found a wide spread of $r$'s from +.77 to —.46, with a mean of +.20. Evidently the children differed in the effect of current emotional state on performance, and for some children other factors (presumably maturational and hereditary) were of predominant importance for mental growth (5).

As for the long-term effects of emotional climate, Schaefer, Bell and I are now in the process of getting some tentative information from the Berkeley Growth Study (32). Personality characteristics of the mothers, relevant to their behavior toward these children, have been rated. The ratings were made on a scale designed for use with descriptive protocols of the mothers, made at the times they came with their children for the tests when the children were between 1 month and 3 years of age. Some of these behavior traits were found to be relatively stable over a 10-year period that could be compared. The most stable of the maternal traits are: cooperativeness, use of fear to control the child, irritability, tendencies to ignore the child, to reject the homemaking role, to evaluate the child positively, to express affection toward the child, to treat him as an equal, and to be strict with him. The 32 traits could be clustered into two main variables, which we have labeled by the "good" end of the scale, (a) Positive Attitudes, and (b) Autonomy of the Child. The few correlations we have computed with intelligence indicate that there are relations with these maternal variables, but that the nature of the relationship changes with the age of the child. If these preliminary

findings should hold up, it would appear that during the first year of life higher scores tend to be earned by babies whose mothers are intrusive, dominating and punitive, while by the time they reach school age the reverse is true and the high-scorers' mothers are characterized as cooperative, evaluating their children positively, expressing affection toward them, and allowing them autonomy as individuals.

This analysis is still so incomplete that I hesitate to mention it. But there are other studies that corroborate it in a general way, though they are not directly comparable in methods or in the variables used. For example, Macfarlane, Allen and Honzik (24) report correlations between children's mental test scores and the number of their problems (as reported by the mothers). At 21 months and 3 years the *r*'s with IQ tend to be positive, but from 4 through 14, the *r*'s are negative (around —.30), indicating a tendency after 3 years for high IQ to go with fewer problems. There are no direct correlations reported here between the IQ's and parental attitudes. However, in another paper (16) the same authors report individual cases whose IQ's appear to be related to parental behaviors. Wittenborn (43), working on data from the Yale Clinic of Child Development, reports some *r*'s between 5-year Binet IQ and certain characteristics of adoptive parents. These *r*'s also vary around .30, with the adoptive parents' Ambition, Education-Occupation, and Age-Duration of marriage. These variables probably reflect some enduring attitudes and expectations of the parents.

If we were to develop an environmental criterion of conditions fostering high intelligence, it might be composed of scores on some of the following variables: 1. Characteristics of the parents (or responsible adults) in respect to: (a) Understanding of the child's capacities and readiness for tasks of given difficulties, (b) willingness to grant the child autonomy relative to his capacities, (c) ability to offer stimulating experiences without overly strong pressures to high achievement, (d) warm affectionate acceptance of child as an individual in his own right. 2. Environmental opportunities geared to the child's stage of development. These last will include good teaching and varieties of experience, perhaps through the media of such things as radio and television, as well as travel, and discussions of ideas.

Wechsler (40) has recently discussed the possible causes of the differences between the 1939 and 1955 age-curves of intelligence as measured by the two forms of his scale. The 1955 curve is at most ages higher than the earlier one, and scores do not start to drop until after 30 years. In addition to the usual reasons (sampling, educational level and test-wiseness) for this generally increased performance, he offers a fourth of a very different kind. To quote from his abstract: "Finally, one may

posit that the improved performance of the American adult on tests of intelligence could be due in part to the improving general health and virility of the population during the last two decades. Advances in medical and social hygiene have seemingly not only served to increase life expectancy but extended the period of intellectual as well as physical vigor into later maturity."

We find some support for this suggestion of Wechsler's from studies of children's nutritional status. Many studies report low positive $r$'s with IQ of such variables as size, health and physical maturity (10, 20, 36). It is, therefore, quite possible that generally improved health and living conditions in recent years are reflected, not only in generally greater physical size, but also in greater and more prolonged mental vigor.

I should like to suggest to Wechsler that he go a step further, and if his hypothesis is true, then possibly his curve falls off after 30 because, in his cross-section sample, his older subjects have not had the advantage of growing up in this generally more healthful world.

Another step beyond this is to suggest that not only is the physical environment improved in the last two decades but also the psychological environment. Perhaps in addition to more years of schooling for more people we have actually progressed in our educational effectiveness, grade for grade. Also, it seems rather obvious that the general environment in which children are now growing up is richer: there is more knowledge available and better communication of it; travel is easier, and more children can, with less trouble, have varied experiences. Possibly also our knowledge of child training and mental hygiene are influencing parental practices in a healthful way.

If these things are true in any significant degree, and if progress in mental and physical health continues, it may become necessary repeatedly to construct new norms for intelligence every decade or so. Furthermore, in considering age changes in older people, we may need to evaluate their scores according to norms standardized at appropriate calendar years, rather than for age only.

REFERENCES

1. Bayley, N. A consideration of age changes in mental organization. *Primer Congreso Panamericano de Gerontologia* (Book of Abstracts) Ciudad Universitaria, Mexico, 1956, 31.
2. Bayley, N. Consistency and variability in the growth of intelligence from birth to eighteen years. *J. Genet. Psychol.* 1949, 75, 165-196.
3. Bayley, N. Data on the growth of intelligence between 16 and 21 years as measured by the Wechsler-Bellevue Scale. *J. Genet. Psychol.* (In Press).
4. Bayley, N. Development and Maturation. *Theoretical Foundations of Psychology*, (Helson, Editor) New York: Van Nostrand, 1951.
5. Bayley, N. Factors influencing the growth of intelligence in young children. *Yearb. Nat. Soc. Stud. Educ.*, 1940, 39, 49-79.

6. Bayley, N. On the growth of intelligence. *Amer. Psychol.*, 1955, 10, 805-818.
7. Bayley, N. and Oden, M. H. The maintenance of intellectual ability in gifted adults. *J. Gerontol.* 1955, 10, 91-107.
8. Bowlby, J. Maternal care and mental health. *WHO Technical Mono.* Series, No. 2, Geneva, 1951.
9. Corsini, R. J. and Fassett, K. K. Intelligence and aging. *J. Genet. Psychol.*, 1953, 83, 249-264.
10. Ebert, E. and Simmons, K. The Brush Foundation Study of Child Growth and Development. I. Psychometric Tests. *Monog. Soc. Res. Child Devel.* 1943, 8, 113.
11. Foulds, G. A. and Raven, J. C. Normal changes in the mental abilities of adults as age advances. *J. Ment. Sci.*, 1948, 94, 133-142.
12. Freeman, F. N. and Flory, C. D. Growth in intellectual ability as measured by repeated tests. *Monog. Soc. Res. Child Devel.*, 1937, 2, 116.
13. Garrett, H. E. A developmental theory of intelligence. *Amer. Psychol.*, 1946, 1, 372-378.
14. Guilford, J. P. The structure of intellect. *Psych. Bull.*, 1956, 53, 267-293.
15. Hofstaetter, P. R. The changing composition of "intelligence": a study of *t*-technique. *J. Genet. Psychol.*, 1954, 85, 159-164.
16. Honzik, M. P., Macfarlane, J. W. and Allen, L. The stability of mental test performance between two and eighteen years. *J. Exper. Educ.*, 1948, 17, 309-324.
17. Jones, H. E. Age changes in adult mental abilities. *Old Age in the Modern World*. London: E. & S. Livingstone, Ltd., 1955, 267-274.
18. Jones, H. E. Trends in mental abilities. (Paper read in a symposium at the 1955 meeting of the American Psychological Association.) *Amer. Psychol.*, 1955, 10, 405.
19. Jones, H. E. and Conrad, H. S. The growth and decline of intelligence: A study of a homogeneous group between the ages of ten and sixty. *Genet. Psychol. Monog.*, 1933, 13, 223-294.
20. Knobloch, H. and Pasamanick, B. Further observations on the behavioral development of Negro children. *J. Genet. Psychol.*, 1953, 83, 137-159.
21. Lorge, I. Aging and intelligence. *J. Chronic Diseases*, 1956, 4, 131-139.
22. Lorge, I. Schooling makes a difference. *Teachers College Record*, 1945, 46, 483-492.
23. Lorge, I. The influence of the test upon the nature of mental decline as a function of age. *J. Educ. Psychol.* 1936, 27, 100-110.
24. Macfarlane, J. W., Allen, L. and Honzik, M. P. A developmental study of the behavior problems of normal children between 21 months and 14 years. Publi. in Child Devel., Berkeley, U. of Calif. Press, Vol. 2, 1954.
25. Miles, W. R. Psychological aspects of aging. *Problems of Aging*. 2nd ed. (Cowdry, E. V., Editor) Baltimore: Williams & Wilkins, 1942, 756-784.
26. Nisbet, J. Family environment and intelligence. *Eugenics Review*, 1953, 45, 31-40.
27. Owens, W. A. Age and mental abilities: a longitudinal study. *Genet. Psychol. Monog.*, 1953, 48, 3-54.
28. Palermo, D. S., Castenada, A. and McCandless, B. R. The relationship of anxiety in children to performance in a complex learning task. *Child Devel.*, 1956, 27, 333-337.
29. Rheingold, H. L. The modification of social responsiveness in institutional babies. *Monog. Soc. Res. Child Devel.* (In Press).
30. Ribble, M. Infantile experience in relation to personality development. *Personality and the Behavior Disorders* (Hunt, J. McV., Editor), New York: The Ronald Press, 1944.
31. Richards, T. W. and Nelson, V. L. Abilities of infants during the first eighteen months. *J. Genet. Psychol.* 1939, 55, 299-318.
32. Schaefer, E. S., Bell, R. Q. and Bayley, N. Quantification of maternal behavior and consistency of mother-child interaction. *Amer. Psychol.*, 1956, 11, 404. (Abstract)
33. Sherman, M. and Henry, T. R. *Hollow Folk*. New York: Thomas Y. Crowell Co., 1933.
34. Spitz, R. A. An inquiry into the genesis of psychiatric conditions in early childhood. I. Hospitalism. *Psychoanalytic Study of the Child*, 1945, 1, 53-74.
35. Terman, L. M. *Concept Mastery Test* (Manual), New York: The Psychological Corp., 1956, 10.

36. TERMAN, L. M. AND ODEN, M. H. *The Gifted Child Grows Up.* Vol. IV, *Genetic Studies of Genius.* Stanford Univ. Press, 1947, 448.
37. THORNDIKE, E. L. The measurement of intelligence. New York: Teachers College, Columbia Univ., 1927, 616.
38. THORNDIKE, R. L. Growth of intelligence during adolescence. *J. Genet. Psychol.*, 1948, 72, 11-15.
39. THURSTONE, L. L. Primary mental abilities. *Psychometric Monog.*, No. 1, 1938.
40. WECHSLER, D. Recent changes in rate of decline of intelligence test scores of the American adult. *Primer Congreso Panamericano de Gerontologia* (Book of Abstracts) Ciudad Universitaria, Mexico. 1956, 267-268.
41. WECHSLER, D. *The Measurement of Adult Intelligence.* Baltimore: Williams & Wilkins, 1944.
42. WHEELER, L. R. A comparative study of the intelligence of east Tennessee mountain children. *J. Educ. Psychol.*, 1942, 33, 321-333.
43. WITTENBORN, J. R. AND OTHERS. A study of adoptive children: III Relationships between some aspects of development and some aspects of environment for adoptive children. *Psychol. Monogr.*, 1956, 70, No. 3, 93-115.

# Implications for Test Construction

THELMA G. THURSTONE

Studies of the growth of human intelligence are of interest and importance for two fields. The first of these is the general theoretical psychological problem of how intelligence develops throughout the entire life span, and the second concerns the implications of the scientific findings for application in all of the fields in which intelligence tests are used. It is the purpose of this paper to report the findings of several studies on the growth of the separate mental abilities and to try to make some suggestions for their application in test construction.

Dr. Nancy Bayley's paper in the *American Psychologist* for 1955 presented an excellent summary of the findings of a large number of studies of the growth of intelligence. Most of the studies were done at Berkeley and represent work with individual testing of people of all ages. Dr. Bayley's paper which was just referred to, contains a note of dissatisfaction. This is not surprising because she set herself an extremely difficult task to study intelligence from birth or even conception through senescence. Though the data on which some of the conclusions are based were not satisfactory to the author, I think that we must all admire the extension of her thinking beyond the actual data. It is from this kind of thinking that we shall discover new ways of attacking the problem of the growth of intelligence in the age ranges where it has been most difficult to study, that is, the earliest years of life and the last years of life.

In this paper I shall restrict my comments to a much narrower age range and perhaps a narrower subject. Specifically, I shall refer to the measurement of primary mental abilities or factors of intelligence. So far the studies in this field have been concerned with children of school age and early adult life and have been based almost entirely on group tests. This does not mean that we have not been interested in the problem of measuring intelligence at the early ages and the later ages but simply that we have not had suitable tests for the youngest years nor tests nor time for the later years. Perhaps there is some justification other than the ease in approaching the problem for having confined our studies to the school age range. It is in the schools and in the military service and employment where we find the greatest application of the results of test construction.

A first area of study in which the results of studies of intelligence at

various age levels have some implication for test construction is in the results of the factor analysis studies at these different ages. Factor analyses have been made of the results of administering large batteries of tests to large numbers of children from kindergarten age through early maturity. In these attempts we have worked our way down beginning with the college age group, later working with high school children, then with eighth grade children, and somewhat later with kindergarten and first grade children. One of the most interesting results of all of those studies was that similar factor patterns of abilities emerged in all of the studies. This finding has allowed a type of study of rates of mental growth which was not possible when we were dealing with a composite or average score such as mental age or IQ, which were involved in earlier studies of patterns of mental growth.

A related question concerns the pattern of abilities to be found among people of varying ability levels when age is constant. The study made by Bechtold at the Psychometric Laboratory at the University of Chicago showed similar patterns of mental organization for groups of children varying widely in intelligence but who were all at the fourteen year age level.

This invariance in the factor pattern of abilities has very definite implications for the construction of intelligence tests and their use. It implies that we should be able to construct tests of the same abilities at the successive age levels in which we are interested. Certain limitations in the kind of group tests which can be administered to very young children who cannot yet read or write prevent a complete application of this principle. To be more specific, the word fluency factor was clearly present among kindergarten and first grade children, but the use of group tests of this ability at a very early age level has not been accomplished yet. We have so far found no way to measure this ability except by giving the child a chance to write or to say words as fast as he can. In the first case the young child cannot write fast enough to give us an indication of his fluency, and in the second case we are turned back to the use of individual tests. Both for theoretical studies of the growth of intelligence and the development of further testing methods these individual tests have been used, but for any sort of practical use in the school the individual testing techniques have not yet received attention.

A second area of research for the primary mental abilities involves a longer range predictive study of tests than is ordinarily made. The study which will be mentioned here will be reported fully in a Psychometric Laboratory Report in the near future. It is concerned with the correlations between the primary mental abilities scores of elementary

school children and their success in the school subjects of reading, spelling, and arithmetic. These achievement variables were selected because they were the ones which were available in the school records. The study was carried out partly because of a need to make a recommendation about the frequency at which testing needs to be done in the school situation and partly to decide how many levels of group tests need to be prepared for an adequate guidance program in the elementary schools.

The data for this study were obtained from about seven hundred children in seven elementary schools in Chicago in the spring of 1952. At that time the primary mental abilities tests for first grade children had been a standard part of the elementary school guidance program since the fall of 1946. This meant that the children who had entered first grade in 1946 were now in the sixth grade. The primary mental abilities test for the intermediate grades had been given in the fourth grade since the fall of 1949. This means that those fourth grade children who had taken the test in the fall of 1949 were in sixth grade. All of the children whose records were studied had taken the primary mental abilities tests in first grade and in fourth grade and were now in the sixth grade. Reading scores for all of these children were available for grades two through six. Spelling and arithmetic scores were somewhat harder to obtain, but spelling scores were available for some children in grades four, five, and six while arithmetic scores were available for most of the children from grades two through six. It is not the usual practice to administer achievement tests in the elementary schools every year but for the purpose of this study, certain schools where that was the practice were selected. Thus we were able to study the correlation between total scores on the intelligence tests and scores in the separate abilities for five different time intervals. We had anticipated that the correlations would drop markedly with a lapse in time; that for example, the correlation between the first grade intelligence test score and reading in the second grade would be much higher than for the correlation between intelligence test scores in the first grade and reading scores of the same children in the sixth grade. It was a very surprising finding that there was on the average no loss in predictive value of the test scores but that actually there seemed to be a slight rise. The correlation between total test scores in the first grade and reading in the second grade was .42; the correlation between total scores and reading for the same intelligence test taken in the first grade and reading scores in the sixth grade was .51; to take another example, the coefficient of correlation between score in the quantitative part of the first grade intelligence test and second grade arithmetic scores was .31, while the

correlation between these same test scores and arithmetic test scores made in the sixth grade was .51.

For the intelligence test which was given at the fourth grade level when the children were nine years old, we have three years of achievement test scores with which to compare them. The correlation between total score on the intelligence test taken in the fourth grade and the reading test taken within two weeks of the same time was .56; while the correlation between these test scores and a reading test taken two years later when the children were in sixth grade was .61. For arithmetic tests and intelligence test scores the correlations rose from 59 to 67. To take a single one of the primary mental abilities' scores, the correlation between the *V*, verbal comprehension, score at the fourth grade level and reading, the correlation was .71. Two years later the correlation between these intelligence test scores and sixth grade reading scores was .78. For spelling the shift was from 64 to 71 and for arithmetic from 54 to 60. The interpretation of this consistent although slight rise in the correlations between the test score and achievement scores over a period of time is difficult. The conclusion that intelligence is a more important factor in later school achievement than in early school achievement is suggested, but we cannot make a definite commitment on this point. However, it does seem perfectly clear that there is a continuity in the intelligence functions tested and that the early measures are significant for a longer period of time than we would have anticipated. The results give us justification for believing that guidance of elementary school children on the basis of tests is a continuous affair, that tests can be useful throughout a wider time range than they once were considered to be. For test construction the implication is definitely suggested that it is worthwhile to try to construct the tests for the various age ranges which have overlapping abilities. The tests for some age groups may contain tests which are not present in others, but an adequate guidance program seems to be impossible unless some of the same tests occur continuously throughout the age range.

A third area of investigation which throws some light on the problem under discussion is a study of differential rates of growth of several of the primary mental abilities. We were fortunate to have data for thousands of children covering an age range from five to nineteen. Data were not available for all children on all of the abilities, but the data were so extensive that we were able to plot growth curves on an absolute scale for seven of the primary mental abilities. These have all been described elsewhere and will be only mentioned here. They are perceptual speed, space, reasoning, number ability, memory, verbal comprehension, and word fluency. The first report on this study was

made in the spring of 1954 and later published in *Science* by L. L. Thurstone. The seven curves plotted on the same chart show very clearly a difference in the rate of maturation of the separate abilities. We know, of course, that all of the abilities start from zero ability at approximately the same age. If we take 80% of adult performance as a criterion, we find that the perceptual speed factor reaches this criterion at the age of twelve years, maturing rather early as compared with the other abilities. The reasoning factor and the space factor obtained about the same level of performance at the age of fourteen years. Rote memory and number ability reached the same level at about sixteen years. The verbal comprehension factor develops considerably more slowly and does not reach this criterion until the age of eighteen. The word fluency factor matures even more slowly and has not reached this criterion at age twenty. Remembering the criterion chosen, 80% of the adult level, it is clear that considerable growth takes place in all of the functions after the year levels mentioned here.

A number of implications for test construction can be read from these results. The first of these is that we should be able to construct tests of the several abilities at all of the age levels with which we are concerned in the age range of school children and early adult life. Certainly it is not difficult to look forward to the construction of tests of these same abilities at the more advanced age levels. So far as our growth curves have been determined, we have not yet found any indication of the age at which the downward trend or decrement in any of the abilities may start. We know that it is there, but it seems that for the present we must think of it as a question for experimental determination. At the other end of the age range, the very earliest years, we shall have to look forward to the development of other techniques in measuring the primary mental abilities in order to extend the growth curve downward experimentally. It is going to require considerable ingenuity in observation in the behavior of young children and in test construction to invent the kinds of test items which will be appropriate for measuring the primary mental abilities at the earliest age levels. It is the conviction of this writer that such an attempt may not be so far away.

So far we have discussed the implication of three experimental studies for the construction of intelligence tests. Certain general implications should be discussed now. It seems that although we feel that we have made progress in the measurement of intelligence by analyzing intelligence into a number of components and constructing tests to measure these separately, we should look upon our present tests as in need of considerable refinement. A few definite examples may be mentioned. In the earliest factor study an ability which we called per-

ceptual speed was clearly identified. Later experiments with numerous tests which seemed to be related to this ability somewhat weakened our satisfaction and we found it necessary to make a very detailed study of the perceptual abilities, issuing in a monograph devoted entirely to that subject. The field of memory is another one which seems to be due for the same kind of revision. Although we seem to be justified in speaking of some people as having a good memory and others as having a poor memory, it seems that we must go much farther and define the kind of memory we are talking about in each instance. A very detailed factor study of memory abilities using about thirty tests with a large group of high school children has been carried out. The data have been recently reanalyzed, and publication will follow this later work. It seems now that we must regard memory as a complex of several abilities, though for practical testing purposes we may restrict our tests to a rather limited number.

The ability to think in visual terms which we have called space thinking has been broken down into at least three space factors. These have been investigated intensely by Mr. Thomas Jeffrey of the Psychometric Laboratory and the results are being submitted to the University of Chicago in a Ph.D. dissertation. Quite recently, we have had some interest to question the relation of various attributes of space test material to the ability of subjects to handle such material well. We are wondering whether the amount of rotation in the figures is a factor which may even be related to success at different ages, whether good balance in the figures may be important, and whether a number of other qualities of the test influence the validity of tests at the various age levels. It seems that the time is already here when we need to go into refinement or polishing of some of our present test techniques.

Perhaps the most important implication of such studies as I have been discussing lies in the direction of recognizing the need for the construction of individual tests of the mental abilities. The group tests we now use have been useful in studying of growth curves for numbers of children, but the thorough study of mental growth seems to point in the direction of studies of individual growth, as it has even in the case of physical growth. It is very clear that for very young children, for people with certain types of handicaps, and for the later age ranges we shall find it difficult to make large group studies as we do with children of school age, and that individual tests of the primary abilities will be of more use to us.

When we have succeeded in isolating and describing thoroughly the most important of the human mental abilities, we shall have better tools for studying the complete range of human ability from infancy through

childhood, adolescence and maturity and through senescence. Such curves of growth based on studies of the fundamental human mental abilities will be a distinct contribution to psychological knowledge concerning the nature of the abilities and the rates of their growth and decrement. Application of these psychological principles in the construction of tests will provide serviceable tests in helping to deal with the problem of people of all ages.

# DISCUSSION

*Participants*

Nancy Bayley, William E. Coffman, Herbert S. Conrad, John T. Dailey, John W. French, Irving Lorge, Elliott Metz, Thelma G. Thurstone, Joseph Zubin.

MR. METZ: I would like to address this question to Dr. Bayley. You mentioned studies concerning the differential effects of praise and blame on learning, and consequently on growth and intelligence. I should like to ask whether the effect of acceptance and understanding without evaluation has been studied with regard to growth and intelligence.

It would seem to me that the adequacy of this technique in the clinical therapeutic situation as developed by Carl Rogers and others could be validly applied to the concepts of learning and intelligence. I would like to know what you think of this.

DR. BAYLEY: I think this is a field that has not really had any good research done in it. It is something that I hope will be studied much more carefully. There are a number of studies that show that praise and blame have differential effects on learning.

Now, the relationship between these and intelligence as such comes out suggestively in clinical studies of children's capacities, and I think there are some studies that show changing, increasing IQ's, or intelligence scores in children as they attain emotionally better states.

There are also evidences of intellectual retardation that seem to go with continued unhappy emotional states. However, in general, these correlations are not very high. There are so many other factors that make a difference. I have some good examples from my own Berkeley Growth Study, of children who were emotionally very disturbed, but whose way of handling their problems was to direct their energies to intellectual activities. The highest IQ I ever got in this Berkeley group was for a child who was disturbed, but who was strongly motivated to handle his emotional problems by becoming a highly intellectual, intelligent person. I think we all know a great number of such people in academic life.

DR. DAILEY: This is a very short point but I think it will help in thinking about a lot of this discussion. I would like to quote something that was done in the Air Force several years ago. They set about systematically to calibrate all the aptitude tests of the Services, the USES, and a number of common commercial aptitude and intelligence tests. Just for fun, we threw in the Stanford achievement test, and a couple of other batteries of achievement tests. We took the reading section,

the arithmetic section and the corresponding aptitude tests, and so on, and we did something that shouldn't be done. We took the labels off and looked at them as Test 1, 2, 3, 4, etc. The interesting thing is that without the labels you couldn't tell the achievement sub-test from the analogous aptitude sub-test. If we are talking about something like GCT or Otis IQ, we might just as well be talking about grammar school level reading and arithmetic achievement.

That might help explain a number of things. I am not trying to say that IQ is the same as reading and arithmetic, but the correlation is high enough to indicate such, and when you break it down into the IQ sub-tests and you do a factor analysis, including achievement tests, they come out in the same factors. In other words, a lot of what we are talking about in terms of group test intelligence is very much the same as educational achievement.

CHAIRMAN LORGE: Kelley has clearly indicated that from the earliest days the relationship between the so-called test of aptitude and the test of achievement has always been very strong.

Are there any other comments? Certainly our textbooks will have to be rewritten on the basis of what we have heard today.

DR. FRENCH: My question is for Dr. Thurstone. I was interested in the rising validities of some of your tests, and wondered whether you had any indication that such rises may have occurred because of rising reliabilities of the criteria.

DR. THURSTONE: I am sorry I haven't that table of correlations with me. Undoubtedly there is some improvement in the reliability of measuring reading as you progress from second grade to sixth grade. I think it wasn't anything like enough to account for shifts of as much as twenty points. That will be in the published paper, but I have forgotten the figures.

DR. CONRAD: Is it possible that the schools have an equalizing cultural effect upon the pupils so that initial home differences are reduced? That is to say, the child at home in some cases will get more reading experience; but later on this home influence may be reduced in comparison with the cumulative school influence so that the home difference which affects the achievement in the early years has a smaller effect later on, and the uniform school experience has a larger influence.

DR. THURSTONE: I am inclined to say that just the opposite happens. The differences among the children are spread out. There is more variability with age as they go through school. We find that in every study.

There are some other bits of evidence that I always hate to mention because they are so difficult of interpretation, and that is the comparisons between white and Negro school children, which have been rather popular in the last few months. Some of those seem to show greater difference between these groups as they progress through school. In other words, if you want to emphasize the effect at home, you can say—and I think we need to do it—that although the school tries to do a great deal for these children, we cannot keep up with the differences in the home environment. It continues to be important all the way through school, and both influences tend to increase the differences and also increase the variability of the groups.

DR. COFFMAN: I would like to address this question to Dr. Thurstone. You mentioned that your testing had been carried out over a wide range of ages, and later you mentioned that you used as a base for determining rate of growth, the mature adult level. How did you determine that the measures for the mature adult level were a maximum?

DR. THURSTONE: Eighty per cent of the asymptote on our growth curve, which is not projected beyond the freshman and sophomore age in college.

DR. COFFMAN: What I meant was, is it possible that the base you took was not the proper one, that actually some of those scores would have continued to rise so that the difference you found at the eighty per cent level might have been different if you had tested older people in order to get the maximum.

DR. THURSTONE: I think so. We chose eighty per cent because we were projecting those curves beyond our experimental data, and that gets pretty risky when you get up into the thin part of the curve. That is why we chose the eighty per cent rather than the one hundred per cent.

DR. ZUBIN: I think perhaps Dr. Bayley can answer this. I am referring to a study by Ben Pasamanick in which he held the weight at birth of Negro babies and white babies constant, and discovered that the differences in intelligence disappeared. He pointed out that there are three possible scientific models for explaining intelligence: heredity, environment, and the interaction between the two. Pasamanick adopts the assumption that genetically all men are born equal, which is a perfectly good model to start with. Most of the data which he was able to collect on newborn children, he says, seems to fit that model, and that life as represented by environment eats away the initial equality and produces the individual differences which we observe, by a process of attrition.

I was wondering whether you have any evidence regarding this point from your own studies.

DR. BAYLEY: In early infancy, weight at birth is one indicator of degree of maturity. This is certainly one measure that is used to determine whether an infant is premature, and if you work with infants around a month of age, you feel there is a great deal of difference in the stage of maturity of the child, and at that stage physical maturity does make a difference as to the kind of scores infants earn on the tests. I know Pasamanick has found for one and two year olds that if he divides the group in half on the basis of weight, the heavier children have higher scores and the lighter ones tend to have somewhat lower scores.

I think the differences are relatively small, and I think there are a great many factors, that is, there may be other than racial differences. We don't know whether characteristically in the early ages, weight for weight two different races are the same in general maturity. It might even vary with the particular racial origins of the Negroes. (African tribes vary in size from pygmies to giants.) I don't think we have it too clearly worked out.

CHAIRMAN LORGE: May I again thank Dr. Bayley and Dr. Thurstone, as well as the discussants from the floor, for this auspicious initial part of our program.

## GENERAL MEETING

**Changing Concepts of Personality**

# Concepts of Personality— Then and Now

GARDNER MURPHY

## I

The modern era in thinking about personality may well be defined as beginning with the general acceptance of the evolutionary principle during the third quarter of the nineteenth century, and its first fruits in the concepts and the methods of Francis Galton. The conception that traits are ingrained attributes of individual personalities reflecting the basic stuff of which the person is made, clearly formulated in Galton's era, still has theoretical value and practical utility. The conception that traits of *temperament and personality* can be measured on objectively defined scales was just as directly related to the evolutionary way of thinking as was the development of psychometric scales as measures of *intellectual individuality*. If you ask therefore what period I wish to define, when I use the word *then*, as assigned to me in this title, in making the contrast between then and now, I will say the era of Francis Galton.

The first concept defined in that era which is of special importance to us here is the concept of behavior traits as given directly in observation; traits involving clusters of behaviors which can be bracketed together under single names like courage, complacency, passivity, timidity. Traits from this point of view are conceived to be attributes so clearly identifiable and so easily capable of reliable and valid measurement that we may regard them as building stones of a personality architecture. Usually associated with such assumptions are assumptions regarding the role of heredity. Most such traits are held to be hereditary and therefore stable attributes of individuality. Some, however, are held to be environmentally acquired, and therefore subject to such temporal variation as processes of learning and unlearning may entail. An attempt has usually been made to treat both the hereditary and the acquired traits in terms of orthogonality; or, when traits obviously overlap, the attempt has been made to set up new traits in their places which will be conceptually and practically orthogonal. My attempt in the time allotted will now be to show that these central, simple ideas, while continuing to thrive and be useful, have had to compete more and more with newer ways of thinking. It will be my

task to define in these few moments a few of the major new conceptual systems of the last fifty years.

II

I will begin with ideas from contemporary biology. Genetics has now rendered very clear the fact that what is genetically simple is at the behavioral level often very complex; indeed, observed phenotypic traits are usually reducible to networks of relationships in which many genes pool their forces or interact; and during the same time in which genes are busy in forming a given phenotypic trait the same genes are busy in forming another trait. What is therefore conceptually simplest at the genetic level is quite different from that which is simplest at the phenotypic level. This will mean that if one does not know the genetics of a trait (and we seldom *do* know the genetics of a trait), we shall inevitably define attributes which in one population behave in a very different fashion from that in which they behave in another. There will be unexplained correlations and failures of correlation. Attempts at nature-nurture studies based on family resemblance, social class differentials and so on, will usually bring one headache after another. That which is conceptually a trait from a psychological point of view will not necessarily behave biologically like a trait, except under certain specific circumstances.

W. R. Thompson (3) has recently shown that in the use of factorial methods it is sometimes possible to disentangle what is genetically complex, and to show that when factor-analysis is applied four quite different types of genetic pictures may underlie what appear to be the same essential factorial structures. Of course if all that one is trying to do is to spot a trait at a given time and not even to be concerned with its development in the course of the individual, or the relation of the trait in one person to the same trait in his brother, no great harm is done. I think, however, that psychology can expect to do better than this; and that it can move forward in the light of modern genetics to a more sophisticated understanding of biological traits.

Closely related to this point is the fact that genetics has likewise shown in recent years that the temperamental and impulsive traits are subject to the same complex biological predetermination as the cognitive traits; but all of them, so far as we know, are neither simply given by heredity nor simply imposed by the environment. They represent very complex developmental entities which require both biological and environmental analyses, and of course highly sophisticated statistical analysis of the components. It is, to be sure, likely that in some cases a trait may represent to a very large degree a characteristic

determined by a single gene. But in other instances it may be better conceived as a conditioned response transferred or generalized from a response acquired in some other setting. In still another context a trait may arise as a result of the impact of massive social forces acting upon almost all members of a social group in such a fashion that specific conditioning experiences are harder to identify than the full impact of the social pattern which is being imposed upon the members of a group. Consider, for example, the sudden impact of industrialization upon a village in Tennessee or in India. This all means, then, different conceptual schemata will have to be utilized if a given trait is to be defined and effectively used.

This is, I think, likewise the context in which we must use the enormously suggestive and valuable contributions of cultural anthropology and of the cultural sciences generally in the last thirty or forty years. It is proving to be of very doubtful value to us to resort to generalizations that such and such attributes of personality are culturally acquired; for such statements leave out of account the nature of the raw material, the biology of the individual which undergoes the cultural impress; it leaves out of account the sub-cultural, family, and personal factors which are likewise involved in such impress; and it leaves out of account the attempt to specify the mode of interaction between biological and social factors in the person. Nevertheless, we must add cultural anthropology as one of the central factors which have sensitized us to the necessity of the bio-social approach.

So far, we have been willing enough to think of many traits as identifiable components in personality structure as a whole. Some temperamental traits, like threshold for rage, may actually be pretty stable in the adult, and it is conceivable that there may even be a few personality traits which are Mendelian. The term personality structure, however, has proved, in the light of cultural and clinical findings, to be a very rich and challenging notion. There are many different ways in which the simplest traits of the individual may be put together; some operate summatively, others subtractively. Sometimes when one trait is present it acts almost like an enzyme, allowing the more effective utilization of another trait. Sometimes several traits seem to form into an emergent structure, just as we may pass from the level of elements to the level of compounds, and from the level of compounds to the level of cellular structure. There are, in other words, various modes of organization of traits empirically discovered. Recently, biologists like von Bertalanffy (4) and other originators of "general system theory," have shown us that wherever living systems are involved, the problem of organization and emergence takes over and complicates the problem of showing how

the individual trait reflects the system of which it is a part. This is, of course, the problem which Gestalt psychology had likewise raised some decades ago, notably in the Gestaltist doctrine of membership character. It does not shock us today to be told that a patch of red in a landscape will look differently if the context is altered. But it still does seem to bother us, conceptually and methodologically, to be told that a *trait* is operationally different when it appears in different contexts. Coolness, for example, or the maintenance of a low level of affect, is a very different thing in a danger situation and in a social gathering.

This is of course a tremendous nuisance, both to the theorist and to the practical worker. It is, however, the price we pay if we wish to go biological and to orient our psychology to the principles which, as general system theory shows, sweep through all the sciences and hold for all types of scientific work as such. The insularity of separating psychology from biochemistry, biology, and on the other hand the cultural sciences, is one for which I should *not* like to plead.

We are then dealing with the development of ways of thinking about personality which have been repeatedly and rather successfully verbalized, and often put into a form convenient for those who think in terms of visual images. It must, however, be admitted that in terms of conceptual rigor necessary for the formulation of methods of measurement an enormous amount remains to be done. It is so much easier to go on taking directly observable forms of response, bracketing these into families, treating them as if they were entities, and then setting up various types of axes to represent their relationships and emerging from this with the pragmatic generalization that somehow it works. This can indeed be made to work if the frame of reference remains constant. It is, for example, entirely possible to set up a social situation in which an IQ will remain constant, year by year, just as it is possible to set up a situation in which the Thurstone primary mental abilities will remain unaffected by the measured environmental factors. Or indeed we might say it is possible to set up a Euclidean geometry by taking certain axioms and ruling out others. This is important and valuable for certain purposes. If, however, we wish to work within a frame of reference in which mixed and varying demands are made upon persons in mixed and changing social relationships, and with stresses within them which bring into activity various latent potentialities which had earlier been unobserved and unmeasured, and if we wish to get a conceptual system rich enough to be ready to move in new directions as new environmental factors come into our preview, we shall be forced to go on complicating our scheme. If traits and the systems within which they appear actually vary greatly with growth and with environmental conditions, what will

be the result of treating them statistically? What will happen if we complain of the low reliability of a test when in fact the flexibility, the versatility is due to the organism's varying behavior in different situations?

To the plea that our scheme has become already too complex, the only possible reply is that it is still much too simple to do justice to the kinds of complex challenges to which children and adolescents growing up in our culture react, worrying and delighting at the same time the teachers, the counselors and the parents who wish that they could understand and help them; and still too simple for the educational system of a nation which needs not only to supply an abstractly valuable way of sorting and coding the latent characteristics of its population, but needs a way of being specifically helpful to individual citizens whose complex attributes and demands are not adequately mirrored in available testing and classifying systems.

Along with the challenge already quoted from Gestalt psychology to the effect that each attribute shows membership character in the whole, goes the still greater challenge, already hinted in other words, that each person in each situation is in vital respects a new reacting entity. Certain potentials within the person are drawn out by the situation which had never been observed before, and certain potentials of the situation, if you will allow me the phrase, are grasped and drawn out by the individual which had never been grasped before. This is the problem to which Sears referred in the discussion of dyadic relationships, and a corner of the problem to which Kurt Lewin referred in his definition of the life space.

The concept of discontinuous life spaces, each of which has its own logic which must be empirically defined for each situation in terms which reflect the meaning of the testing situation to the individual testee, is one which becomes intensely practical if you are trying to describe (as, for example, Stern, Stein and Bloom (2) did in their recent volume), the relation of projective test findings to particular situational demands in particular persons. We know for example that a Rorschach is of very limited value in predicting sheer success in naval aviation. We wonder why anyone ever thought it was worth trying. We know, however, that when the relation of the person to the detailed structure of the situation to be met is fully described, the field situation can and does permit some prediction. And here with a vengeance the theoretical advances in organismic and field studies in biology and those few which have been attempted in psychology are paying off on a large scale.

I think the same appeal to complexity rather than simplicity is forced upon us as we now come into the dimensions of personality which have

been investigated most fully by clinical rather than by experimental or by sociological methods. I have in mind, of course, particularly the problem of depth dynamics, unconscious motivation, and the surface manifestations of deep level intrapsychic conflict. It has gradually become clear to non-clinical psychologists that the psychoanalytic system did not deliberately court complexity for the sake of complexity but that human beings are actually that complicated. The psychoanalytic system has become, in a few decades, as complex, one might say, as Catholic theology, Marxism, or literary criticism. We are dealing with bio-social products where to be sure we often over-complicate, but likewise, in our ignorance, often over-simplify. Now many of the traits which emerge are traits which describe conflict or the interdependence of motives, or the three-way relationships between Id, Ego and Super-Ego. There have been in the last few years—and Dr. Krugman will undoubtedly tell you more of this—attempts to re-think the methods of personality diagnosis itself to get them into alignment with the clearest and least controversial aspect of psychodynamic theory.

Insofar as this purpose can be achieved, it opens the way to conceptual clarity and the development of clearly defined test methods of many sorts, and offers the possibility that the test methods so devised can be used in actual experimentation on psychodynamics. This could, if successful, not only yield much practical knowledge as to where psychoanalytic theory is correct and where it is in error, but could likewise serve by meeting the ordinary canons of simplicity, consistency and prediction, with devices for testing our tests. If, in other words, our test systems incorporate ruthlessly the logic of observations derived from clinical practice, they will fall down whenever the clinical observations are in error, and will stand up and grow stronger with the years whenever reality rather than error is contained in the clinical observations. Hand in hand with this goes, of course, the fact that day by day clinical practice yields constantly new ideas and systems of ideas which in themselves will ultimately suggest new tests.

One serious difficulty, of course, in much clinical thinking is the tendency to think normatively and scientifically at the *same* time, making value judgments about health and illness at the same time that one strives to see what is. A child's shyness or aggressiveness, troublesome at the time, may be a phase in growth, which we see less clearly rather than more clearly, when we call it "problem behavior;" similarly with adults. A man's capacity to complete a task may depend in part on traits which clinically get very bad names.

Am I asking for too much? Am I setting sights which are too high? I am certainly not demanding everything at once. I do believe, however,

that along with certain limited uses for simple abstractions regarding traits, justified because they seem to work, at least in short term situations, there is a need for developing a system of trait conceptualizations which are in harmony with the best knowledge we have derived from general system theory, from Gestalt psychology and from clinical practice. Such conceptualization will have to take account of the inherent flexibility of personality. It will have to find out what is fixed, what is changeable, and find ways of identifying and measuring the long-term trends which often underlie the changes. Relatively little, I think, is going to be accomplished by just deriving more names and studying mathematically the relations between the traits to which these names are given—what we might call very briefly the adjective-checklist way of doing personality research. More and more, I believe, will be accomplished by studying biological realities, bio-social realities, the complex structural wholes which the biological entities produce under the impact of various bio-social forces.

I would throw in for good measure a concept which B. F. Skinner (1) has well represented, namely that it is often better to pursue one's analysis until one can really predict the behavior of a single individual. Then one may dare to say that the determining circumstances have been recognized and controlled. It is the conceptualizations about personality based upon simple building stones and simple combination of these in summative and subtractive form, that have given us the specious and I fear, short-lived, control over practical situations which schools and industry, and even the Armed Forces have been glad to pay for. The uneasiness of the last few years which these same patrons of psychological research have begun to show are signs which it will be worth while to heed. If the Michigan study failed to select good clinical psychologists and if the enormous labors of the Armed Forces have only slightly reduced the errors involved in picking recruits who are likely to find themselves in grave emotional difficulty, perhaps it is because what we know in our hearts to be necessary conceptual complications have not been courageously followed through in the building of tests.

REFERENCES

1. Skinner, B. F. A case history in scientific method. *Amer. Psychol.*, 1956, 11, 221-233.
2. Stern, G. G., Stein, M. I. and Bloom, B. S. *Methods in Personality Assessment.* Glencoe, Illinois: The Free Press, 1956.
3. Thompson, W. R. Conference of the Milbank Memorial Fund. Oct. 31 - Nov. 1, 1956.
4. von Bertalanffy, L. *Problems of Life.* London: Watts, 1952.

# Changing Methods of Appraising Personality

MORRIS KRUGMAN

In his volume on personality, Gardner Murphy places considerable emphasis on autism as an aspect of personality. In this brief discussion of changing methods of appraising personality, the treatment will be purely autistic. I shall attempt to trace crudely the development of methods of appraisal as seen through the eyes of one psychologist over the past thirty years.

In 1926 psychologists were generally more certain of their diagnoses, and had more confidence in their measuring instruments than they have today. In retrospect, they had quite a few instruments at their disposal, and took them seriously. The Stanford-Binet, though less than ten years old, was almost universally the backbone of the psychological examination, although there were some centers where the Goddard or Herring or Yerkes revision of the Binet was preferred. The IQ obtained by these tests was considered definitive and was accepted as the most important index not only of intelligence, but of personality. A performance battery like the Pintner-Paterson or the Arthur Point Scale was considered essential not only to supply a measure of non-verbal intelligence, but as a vehicle for observation of behavior and personality traits such as volition, perseverance, suggestibility, self-confidence, planfulness and others. Clues to neurotic traits like anxiety, obsessive-compulsive tendencies and conflict were sought, particularly in the performance tests; likewise, bizarre behavior, autism, unrealistic approaches, hallucinatory material, confusion and other indicators of psychosis, as well as observations on irregular functioning, motor incoordination, gaps in function, and other indicators of organicity. Psychologists were beginning to employ projective approaches by adding qualitative approaches to objective measurement, although the term "projective methods" was not coined by Lawrence Frank until 1939.

Achievement tests, diagnostic tests, aptitude tests, and a great variety of special tests were available to the psychologist thirty years ago. Personality tests were in existence, although still used sparingly. In vogue at the time were the Pressey X-O and the Downey Will Temperament Tests. The Pressey was one of the forerunners of the personality questionnaire in which "like" and "dislike" were used, while the Downey was a projective test without the name, using normal

handwriting and handwriting under stress to judge probable behavior under stress. A much older technique employing a stress situation was the mirror-drawing test, which was widely used at the time. Personality inventories were also in common use. These generally were of the bipolar type, purporting to measure such traits as introversion-extraversion or ascendance-submission. Pictures, cloud pictures and inkblots were available for the study of specific traits rather than for total personality appraisal. The Porteus Test was used to measure planfulness, the Knox Cube, immaturity, the Kohs Block Test for organicity and frustration tolerance, and a great many others for specific purposes. The Rorschach was in use in Switzerland and Austria by psychiatrists, and was known in this country only to a few until the publication of the blind analysis by Oberholzer, and the articles and dissertation by Beck. Rating scales had a tremendous vogue for personality appraisal. Attitude scales were also widely used. All in all, the psychologist of thirty years ago had quite a battery of psychological instruments to use.

The psychologist of 1926 was generally employed in a mental hospital, in an institution for mental defectives, in a psychological clinic connected with a university, or in a child guidance clinic. A few served in school systems, mainly in connection with the education of the mentally retarded. When a psychologist functioned in a research department of one of the institutions listed, he often was in a position to do independent creative work, and many of the developments in clinical psychology came from this source. However, when he was engaged as a clinical psychologist, he was usually expected to do mental testing only, except in the newly developing child guidance clinics. Only when psychologists were well established were they permitted to delve into the mysteries of personality. Psychiatrists were, as a rule, less dynamic thirty years ago than they are today, and psychotherapy, even for psychiatrists, was limited to the private practice of psychoanalysts or to experimentation in some of the more advanced child guidance clinics. Most psychiatrists limited their function to diagnosis, not only of personality deviations, but of intellect. Psychiatric textbooks of the time usually contained a chapter on the measurement of intelligence, which consisted of a translation from the French of the 1905 or 1908 Binet-Simon. Twenty-five or thirty years ago, psychologists fought for the privilege of doing mental testing. Today it is almost impossible to employ a psychologist who is willing to perform diagnostic services; everyone, including the most recent graduate, seems interested only in therapy. This is not stated in criticism, or in judgment, but merely as a fact to indicate the tremendous change that has occurred in the status of the psychologist.

As psychiatrists shifted from mere classification of mental disease to the more dynamic Freudian psychology, they abandoned the field of mental testing to the psychologist, and, as psychologists improved their training and developed new approaches that commanded respect, psychiatrists and psychologists shared responsibility for personality appraisal. The concept of the clinic team and orthopsychiatry was developing. Even the Rorschach, which was originally a psychiatric tool, was being taken over almost entirely by psychologists. In the meantime, psychologists were experimenting with innumerable techniques and devices in the hope of discovering an all-purpose instrument for appraising personality. The psychological literature of the past twenty-five years is replete with preposterous claims, and many instruments had brief vogue, only to be replaced by others that ultimately proved no better. In the main, efforts at devising methods of personality appraisal were in two different directions, each with several subdivisions: one was the experimental-statistical approach, with emphasis on standardization and validation by one of several methods, while the other was the intuitive approach, with heavy reliance on clinical intuition, using unstructured or partially structured situations as stimuli for reaction. Under the first, we saw the development of hundreds of different personality inventories, trait inventories, preference questionnaires, attitude scales, and self-rating scales. Except for the very early ones, the tendency was to standardize psychiatric concepts in simplified statements, with simple administration and scoring, and, usually, with group administration possible. The theory back of these questionnaires was acceptable, except for the assumption that the questions were truthfully answered. These instruments were to be so objective that scoring and interpretation could be automatic. The major difficulty with them was that, in spite of good statistical and surface validity, most of them showed little clinical or life validity. Nevertheless, these instruments had considerable usefulness for the sensitive clinician who employed them qualitatively rather than quantitatively, in conjunction with other techniques, particularly with the interview, the case study, and observation. On the other hand, these instruments did a great deal of harm because of their widespread use by unqualified persons in group situations who often made serious unwarranted diagnoses on the basis of numerical scores.

In the second category, that of the partially structured instrument, later to be called "projective," many clinical psychologists felt that at last they had unlocked the door to the mystery of personality appraisal. Inkblots, cloud pictures, picture interpretation, incomplete drawings, figure and other drawings, painting and fingerpainting, sentence completion, story completion, picture completion, mosaics, situational tests,

handwriting, expressive movements, and a great many others developed with bewildering rapidity in the decade from 1930 to 1940. Seldom did test constructors bother with validity, except to indicate that "clinical experience" proved the validity of the instrument under discussion, and that further evidence would be provided in the vague future. The inevitable happened and reaction set in when workers began to test the validity of these instruments, and found most of them weak. In many cases, the reaction was much more violent than was warranted, since the criteria for judging the projective instrument were those of the personality inventory intended for use by relatively untrained workers rather than those of complex clinical instruments to be used only by highly trained clinical psychologists. After twenty or more years of violent debate, psychologists are still sharply divided into two camps on the question of personality appraisal: the objectivists, who clamor for statistical validity, and who consider other clinical approaches unscientific and soft headed, and the subjectivists, who consider statistical approaches undynamic, and emphasize the complexity of human behavior and the need for intensive training and clinical insight for personality appraisal. I have no convincing evidence for this, but my impression is that the majority of clinical psychologists are in the latter category.

I have been doing little more than reminiscing, and at this point it occurred to me to check my reactions against the opinions of accepted authorities in the field. In Buros' Fourth Mental Measurements Yearbook (1953), 308 reviewers reviewed 793 tests, grouped in 13 categories. The categories with the largest number of tests reviewed are those of "character and personality" and "vocations," each with 121 tests. Seventy-two fairly well known psychologists reviewed the 75 non-projective tests and 16 projective instruments in the section on tests of character and personality, and in scanning this large section of the volume, the impression is gathered that the majority of reviewers found both the projective and the non-projective tests seriously lacking in validity. Here are samples of many similar sentiments from reviews of different tests by different reviewers in the Mental Measurements Yearbook. Most of them are from the summary statements of the reviews. The first group of quotations is from the section on non-projective tests:

> "There is no indication in the literature that it can be depended upon to pick out of a group all maladjusted individuals . . . Its greatest usefulness would appear to be as a guide to interview and as an aid to better understanding of the individual."
>
> ". . . does not give the impression of being an instrument that

can measure objectivity in children, nor are the author's claims that it does so substantiated by the evidence presented . . ."

". . . the inventory does not fulfill the reviewer's specifications for a useful diagnostic aid in counseling, although this is the expressed purpose of the inventory."

"This technique continues to be best characterized as an experimental device rather than a test method."

"In my experience, questionnaires are not readily suitable instruments for use in industry."

"Any use of (personality) questionnaires in research altogether is of doubtful value; I do not feel that this particular questionnaire is an exception to the rule."

"As to differentiating persons with serious personality difficulties from the normal population, it appears to be, like other inventories of its kind, not very valid for the individual case."

"Where the individual is concerned, it should be used as an adjunct to the interview, not as a substitute for it."

"The illustrations given of the use of the inventory in educational practice do not convince this reviewer that this test has a useful function in general education."

"In sum: . . . . . Inventory, while as good as most of the similar tests of this type, seems to have only a screening usefulness; and it is depressing to see our counselors being encouraged to use any instrument of this sort in dealing with individual counselees who have gone beyond a screening process. All contemporary personality inventories, including this one, have doubtful individual diagnostic value."

"It is difficult to know the extent to which this type of questionnaire can be used because there is no evidence presented that the alleged areas measured have any meaningful existence, or, if they do, that they can be measured by questionnaires of this type. The path of research in personality measurement is much more thorny and difficult than is suggested by the authors of inventories and questionnaires; and, until they face more realistically the problems of taxonomy and validity, their products will be of doubtful scientific value."

"The test would appear to be of greatest value when utilized by an experienced clinician with emphasis upon the qualitative aspects of the test results."

"The general advantage of the inventory seems to be based on the future promise of the general factorial approach. Regardless

of the general promise of this approach, data are not yet available showing the practical utility of this inventory."

"This test must be considered to be of research interest only."

"The inventory will probably be used by guidance officers in . . . schools where psychologically trained personnel are not available and where there is a basic philosophy of adjustment by conformity."

"There is no evidence that the test can be used for the prediction of human behavior."

"Aside from the fact that . . . is no worse than most conventional personality inventories, there is little that can be said in its favor."

"At this writing, the validity data are too sparse to warrant drawing general conclusions with regard to the applicability and usefulness of the test."

"This is just another one of those toss-a-circle-around-the-symbol-for-your-problem inventories which holds no promise of contributing any more to our understanding of pupils' problems than the scores of others which represent the lowest form of the testmaker's art . . . It is hoped that we will soon pass that naive stage in educational and psychological thought in which we expect to get at important problems by a hurried mass approach. Perhaps these inventories will find their best use in bonfires celebrating our emergence from the ruts that the personality and adjustment testers, ably abetted by high pressure salesmen, have carved out for us. *Caveat emptor.*"

These quotations from test reviews are typical, and while there are other neutral and even positive reviews, the majority opinion of reviewers seems to be exemplified by these statements.

The projective techniques fare no better. To present only a few quotations, by way of example:

"At the present time the test is sadly in need of supporting validational studies, and until these appear its diagnostic employment should be more experimental than routine."

"The author wisely suggests that it be used only by clinical psychologists with general competence in other projective tests and with a broad knowledge of psychoanalytic theory."

". . . the extent to which a protocol is actually representative of the subject's psychosexual development, has not as yet been demonstrated. This problem, however, exists in relation to all projective tests."

"At its present stage of development the (test) is a very promising tool for the experimental study of children's attitudes and personality development and for the clinician who is ready to use the results of this test . . . to build up a clinically shrewd and (hopefully) consistent concept of a personality; it is not ready for those who wish quantitative or even qualitative guideposts to normative status or to differential diagnosis."

"It may be well to point out, however, that, since it is a projective technique, its value in this field is directly proportional to the ability of the individual attempting its interpretation."

"In line with a tradition stronger in European psychology than in the United States, he seems to imply that emphasis on interpreter skill is incompatible with emphasis on objective norms. Understanding, intuition, and internal consistency are sought, rather than objective evidences of validity. Thus, though a skilled examiner may find the test very useful, we have no scientific indications of where its limitations and greatest usefulness lie."

". . . it seems worthy of wide experimental trial in the hands of persons with the requisite interpretational skills."

". . . original manual for the (test) is perhaps one of the worst horrors ever perpetrated in the field of clinical psychology."

"There is no doubt that the test has possibilities as a projective technique; but as the author implies in his manual, it needs much more work before it can be considered a proved clinical instrument."

"Its interpretive scheme requires theoretical assumptions which themselves need to be tested at a much more basic level. This reviewer does not feel that the assumptions are invalid, but that, since they are assumptions, they do not justify the formal structure of the scoring system here proposed by the author of the test."

"In view of the lack of reliability and validity data, the use of the test as a clinical instrument at this time might be premature."

"On a research basis, its use should be strongly encouraged. As a test, it is still essentially unproven."

"This contribution to . . . projective techniques will be received enthusiastically by those to whom the theoretical assumptions of dynamic psychology are congenial, and examined with

suspicion, probably, by those who look for statistical or other quantitative evidence of validity for a test of personality."

"One may hope that the future of this promising technique will be marked, not by slavish adherence to formulas, but by sound research to determine its validity—and its limitations."

"Whether the test has any merit as a personality test cannot be assessed since the only evidence of validity given is the relatively low correlation with a test which can hardly be considered a criterion."

"(This) test is a clinical technique, not a psychometric method. As such it has the advantages, and is subject to the limitations, of other complex, flexible clinical tools. In its present stage of development it is an aid to psychological investigation and interpretation, the usefulness of which depends upon the clinician who applies it."

"At present it appears to the reviewer that the clinical use of both forms of the (test), except in relation to other tests or case history material, is hazardous."

"Pending more positive statistical and clinical validating findings, it will have to continue its precarious position in the field of projective techniques."

"Those who look for validation data will see only large claims and little evidence."

"Pending the publication of adequate supporting evidence, the psychologist using this method is proceeding on the basis of faith and should be willing to recognize the perils of his course."

"Indeed, even the friendly observer cannot escape the impression that in many instances the utilization of this rather tedious and time-consuming technique serves no purpose that could not be accomplished more efficiently by personal interview."

"Any step of true research in the projective field is encouraging, . . . but much of the path of progress lies ahead."

"Here is another of the all too numerous family of prematurely published, invalidated projective tests."

It seems no exaggeration to state that, in spite of thirty years of extensive experimentation and in spite of a vast array of instruments, projective and non-projective, that are available today, we do not now have instruments for personality appraisal that are generally acceptable to psychologists. There seems to be a growing belief among many psychologists that we have, perhaps, been on the wrong track in seeking such instruments; that personality is so complex, and personality factors, at best, so unstable, changing rapidly from situation to situation, that

simple approaches to personality appraisal can yield little that is valid. Much of the experimentation in personality appraisal has been empirical, ignoring to a great extent the great development, over the past 25 years, in theory of personality growth and development, and in the basic concepts of psychology, psychiatry, sociology and anthropology. There are, however, many encouraging signs to indicate that we are emerging from the empirical stage, and coming closer to the utilization of current theory and knowledge about personality, in our approaches to understanding and appraising personality and behavior. I shall close by merely listing a few of these trends, without presuming to guess what the future holds in store for us.

1. Psychologists are much more critical than they were, and do not take on faith everything that is published purporting to measure personality. They are infinitely better trained in fields related to personality development than ever before, so that their judgments are more trustworthy.
2. Opportunities for using this training are greater than ever since psychologists have demonstrated the contributions they can make in diagnosis and therapy.
3. Concepts of personality and behavior are more dynamic than heretofore, and are more acceptable to psychologists of all shades of opinion. More and more are psychologists deserting the hope for simple approaches to personality appraisal. Even the fantastically elaborate appraisal programs of the Office of Strategic Services and the V A program for appraising clinical psychologists proved disappointing. How, then, can a single inventory or a projective test serve the purpose?
4. In spite of the rash of new instruments for personality appraisal over the past twenty years, there is now a tendency toward consolidation, critical review, and research with existing techniques, using dynamic theory as a basis, rather than haphazard production of new tests.
5. While experimental psychologists are becoming more dynamic, clinical psychologists are becoming more research minded.
6. In some recent publications on personality appraisal, the trend is away from attempts at omnibus or global measurement of personality, and toward a very specific look at a limited sample of behavior in a specific situation. Gardner Murphy warns us that the study of the situation in which an individual functions is as important as the study of inner drives.

7. Concepts about the dimensions of personality are undergoing radical changes even among statisticians. We are beginning to see objections raised to the unidimensional profile of personality presentation, and tendencies in favor of multi-dimensional approaches.
8. More and more do we find criticism of all types of instruments for independent use for personality appraisal. When such instruments are used, they are used to supplement rather than replace interview, observational, and other techniques.
9. More and more do we find recommendation of the case study and observational techniques by qualified psychologists rather than reliance on test scores or formulae for personality appraisal. (In a very recent volume the "anamnestic study" was recommended as a new approach to personality study. Twenty-five years ago the "anamnestic study," commonly used in mental hospitals, was discarded as undynamic, and the social history substituted.)
10. More and more do we find multi-discipline approaches to the study of personality. Simple approaches are becoming suspect.

This, then, is what it looks like to one psychologist scanning the thirty year period. We seem to have come the full circle, beginning with the psychologist using available devices as tools rather than goals, going through long periods of search for simple objective measures of personality, emerging recently with much more mature, sophisticated understanding of the complexity of personality and its interaction with complex and fluid situations, again apparently ready to use the best available instruments as tools rather than goals. There is, however, a tremendous difference between the psychologist of today and that of 1926. The modern psychologist has a keen appreciation of his own limitations as an observer of human behavior, and the limitations of the instruments he uses. He does not rely unduly on personality appraisal instruments, although he uses them constantly. He knows there are no shortcuts, no royal roads, no panaceas—at least not at this moment. He uses everything he can find—objective, projective, subjective, to meet the needs of the moment, searching for better ways, but he is not satisfied with what he has available.

# Testing—Then and Now

BEN D. WOOD

The most visible difference between "then" and "now" is represented by the size of this Invitational Conference as compared with its size when its first annual meetings began about a quarter of a century ago. In those times we were not so numerous; nor were we opulent enough for such a sybaritic feast as we have just enjoyed: in those days the more fortunate lunched in the Coffee Shop, while the rest of us patronized a drug store or a lunch wagon on a side street. I am curiously reminded of what Will Rogers said to an unexpectedly large crowd that greeted him at a southern airport during the great depression that marked the first years of the Cooperative Test Service, of the Educational Records Bureau, of the Pennsylvania Study, and of this organization. Looking over the vast crowd that had gathered to greet him, he said, "I had no idea unemployment was so bad in this area." The size of this gathering is a highly deserved tribute to the leadership of Educational Testing Service.

Perhaps a brief digression on the origin of this conference would be in order, especially as there may be a lesson in it which we should take to heart. When the joint conferences of the Educational Records Bureau, Cooperative Test Service, Progressive Education Association, American Council on Education, etc., began, our programs were liberally infested with testing leaders and test technicians; and during the discussion periods, there was a slight tendency for the technicians to display their technical erudition in arguments that were usually polite and often too long. The schoolmen who were the bulk of our audiences quite frankly admitted they did not understand the technical lingo; and only politeness led them to conceal their suspicion that the technicians didn't either. Some disputations that were heard in those days indicated that their ill-concealed suspicions were not without foundation. Be that as it may, our technical vocabulary put them to sleep or drove them to the bar for transfusions of one sort or another. Thus fortified, they descended on the Program Committee and let it be known that such nonsense had to be eliminated or reduced, or else.

Well, of course, when your paying clients show such inadequate appreciation of pearls of wisdom, you have to remove the pearls from their programs. But still the Program Committee was convinced that such pearls of wisdom were desirable; indeed, were necessary for progress

in the testing movement whose leaders at that stage were mostly working in isolation, without any national meetings specifically designed for their interests. Therefore, the Program Committee kept its technical speakers, but scheduled them on Saturday after the schoolmen had returned to their schools.

At first we were deeply concerned about the problem of finding an audience for these gifted speakers. With only a dozen or so showing up for the first few Saturday sessions, we were afraid our speakers would quit. But the dedication of our speakers made our anxiety seem childish—they came back year after year with longer speeches and constantly increasing technical vocabularies. One of our elder schoolmen who witnessed this phenomenon said it reminded him of the character in a Bernard Shaw play who explained that he had to preach even though he had none to hear him and nothing to say.

But our speakers did have something to say; what they said became increasingly significant every year. The size of this audience today is ample vindication of the plan adopted in the early 1930's. When the success of the Saturday conference was pointed out to one of our elder schoolmen, he not only offered congratulations, but expressed regrets at the necessity for reducing the number of technical speeches in the joint conference programs that have always been scheduled on Thursday and Friday. His expression of regrets was sincere, but understandably not quite as effusive as his expressions of congratulations.

Another difference between "then" and "now" is that certain criticisms against objective tests then frequent are now not so often heard, at least not from the same sources. In the early 1920's it was a serious problem to get anybody to use objective tests at all, even experimentally.

In those years there was a widespread reaction against the Army Alpha and its vast brood, not only because of the limitations of those early tests but even more, I think, because of their extravagant claims and their misuse, which often included trying to do more accurately what ought not to be done at all in schools dedicated to the conservation of human resources and to the general welfare, as against blind adherence to rigid curricula and punitive doctrines of discipline. This partly laudable and largely temporary reaction was, however, only a minor incident compared to a type of antagonism based on almost instinctive opposition to anything new, and an equally tropistic faith in and allegiance to the old and familiar. Allergy to new ideas is an almost universal human disease. If I fail to include any palinodes here, it is not because I have been immune to this endemic malady, but because I wish to stay within the time and space limits assigned by the chairman.

Typical of the objections to objective tests were those made when a

proposal to use "new type" placement tests in Columbia College was being discussed about 35 years ago. A French vocabulary test was declared unacceptable mainly because the examinees would "beat" the test by the simple trick of memorizing 1,000 or 2,000 words. I was urging the experimental use of the proposed placement test because I thought it had some important potential values; but I had to confess that in my wildest flights of confidence, I had never dreamed that our innocent little vocabulary test could ever produce such a miracle. Unfortunately, the fears of the critics were found to be groundless: after 30 years of using such vocabulary tests, teachers are still lamenting the inadequate vocabularies of their students—and their lamentations are amply justified by the results of tests, both old and new type.

These and similar objections to what we then called "new-type" tests were nearly always accompanied by enthusiastic approval of the essay type of examinations. Some of the expressions of approval were quite unrestrained: the essay examinations "forced pupils to think creatively, to organize their thoughts logically, to marshall their arguments coherently, and express themselves persuasively." Some of the panegyrics were more cautious: the essay examinations "gave opportunity to think, to create, to organize, to be logical, to write coherently, persuasively and artistically" and thus "positively invited" the pupils, one and all, to indulge in these unquestionably desirable, but to some pupils slightly vague if not esoteric blessings. These claims were made at a time when complaints were widespread, *in* as well as *out* of academic circles, that the writing of many college graduates was at best semi-literate.

These complaints were accepted as amply justified after centuries of exclusive use of essay-type examinations throughout the schools and colleges of the world. The essay-type examination undoubtedly affords opportunity and offers invitation; but its sometimes alleged power to *force* creative thinking, effective writing, etc., is too often an example of an irresistible force coming into contact with an immovable body. Lest I be misunderstood, I hasten to interject that I have never knowingly advocated the elimination of the essay examination, which some formerly fervent protagonists have apparently abandoned. I have tried to maintain the values as well as the limitations of all kinds of tests and of other sources of information for guidance purposes. In general, I have tried to emphasize the importance of distinguishing the measurement and instructional functions of tests, examinations, and other devices. Both functions are necessary, and neither should or need be sacrificed to the other. Since examinations have traditionally been used for competitive and often invidious purposes, and are still thought of too exclu-

sively in relation to such purposes, it would be ideal if instruments that produce the best measurements could also serve as the most effective teaching devices. Although substantial progress has been made, and continues to be made, in producing objective types of tests that measure elements and combinations of elements that more nearly approximate some of the important goals of instruction, fortunately no one is as yet satisfied, so far as I know.

But that the progress thus far made is decisively important cannot be questioned. While I have always felt that the bad effects of objective questions and the good effects of essay examinations on learning were somewhat exaggerated, there was a time when some of our wisest educators and leading experts in the field of measurement were so fearful of the evil influences of objective tests that no reconciliation seemed possible. It is gratifying to note that the improvement in test production has been great enough to permit almost exclusive use of the objective types of tests, especially those that can be scored by machine.

This mention of the test scoring machine, for the development of which we are indebted to the educational interest and multiple genius of the late lamented Mr. Thomas J. Watson, and to the great corporation which he created, recalls another difference between "then" and "now" which involves some vivid and amusing memories for some of us who were associated with that development.

From the beginning of my work in the testing field, my chief interest has always been in the long-term systematic and constructive use of tests and other sources of information about the abilities, interests and needs of pupils as GROWING individuals, rather than in one-shot efforts to separate the sheep from the goats at any one point on the educational ladder. I have always supported college entrance, scholarship, and single-shot competitive examinations as necessary and desirable; in fact, I have helped construct and administer some tests of that type; but I never considered them sufficient for constructive educational and conservation purposes. From the beginning I felt that systematic testing and guidance should be an integral part of the whole school procedure, from kindergarten through university and on into appropriate gainful employment in the work of the world. The inscription over the door to Dr. Colton's famous museum in Flagstaff, Arizona, says: "This museum exhibits, not things, but ideas." My efforts from the beginning have been directed at selling, not tests, but the idea and practice of individualizing education by continuous guidance based on cumulative records of all kinds of relevant information about individuals, including scores from tests that are as nearly comparable as they can be made. Selling this idea, which is at least as old as Plato,

was and still is a formidable task, requiring nothing less than a fundamental reorientation of the masses of our teachers and school administrators.

Our early experience with placement tests in Columbia College, and with similar tests in the Regents Experiment with New-Type Tests, suggested that even with the economies and other advantages of objective tests, and even if by some miracle all schools could be persuaded, univeral testing and guidance of the type that obsessed me as necessary would be prohibitive in cost. A bit later, our experience with the Eastman Teaching Film Experiment, the Classroom Typewriter Experiment, and especially the Carnegie Foundation Pennsylvania Study, the Educational Records Bureau and the Cooperative Test Service, convinced me that the cost and errors of scoring tests and processing their results had to be drastically reduced in order to make their values available to all children in all schools.

The notion supervened that IBM tabulators could be adapted to statistical procedures. By Mr. Watson's generosity and faith we secured the loan of a large installation of tabulators and related equipment. Several time- and money-saving adaptations of existing tabulators were made by Warren and Mendenhall in a few months, and within a year they, in collaboration with IBM engineers, designed and the company produced the first statistical tabulator. A year or two ago, Mr. Watson referred to this machine as the first of IBM's mechanical "brains." Two or three years later, IBM engineers mechanized Rulon's method for scoring Strong Interest Blanks on the standard tabulator. But from the beginning, our greatest efforts were devoted to developing a test scoring machine. The first machine, using the electric-eye sensing method, turned out to be too slow as well as too expensive. The solution finally came through the brilliant inventiveness of Mr. Reynold B. Johnson, an IBM engineer who is now head of the IBM Research Laboratory in San Jose, California.

One of the more amusing incidents following the announcement of the test scoring machine was the rumor, which I was never able to verify, that some highly placed professional educators had said we would mechanize and ruin education with that machine.

The addition of the item analysis unit to the scoring machine, which I am told was the first analogue computer ever built anywhere for any purpose, and which was intended to reduce the drudgery, cost and errors of scoring objective tests, widened its scope to test construction; and the addition of the Thurstone-Tucker Matrix Multiplier unit made it into a research instrument of substantial importance.

This machine came along just in time to make substantial contributions to the testing work in our armed forces during World War II.

In addition to the alleged fears regarding the danger of mechanizing education, there were some other unfavorable criticisms, but in recent years the only criticisms I have heard are that the machine is not fast enough nor versatile enough to meet the needs of the increasingly large objective testing programs in our country. I am confident that the existing and still greater needs of the future will be amply cared for by two types of machines. The first is Dr. Lindquist's giant scoring and processing machine, which represents an achievement as great as Dr. Lindquist's notable contributions to the art of writing test items that tap abilities more complex than memory and single-step reasoning. The second is the forthcoming new model of the IBM test scoring machine, which can be used in combination with one or more of the new IBM electronic data processing machines.

It would probably not be expedient to speculate too deeply on the real reasons for the currently more favorable, or *apparently* more favorable, attitude towards objective tests and their mechanical or electronic processing. As you may recall, the General Motors diesel-electric locomotive was not received with unanimous approval. Some of the major railroads questioned the adequacy of its drawbar pull, until they observed that it was great enough to pull 37 railroads out of bankruptcy in three years; then under irresistible financial motivation they swamped General Motors with orders. I can only express the hope, timidly, that pecuniary motives are not the only reason for the apparently more favorable attitude toward objective tests and scoring machines than obtained 20 to 30 years ago.

One of the distinguished speakers at the conference yesterday reminded us that Mark May once expressed the wish that the IQ had never been born, because it had been so widely misused. Thirty years ago I heard much the same kind of wish about objective tests in general. It would be more logical, if one insists on making drastic wishes, to wish that those guilty of misusing test results had never been born. But since the shortage of teachers is already alarming, such wishful depopulation should be avoided. It would be more constructive as well as more logical to try to provide more adequate education on the interpretation and use of all kinds of guidance data, with special emphasis on the limitations as well as the values of each kind of data.

In addition to the limitations and ambiguities inherent in even the best existing tests, the testing movement has, in the opinion of some wise and sympathetic critics, lost effective contact with the teachers who, in last analysis, are the agents through whom the potential values

of test data and other guidance information must reach the individual pupils. My distinguished colleague, Dr. Lorge, has repeatedly called attention to this gap, and has warned against the possibility that technological advances may tend to widen the gap by concentrating all testing matters in the hands of specialized experts who are necessarily remote from classroom situations, or who at least appear to the teachers to be remote.

Since preparing this paper, I was highly gratified to learn that Educational Testing Service has launched a project aimed at enlarging the classroom teacher's role in student guidance.

Many statewide testing programs, as well as several programs made available on a national basis, are helping to offset this danger. Some of the earliest of the state testing programs, such as those in Iowa, Minnesota, South Carolina, Connecticut, Virginia, New York, Ohio, Wisconsin, Florida, Kentucky, and Texas, have regularly been represented in this conference. Representatives of more recent statewide programs, such as those of Illinois, Michigan, and Tennessee, also contribute their ideas to these meetings. These programs have always tried to foster close relationships between test specialists and the teachers and guidance workers in the schools.

As a concrete example of the kind of continuous testing in guidance programs that has always been my chief professional interest, I would like to give you a brief sketch of the history, operations, and support of the Educational Records Bureau. I choose this example because I am more familiar with it than with others which illustrate the same emphases.

The Bureau was started during the school year 1927-28 by a small group of fifty schools or less which were interested in obtaining objective records on their pupils from year to year. It had a small grant from the Keith Fund which enabled it to plan a limited testing program based on a budget of some twenty thousand dollars. It had no reserve fund, and its financial outlook was not promising, even in the eyes of its most ardent supporters. But within eight years, all of them depression years, this struggling little organization had grown to a membership of 300 schools, with an annual budget of $65,000 and a reserve fund of about $40,000. Within another ten years, by 1946-47, the Bureau had, notwithstanding the set-back occasioned by World War II, expanded to a membership of 440 schools, with an annual budget of $150,000 and a reserve fund of approximately $130,000. Nine years later, at the end of the school year 1955-56, the Bureau had a membership of 650 schools, an annual budget of $270,000, and a reserve fund of around $260,000.

In 1927-28, the tests scored by the Bureau for its independent school

members, which could then serve as a basis of norms for independent school, or college preparatory, pupils totaled just 8,900. Ten years later, in 1937-38, a total of 83,400 tests were available for independent school norms; in 1947-48, the number was 112,500. Last year, the results of 153,000 tests were available for use in the norms for the independent school group. This number was approximately half the total number of tests scored by the Bureau for all kinds of institutions.

What are the reasons for the growth of the Educational Records Bureau, without any attempt to sell its services, and for the growth of other similar test service organizations? It seems to me that the basis of this development is found in three procedures which have constantly been emphasized and increasingly put into practice.

The first of these is continuous testing of every pupil and measurement of growth in terms understandable to teachers and counselors; and analytical reports to the schools, including detailed interpretations of their test results, not only for class groups but for many of their individual pupils.

The second is the use of comprehensive cumulative records, not of test results alone, but records arranged in such a way that test results are related to, and interpreted in the light of, much other information about individual pupils.

The third of these is continuous research on the values of the tests, and publication of the results, with the use of as few technical terms as possible, so that the schools would be apprised of both the values and the weaknesses in the instruments used, and would be stimulated to ask the test publishers for improved measures.

These three emphases, under Dr. Traxler's dynamic leadership, mainly account for the gratifying growth of the Educational Records Bureau.

It is a pleasant duty to pay homage to those who are not only advocating the production of better tests but are unceasingly slaving at the most difficult of all testing chores, that is, the writing of test items that with increasing accuracy and fullness measure the dimensions of intelligence, learning, and personality that we need to have measured for the greater conservation of human resources and the greater happiness of all individuals during and after school days. Those whose unremitting efforts are bringing us closer to this goal are, in my view, the real heroes of the testing movement. Although I still occasionally try my hand at this basic task, recently even to the point of playing with palindromes for a projected language test, I confess that for many years I have quite fallen away from this onerous type of work. My own fall is quite unimportant, but I do regret to find that, in devoting myself to softer

kinds of jobs, I am still not without good company from the ranks of former test makers!

While offering homage to the many devoted workers in various programs and in various institutions, it is a particular pleasure to congratulate President Chauncey and his staff in Educational Testing Service on their leadership, on the high professional level at which their test construction is being done, and on their broad research program, which amply justify the high hopes of the architects, founders and supporters of Educational Testing Service.

My concluding remarks may be epitomized by reminding you that the cumulative educational record form which we prepared for the American Council on Education in 1928 allocated more than three-fourths of its space to information other than test results. This proportion in the allocation of space to records of behavior other than on paper-and-pencil tests still seems reasonable to me. I have neither the time nor the talent to discuss Ludwig Bolk's conclusion that in the process of human evolution Man had been arrested at the level of growth of the embryo of an ape, or the more recent argument of Lucien Cuenot that Man is "an example of neoteny in a chimpanzee or gorilla," or the still more recent opinion of Professor Carleton S. Coon that "The Neolithic mentality is still with us in most of the world, a social coelacanth on the beach of time and an impediment to global peace and unity." But we do not need the authority of these gifted students of anthropology and evolution to convince us of the critical importance of devoting more time and effort to the study of human behavior. Since one of the most productive workers in our field in now heading a project in the behavioral sciences, I would like to express the confident hope, which I am sure we all share, that Ralph Tyler and other students of human behavior will forge ahead far enough to improve Man's basic education, and fast enough for Man's survival through the age of the hydrogen bomb.

# DISCUSSION

*Participants*

Robert G. Bernreuter, Richard Hill Byrne, S. S. Dunn, Hanna F. Faterson, Warren G. Findley, John C. Flanagan, E. Z. Friedenberg, Newman L. Hoopingarner, Morris Krugman, Irving Lorge, James L. MacKay, Gardner Murphy, Edward Silverberg.

DR. MacKAY: It looks to me as though Dr. Murphy was just bursting with some things to say during that last paper. I wonder if he was.

CHAIRMAN LORGE: They are doing personality appraisals out there in the audience.

DR. MURPHY: I think Dr. MacKay saw me taking notes. I can throw this either to you or Dr. Krugman, depending on the mood. I wonder if I was more optimistic or less nihilistic about the actual state of personality testing.

It might be interesting to see what you think of the realistic possibilities of four types of things I put down. First, what do we think about following one projective test systematically with procedural variations within a population with new types of cross-checks year by year, the way McClelland has used the TAT? Should we say this is simply a way of testing a test, or should we conclude that something about the basic validity of our personality theory is obtainable from this sort of thing?

Second, what do we think about the attempts to develop projective tests into a more standardized and objective form, like the attempt of Wayne Holtzman today that we make the Rorschach into a systematic and pretty objective test? Third, what do we think about the success of Witkin and his collaborators in systematic comparison of experimental procedures with projective procedures? I have particularly in mind your life situation of children in relation to field dependence.

Fourth, what do we think of straight laboratory procedures, like George Klein's studies, as ways of getting at personality dimensions under tightly systematic safeguarding of experimental procedures?

Would you be willing to take a minute or two and talk on one of those?

DR. KRUGMAN: We are here for the rest of the day. First, I hope I did not appear nihilistic. I was trying to present my own experience and the feeling of many people that the simple approaches to personality do not yield the desired results. I hope I am optimistic about training clinicians, using all the knowledge of the basic sciences and all of their training, using the best method of research, combining the two and not looking for scores or quotients or profiles or simple techniques, but looking at personality and its complexity.

Let's say in this situation as in the study you mentioned earlier, Dr. Murphy, that the teaching situation or the graduate student situation is considered in relation to certain members of the faculty. The question becomes: How do these students react to particular faculty members in particular situations?

In any approach, the follow-through systematically and chronologically is the kind of approach we are all pleading for and hoping for, rather than the panacea to give us a very rapid picture, or give us nothing at all in the ultimate.

CHAIRMAN LORGE: I would like to suggest that some of the questions Dr. Murphy has proposed are questions that some of you ought to discuss. I would like to remind you that F. L. Wells, in his textbook, put a tremendous amount of emphasis on the Army Alpha, not so much on the test, but rather in the course of working with many individuals over many years he developed an understanding and a feel of how Alpha went with certain kinds of broad behaviors. The very fact that the test went with broad behaviors suggested that Alpha was a useful instrument. He found difficulty in communicating his ideas about Alpha to other clinicians.

The problem becomes one of why are such intuitive and valid judgments so difficult to transmit? The test maker has been working over the years trying to find an instrument which will simplify diagnosis in terms of interpretation. I think the line suggested is one that would be of great value. If you worked with the same test with many individuals and made conscientious follow-ups, you would be surprised at how much you could learn.

Are there further questions or discussions?

DR. BERNREUTER: I have recently made a very crude estimate of the number of people who would be required to carry out psychological services that seemed to be demanded by the general populace in America. A very crude estimate indicates practically every person in America with an IQ over 130 would be required in the field of psychology if all the work that needs to be done is to be done by people with Ph.D. degrees. That, to me, is a very sobering fact. I am not at all in sympathy with the concept that says you should not try to develop an instrument that can be used by a person who does not have a Ph.D. degree. I think it is extremely important to us that we do have instruments that can be used by assistants.

I think that when the younger people in our profession are given the impression that it is foolhardy or unwise to work on an instrument that can be used by someone without a Ph.D. degree they are being misled.

I think a disservice is being done to psychology when any young person is discouraged from attempting to work in a field which would result in an instrument that could be used by someone without a Ph.D. degree.

CHAIRMAN LORGE: I have one hope that people like Bob will have taken their own tests systematically at two-year intervals to see what the impact of the external situation is on the internal dynamics.

DR. FRIEDENBERG: I would like to address this to Dr. Murphy, whether he sees any specific promise in the sort of approach that William Stephenson has published and advocated. It seems to me that in that you do have a systematic way of going about two of the specific goals that you mentioned, one, designing the external situation into your experimental study so that you are, in fact, examining an interaction rather than a set of traits, and two, providing a basis for examining your orthogonal effects, so that they just don't crop up out of the data, but in response to hypotheses that are included in your experimental design.

DR. MURPHY: I am very sympathetic and very enthusiastic about this approach. I think it is relatively untried, but it makes a lot of sense in offering an opportunity for clinical people to get together with sophisticated non-clinical people in a procedural way, and work back and forth. They learn a lot from each other. I have seen this work where the statisticians were rather naive clinically and the clinicians naive statistically.

DR. BYRNE: I address this question to Dr. Krugman. I wonder if he has any comment to make about Meehl's recent work in which he says, I think this would be a safe generalization to make, that when it comes to statistical or clinical prediction, the statistical operator can predict personality as well as or better than the clinician. I gather Dr. Krugman was perhaps being negative about the statistical approach through the use of present measurements, and was favoring a greater emphasis on the clinical approach.

I wonder if you would have a comment about Meehl's study.

DR. KRUGMAN: We are not at all sure that the statistical approach gives us a better diagnosis or a better picture of personality. What we are sure of is that, using the criteria established statistically, we seem to get better results if the test user is clinically trained. I say we are just as weak in the clinical field as in the other, and I am not defending it. The point I make is that we have no evidence that in either case we have really been able to appraise personality to the others' satisfaction on a group basis.

Apparently in the clinic situation, when the clinician correlates data, not mathematically but on a case to case basis, with the judgment of the clinician, with the judgment of the life situation, with the total case study, and with prognosis in terms of what happens to a specific individual—of course we are faced with disturbed individuals, too—it seems to stand up a little better than simple instruments, but that isn't saying very much.

Obviously, we have not reached our goals in either category, and much more work needs to be done. I cannot accept the proposition so far that anybody has developed any clinical instrument that gives you complete satisfaction when used with groups.

The Eysenk study of therapeutic procedures, you will recall, proved that it doesn't make much difference whether the therapy is done by a general practitioner, psychologist, psychiatrist or psychoanalyst, the net result seemed to be the same. In research, the clinical approach often fares no better than the statistical.

CHAIRMAN LORGE: The antagonists in the argument "statistical versus the clinical" ought to include in the family the mathematical statisticians. They have not yet given us methods of relating a complex diagnostic indicator to complex criteria, i.e., our criteria are not yet established. Any time that anybody takes a simplex criterion, such as success, adjustment, duration of therapy, and takes a number of variables to relate them to such a simplex, you will find that the statistics are necessarily inadequate because the method is multiple correlation. If, however, you were to ask "what about a complex criterion?" (which is what life is), it isn't simple. This proposes a problem for which we do not have the tools, we seem to lack the requisite methods. That is, we need something that goes beyond our current conceptionalization of mind. This is where Dr. Murphy's presentation is so germane, because it does give us certain indication of direction.

I think in the same terms Dr. Krugman's statement to us is, "where are we and what are the possibilities for going some place else?" I am sure that in this audience there must be people who have thought as Dr. Krugman and Dr. Murphy have, who have more positive things to say than have already been said.

DR. FLANAGAN: In searching for criteria for use in prediction studies in a variety of fields including educational achievement, military and civilian aviation, and children's personal and social development, I have come to the conclusion that Sir Francis Galton's idea proposed nearly eighty years ago is a very sound one. He proposed that we make daily observations on people's behavior, and that these be described

as objectively as possible in written records and tabulations. Through the use of such systematically collected performance records a truly valid and predictable criterion can be obtained. By maintaining these records in a form in which the inferences of the observer are reduced to a minimum we have the basis for developing a set of useful instruments that the younger psychologists can use for prediction studies involving whatever variety of procedure they choose to use whether it be objective, projective, or subjective.

DR. FINDLEY: As I have listened to the speakers, they all seem to be agreed that this matter of appraising and dealing with personality is very complex. I feel that we are now reaching a conclusion that this complexity of appraising personality with tests verges on the order of the complexity of a problem that we haven't done much about, what I would call the area of preventive mental health programs. I wonder whether one of the conclusions might be, not the one that Dr. Bernreuter has suggested, of going on, striving valiantly to make more and better measures than can be used by persons with less facility, but perhaps trying to communicate to a larger and larger body of school administrators, school teachers, parents and others, the rudiments of the ideas about personality that seem to underlie our difficulty in making simple instruments.

I wonder if either of the speakers would care to comment on that as an alternative approach.

DR. KRUGMAN: There you inject quite a different problem. There always will be the problem of diagnosis regardless of how much prevention is done. My comment on that question would be something like this: That in this field of personality appraisal, we are probably now at the stage where physics was before the advent of atomic physics. We thought we knew exactly what matter was; we knew it had its limitations, it was static and stable. Suddenly we are in an upheaval. All our notions are dispelled, and matter isn't at all what it was.

I think right now we are at the opening of an era where we are beginning to understand a little more about what personality is, and from that, I think, will stem better approaches to the measurement of personality.

DR. PATERSON: Going back to a point raised previously, I am one of those soft-headed psychologists who feel that a good deal can be learned from projective techniques, not only about the person with whom one is working, but also about the instrument which one is using.

Dr. Lorge made a comment about Dr. F. L. Wells feeling that even the Army Alpha was an instrument with which you could learn more about persons. It seems to me that one difficulty with many of the

studies on the validity of projective tests which are continually being reported is that such studies so often begin with a mechanical application of traditional scoring schemes and interpretations—*for example, correlating the color score on the Rorschach with ratings of "extraversion," and finding, as might be expected, little or no relationship.*

Nor do I feel that a Ph.D. degree is any guarantee of the level of work performed. I am reminded of a Ph.D. student of Dr. Krugman's who told me "I know all about the Rorschach. I had a course with Dr. Krugman." I am really quite concerned about the attitude of many young people entering the field. Somehow, in the course of their training in clinical work with projective techniques, they develop the attitude that "there are certain things to be learned; I will learn them, and use them." More creative people who have gone more deeply into projective tests, or people who have used objective tests in the way in which Dr. Wells used the Army Alpha, are in a position to look for *new ways* of understanding the material, and thus to put validation studies on a more substantial basis.

In reading published reports of validation studies, I am also often disturbed by the fact that interpretations made by psychologists of very varied levels of competence and insight are pooled together. The results might perhaps be different if psychologists selected for subtlety of understanding and flexibility of approach served as interpreters. If the question of criteria also received serious attention, we would be on the way to better designed validation studies.

DR. HOOPINGARNER: It seems to me that one of the most essential steps to be taken to prove the validity of tests and instruments for checking and evaluating personality should be specific research to establish criteria on what we might call very broadly *success or satisfactory living.* Even though we might have to change our criteria from time to time we would have something that at least was defined as closely as possible, and by checking our various instruments against these criteria we would have some standardization, some real basis for comparison and some basis for judgment as to how valuable the test results are.

As I see it, there has not been very much accomplished in the way of establishing any sort of criterion, or criteria, to judge by. Before we can make any great amount of progress [illegible]ld there not be something definite done to establish criteria for w[illegible] like to call *success*?

DR. DUNN: I am wondering whether some speakers would care to comment on the idea that instead of setting up criterion, we might often accept the test scores as an adequate measure and start to vary the

situation and say that we understand what happens when you vary the situation and reuse the sign test. I think where you make a reading test, you make some sort of change in the reading techniques and then reuse the sign test, and you will get a better understanding of the reading process. Wouldn't it be possible in personality not to talk about some sort of external criterion, but to use the test itself and the shifts in the test and vary the situation so that in that way there is a better understanding?

DR. MURPHY: That is what I had in mind in referring to McClelland's work, which I believe is an example of the suggested procedure.

MR. SILVERBERG: Haven't we reached the stage in the field of personality study, where we could approach these problems more maturely by setting up an organization or foundation, to do the same job for us that similar foundations have accomplished in other areas? This organization could set up many series of studies using cross-discipline and theoretician-practitioner teams. Studies could be assigned arbitrarily to institutions, with assigned goals and periods of duration. Using our own rich human resources, we can then follow these studies over years of testing and practice, and arrive at more valid conclusions than we seem to be arriving at now.

CHAIRMAN LORGE: This is a suggestion that I think I made: a man ought to test himself. When I was naive and young I rated myself on a large number of aptitudes, because I thought I would get a fellowship at Stanford. I never sent the application, but every two years I rated myself. I want to tell you by my self-appraisal I made progress.

The suggestion is reasonable; if we could only follow-up the behavior of youngsters, or oldsters, as the case may be, who have been tested, we may get a good deal of information. The difficulty is that this information is not easily accessible.

I have a set of records on the Strong Interest Blank, from a group of WPA workers, who took the blank a week apart, thirty days apart, a year apart, two years apart, and then WPA went apart. Yet there is something in those records of potential significance, e.g., consistency over time, over the span, which is astonishing. Sixty-six per cent of the items are relatively unchanged over time.

We have the suggestion that Dr. Murphy has made to us, that this is one way, following the same instrument in a longitudinal study in which you ask what happens to the instrument? The other approach is to use the instrument as McClelland does to see what happens to the instrument as a consequence of changes in condition. This is essentially what Skinner is doing. He takes a given method and keeps on working

with it until he can understand it fully. That is Skinner's methodology; if used the way he is using it and the way other people are using it, you will know much more about how people behave in a specific problem.

Are there any other comments for the good of the cause? May I then thank both Dr. Murphy and Dr. Krugman for the wisdom that they have given us.

## GENERAL MEETING

**The Changing Curriculum and Testing**

# The Curriculum—Then and Now

RALPH W. TYLER

Any effort to review the development during the past fifty years of the school curriculum of the United States encounters a confusing complexity. This is not due alone, or even primarily, to the fact that we have no centralized control of education in this country, although it is true that the variations among the forty-eight states are easily marked. These local variations are less pronounced than are the differences of another sort.

One may analyze the development of curriculum theory, that is, the statements of rationale for the curriculum and the related discussions which seek to explain it and to make it more coherent and systematic. One may examine the work of the persons and groups who have designed courses of study and prepared curriculum guides. One may assemble reports of the curriculum in operation in certain schools to obtain a picture of what teachers are actually teaching at a given time in the few schools on which this kind of report is available. These three reviews will give quite different content, yet each is a significant aspect of the American school curriculum.

Another factor in the complexity of this topic arises from the unusually comprehensive definition of the term "curriculum" which is currently employed in American educational circles. The term is not limited in this country, as it commonly is abroad, to refer to the outline of the content to be taught but is used to include all of the learning of students which is planned by and directed by the school to attain its educational goals. This inclusive definition covers the formulation of educational objectives, the planning, use and organization of learning experiences, and the appraisal of student learning. It also includes not only the learning activities carried on in the classroom and laboratory but also those at home or in extra-curriculum situations insofar as these are planned and directed by the school to attain its aims. The line drawn between the curriculum and other activities of the students is that which separates activities designed by the school to contribute to educational ends from those which are provided for recreation or for other purposes or are not part of the school's plan. It is true that the current definition of the curriculum is a more adequate one for thinking about, for planning and for conducting an educational program, but it does mean that a review of the curriculum must include a larger range of matters than

would be required if the definition were limited to the outline of content to be taught.

To bring my task into manageable size, I have chosen to focus attention on the development of curriculum theory over the past fifty years with occasional comments on the ways in which courses of study and curriculum guides diverge from the accepted rationale and with still fewer occasional comments on the discrepancies between teaching practices and curriculum theory. To simplify this complex review still further, we shall examine each of three major aspects of the curriculum in turn, treating the formulation of educational objectives, the selection of learning experiences, and the organization of learning experiences. Although the evaluation of the effectiveness of the curriculum is commonly included as an aspect of the curriculum itself, I shall not discuss it since the two following papers deal with this subject.

### *The Formulation of Educational Objectives*

A major step in most theories of curriculum development is the formulation of the educational objectives of the school, that is, the goals to be attained by its educational program. To be sure, this is not an appropriate step in John Dewey's educational philosophy in which the direction of learning is guided by careful consideration of the quality of the learning experiences. Insofar as the learning experiences provide for continuity and interaction, in Dewey's terms the program is effective. His philosophy does not involve a distinction between ends and means. But the other chief leaders of American curriculum thought emphasize the importance of clear objectives as the basis for planning the learning experiences and appraising the results.

Since the turn of the century there have been several marked changes connected with the formulation of the objectives of the school. One of the most obvious has been the changed conception of the nature of educational objectives. The dominant educational psychology in 1900 was based on the theory of formal discipline and expressed in terms of "faculty psychology." The mind had certain faculties such as memory and reason which could be trained or disciplined by proper exercise. The objectives of the school were stated in terms of the faculties to be trained, and the learning experiences were those exercises in which these faculties were engaged on content particularly rich in opportunities for memorization, reasoning, and the like. Certain subjects by the very nature of their form and content were superior means for cultivation of these faculties. Language, particularly Latin, for example, was a superior subject because the learning of it required the exercise of

memory, while its grammatical structure provided exercise in orderly reasoning.

With the decreasing acceptance of the theory of formal discipline and the elimination of faculty psychology, the prevailing view became increasingly behavioral. Learning was then conceived as the acquisition of patterns of behavior which the student had not previously followed. Human behavior was defined quite generally to include all the reactions of an individual, his thinking, feeling, acting.

Educational objectives are now couched in behavioral terms. An objective is a statement of a kind of behavior pattern which the school seeks to have the student develop. In the first flush of behavioral concepts, roughly from 1918 to 1925, the objectives were commonly stated in highly specific terms, such as ability to add 2 plus 3, ability to use the indefinite article "an," ability to spell "believe," ability to recall the atomic weight of sulphur. This was a natural corollary to the prevailing associationist theory in the psychology of learning. Every number combination, for example, was viewed as a different stimulus to which the student was to learn an appropriate response. This extreme view led to the listing of nearly three thousand specific objectives for arithmetic, and nearly two thousand for English. A student had attained the goals of the curriculum when he had learned to make the appropriate responses to all of the specific stimuli, that is, when all of these innumerable objectives had been reached.

By 1925, this view of objectives had largely fallen of its own weight. On the side of the teacher it required keeping in mind far too many goals to be remembered, and on the side of the student, it denied the development of generalized behavior patterns which quite obviously were developing. The formulation of other theories of learning which took into account the phenomenon of generalized behavior provided terms in which educational objectives have commonly been stated since 1930. For example, in 1936 the Department of Superintendence of the National Education Association published a yearbook on "The Social Studies Curriculum." Among the objectives suggested were: (pp. 320-340)

1. Acquisition of important information
2. Familiarity with technical vocabulary
3. Familiarity with dependable sources of information on current social issues
4. Immunity to malicious propaganda
5. Facility in interpreting social science data
6. Facility in applying significant facts and principles to social problems of daily life
7. Skill in investigating social science problems

8. Interest in reading about social problems and in discussing them
9. Sensitivity to current social problems
10. Interest in human welfare
11. The habit of working cooperatively with others
12. The habit of collecting and considering appropriate evidence before making important social decisions
13. Attitudes favorable to social improvement

These obviously present a conception of generalized behavior. However, although they avoid the piecemeal aims of highly specific objectives, they may be as limited in their value for guiding teaching as the earlier statements of objectives in terms of faculties to be developed unless each of these thirteen objectives is clearly enough defined to have meaning for the teacher so that he can easily think of concrete illustrations of the general aims. The developments since 1935 in the conception of the nature of educational objectives have largely focused on defining in concrete terms aims which are expressed at a similar level of generality as those above. These efforts have been applied to defining the kind of behavior implied by such general terms as "understanding," "applying principles to concrete problems," "ability to interpret reading material," and to indicating the range of content to which each kind of behavior is to be applied. Thus, the objective "to develop understanding of the basic concepts of physiology" has been defined from the standpoint of behavior and of content. The behavior "understanding" is defined as "the ability to recall the concepts, to state them in one's own words, to give illustrations of them, to recognize illustrations given by others, and to compare and contrast related concepts." The content termed "the basic concepts of physiology" is defined by listing some two score concepts which these curriculum makers have selected as basic to this science. This kind of definition helps greatly to clarify the aims of the curriculum so that they can actually be utilized in planning and conducting an educational program in terms of the prevailing conception of the psychology of learning.

A second marked change in the objectives of the American school curriculum has been in the sources used to derive the aims. To some extent all of the five major sources have been used [illegible] period of American history but at a given time certain sources [illegible] dominant in their influence while others are given only minor attention. Between 1900 and 1918, the judgments of subject specialists and the prevailing conception of the psychology of learning were dominant in formulating objectives. At the high school level, the Committee of Ten used subcommittees of mathematicians, historians, language scholars, and the like to outline the objectives of secondary school instruction in these

fields. Although the prevailing educational philosophy had already emphasized knowledge and skill for the layman as a major aim of the American high school, this was given little attention in deciding on objectives. No studies were made of the needs of society nor of the needs of students to help in identifying appropriate objectives.

As a result of the success of job analysis in building vocational curricula during World War I, the process of formulating objectives from 1918 to 1933 leaned heavily upon job analyses, activity analyses, word counts and other techniques for identifying the demands made on the individual by contemporary social life. At this time, curriculum makers also gave attention to the notions of educational psychologists as to what behaviors could be taught. However, during this period little attention was given to the prevailing social and educational philosophy regarding the characteristics of the good man and the good society. The opinions of subject specialists were given much less weight than in the previous period.

From 1933 to 1945, studies of children and youth served as a major source of suggestions for objectives. With an emphasis upon the responsibility of the school for meeting the needs of children and youth, curriculum commissions drew upon child study data and reports of adolescent studies to derive objectives. This largely coincided with the prevailing emphasis in educational philosophy and to some extent the work of educational psychologists was used. But the use of studies of social demands was notably less than in the decade previous while the opinions of subject specialists played a very minor role.

Since the Second World War, the shift in emphasis among the five kinds of sources has been marked. Primary attention is currently given to the opinions of subject specialists, particularly in mathematics and science. Very little weight is currently given to studies of the learner, but the specialists are asked to outline what they believe to be important potential contributions of their fields which will be of value to laymen as well as persons planning to specialize in the field. In this respect the emphasis is different from that in 1900. Today some attention is also being given to an examination of social demands, and to a lesser extent to the current conception of the psychology of learning. Much less use is made today of studies of the learner than was true fifteen years ago. In general, the shifts which have taken place in the primary sources used to derive educational objectives most closely parallel the changes which can easily be seen in the statements of objectives appearing in courses of study and in curriculum guides. Because the actual practice of teaching depends so largely on the habits and outlooks of the thousands

of American school teachers, the shifts in practice are not so easily discerned.

A third marked change has been in the range of objectives which the American schools have not only accepted for themselves but have actively championed. At the turn of the century there was a sharp difference between the claims made regarding the schools' general contribution in promoting citizenship and character and the working objectives of the curriculum which were focused on knowledge and skills and intellectual discipline. The development of many basic attitudes, values, interests and habits was recognized as a primary function of the home and church, and for those habits, attitudes and skills relevant to work the employer was expected to play a strong role. The school today commonly lists the whole range of educational goals required for the induction of young people into effective adulthood. It includes objectives relating to home life, personal-social relations, civic life, occupations, and the like. It includes not only knowledge and intellectual abilities but interests, attitudes, social and recreational skills. Frequently, too, there is no indication of relative weighting. Developing social skills and a cooperative attitude appear to be viewed as jobs as important for the school as developing understanding of basic concepts of science and the social studies or the acquisition of the skills involved in reading.

Since the level of learning required of people today is a high one, a major problem in education is to select wisely among all the possible goals the important tasks which the school can do well and to concentrate its energies effectively. Since the total educational job is very great, the home, the church, the employer and the other potential educative agencies of the community need to be encouraged and strengthened to take their share while the school concentrates on the things it can do best, and in many cases the things that only the school can do. Hence, the present shift in school objectives is toward a more discriminating selection, toward the kinds of learning which involve intellectual skills, which require sequential experiences to reach the necessary level of competence, which involve concepts and principles that are not apparent on the surface and for this reason are not likely to be learned through the guidance of laymen. This shift is likely to reduce the great range of objectives, and to diminish the emphasis upon social adjustment and similar goals which fail to recognize the importance of individuality and individual creativity in responding to experiences and in solving problems. The increasing emphasis upon understanding and thinking as kinds of objectives, with lessened stress upon attitudes and habits as primary goals, may help to revive the conception of the

individual who controls his feelings and actions in terms of his knowledge and thought rather than one who simply seeks to express "acceptable" attitudes and feelings and to do the "proper" thing. This is a shift in objectives which will be interesting to observe.

*The Selection of Learning Experiences*

Among the changes taking place in the learning experiences provided by the American schools, those in the prevailing notions of the nature of learning experiences are particularly significant. At the beginning of this century, the term was not used. Exercises, assignments, examples, problems were the words commonly employed to designate the learning tasks set for students outside of the class session, while the term "recitation" was used to refer to the oral responses expected of the student in the class. No mention was made of the student's mental reactions in the class, although it was clear that he was expected to pay attention, that is, to watch and listen to the teacher's presentations. When I began to teach more than thirty-five years ago, we had to file lesson plans for each week in advance. These plans outlined the content to be covered, what the teacher expected to do and the out-of-class assignments to be made. The focus of planning was on the teacher.

John Dewey and other educational leaders gave wide publicity to the increasing psychological evidence that learning could be most readily interpreted in terms of what the learner was doing. It was his reactions that he learned—not the teacher's. The teacher's role was to stimulate, guide and reward the learner as he carried on the behavior which the school sought to teach him. This view placed attention upon the activity of the learner as the basic factor in attaining educational goals. By 1925, both writings of theorists and curriculum guides were commonly using the term "learning activities" to refer to the basic elements of the teaching-learning situation. Courses of study were listing reading activities, listening activities, study activities and laboratory activities in outlining the day-by-day program of the school.

By 1935, curriculum writers were pointing out certain limitations in the concept of learning activity. For example, two students might both be reading an historic account of the California "Gold Rush," yet each might be carrying on quite different mental reactions and making different emotional responses. One might be thinking of the excitement and challenge involved in the long wagon haul the pioneers made in crossing the country, thrilling himself as he imagined the Indian encounters. The other might be thinking of the rough, lawless life of the early mining community, wondering why people would leave the comforts of civilization to live in such trying conditions. In terms of the

course of study, both were engaged in the same learning activity but each was having a different experience and to that extent was learning something different. This kind of analysis led to the adoption of the term "learning experience" to refer to the reactions of the student in the situation. In 1936, Dewey's book on *Education and Experience* clarified this concept further by emphasizing the notion that "experience" involves the interaction of the individual with the situation. This interaction involves some mutual effects, the individual modifying his reactions in terms of the demands of the situation and he also modifying the situation through his reaction to it. Today, almost all curriculum writers use the term "learning experience" and they seek to plan the learning situation so as to give direction to the experience the student has, that is, to his internal perception of the situation and his own interaction with it. This requires consideration of what the learner brings to the situation, what it will mean to him, how he is likely to respond to it mentally, emotionally and in action.

Beginning with James and Thorndike and exercising increasing influence in recent years is the conception of the learning situation as one which should provide for certain essential conditions of learning. Thorndike's earlier work emphasized two conditions—exercise and affect. Current curriculum guides mention such conditions as motivation, opportunity for practice, guidance of desired behavior, provision of satisfaction when desired behavior is elicited, and the like. Hence, some of the current courses of study are pointing out the need to consider these conditions in selecting the learning experiences for a particular class group from a larger list of suggested ones.

A second marked change in learning experiences can be found in their range. Although the Floyd movement had influenced forward-looking American schools in the late 1800's to introduce manual training, not as vocational training but as a means of "learning through the hands," most of the learning exercises employed at the turn of the century were verbal ones. Listening and reciting, reading and writing represented the ways of learning the "academic subjects," except for the laboratory periods in high school science. Even the laboratory exercises were heavily verbal, with detailed instructions in the manual and a formal plan for writing up each "experiment." Map work in geography and field work in biology were strongly recommended by the writers of the period from 1900 to 1915. Most courses of study advised having children make maps and locate points of geographic interest on them. At this time, too, high school botany courses typically required the student to collect and identify fifty or more plants.

By 1910, high school agriculture was widely offered in rural areas. These were first courses to introduce the project, or "student initiated" enterprise, which, it was hoped, would help him to understand and to apply the knowledge he was gaining in the course. The use of projects spread to other fields and to the elementary school, thus providing a much wider range of learning experiences than schools had commonly used. The writers who urged the introduction of projects conceived of them as involving a range of experiences as broad as life itself, but in the actual use of projects in the schools, activities involving the construction of objects have been predominant. Many teachers think of a project as making, growing, or producing some physical object. The extended inquiry which Dewey thought had largest potentialities as an educational project is rarely found. The intellectual learning experiences are frequently quite minor to the physical manipulations required to complete a "construction project."

During the depression, with its great reduction in opportunities for remunerative work for youth, many secondary school leaders recommended the addition of work experience to the high school program. Although only a small minority of high schools introduced work experience as part of the curriculum, some developed well-planned programs which involved using a wide variety of work activities as means for attaining educational objectives related to science, social studies, mathematics and English as well as vocational fields.

The greatest impetus to extending the range of learning experiences has been the technological developments in communication. Lantern slides were in use at the turn of the century but were not found in many schools. At best they served only to extend the number of pictures which could be employed, to add concreteness, or to give variety to the teacher's presentation. The perfection of the motion picture, however, made it possible to analyze movements, to show time and space relationships much more graphically and to increase the sense of reality in dealing with many subjects which require vicarious treatment. The addition of the sound track heightened the sense of reality and added another dimension of analysis. The sound-slide film gave some of the features of the sound motion picture in a more economical form, but it lacked the distinctive assets of motion. The television set made possible instantaneous viewing of events in a fashion much like the motion picture but with a further sense of the reality of the event, due to the viewer realizing that it is taking place at the same time he is seeing it. These technological developments have gone far in removing the physical limitations to providing as wide a range of learning experiences in the school as those of life outside. But much of the comprehensive, effective

development of these potentialities lies ahead. They still represent a small per cent of the learning experiences provided by American schools.

The selection of learning experiences so as to provide for individual differences among students is another respect in which changes have taken place in the last fifty years. Attention to individual differences has been accentuated by two factors: the psychological studies which have identified the extent of differences among schools, among classes and among students in the same class; and the increased visibility of individual differences brought about by the enrollment in the school of children from heterogeneous ethnic groups and social classes. There are few teachers now who fail to recognize a variety of differences among the students in their classes—differences which affect interests, meanings, efforts and outcomes in school work.

Typical devices to provide for differences among students have involved adaptations in the time given for completing learning exercises, or variations in the exercises themselves, or both. The first type of adaptation requires a plan for students to work at varying rates. Among the early developments were the San Francisco, Dalton and Winnetka plans, all of which involved organizing the school day into two parts, one for group activity and the other for individual work. These plans also required the development of a series of assignments with full directions on paper so that the students could work as individuals on different assignments at the same time. As a student took an assignment it became his "contract" which he undertook to finish before he went on with another assignment in the same field. He might, therefore, complete his assignment much earlier or much later than the average.

Adaptations of the learning experiences themselves were first found in courses of study which marked some of the exercises as those to be required of all students and others as optional for the better students. By 1915, this was common among American schools. By 1925, a number of cities had introduced "ability grouping" in which the course of study was differentiated in such fields as reading and arithmetic into three levels—the superior, the average and the slow sections. These three courses of study differed in the time provided for learning exercises and to a lesser extent in the nature of the exercises. In reading, the amount of material dealing with personal and social activities of children was greater in the slow sections, while the adult material was greater in the superior sections. In arithmetic, more concrete objects were counted and compared in the slow sections than in the others.

The use of individual projects was also a means of adapting to the individual student's interest and ability. This was recommended in courses of study as early as 1915. Learning exercises carried out by

small groups (two to ten students) were first employed in the late 1800's to compensate for inadequate laboratory equipment. The apparatus was insufficient to provide opportunity for every student working individually to carry out the assignment. By 1930, small group projects were being used by many schools as a manageable means of providing for individual differences. The projects themselves could differ in the rigor of their intellectual demands and the division of labor among the students in the small group could adapt further to the abilities and interests of the individual. Unfortunately, all too often the slowest learner was given some handwork which involved little or no new learning. By 1950, with the publication of research on the psychology of small groups, educational writers were recommending the use of small group projects as a means of heightening motivation and increasing the amount of meaningful learning activity. Since 1918, the attention of educational leaders has focused increasingly on the "education of the gifted student." This has led to emphasizing learning experiences which require greater understanding, or skill, or effort than those usually provided in the course of study. It has also stimulated some schools to develop learning experiences that can be carried on as independent work.

The most typical development in the past twenty years found in courses of study to provide for individual differences has been the listing of a large number of suggested learning experiences from which a given teacher may select those particularly appropriate for his class as a whole or for groups or individuals within the class. The uniform lesson plan so common when I started to teach is almost unknown now. Most curriculum guides include a discussion of how to select from among the large number of learning experiences suggested in the course of study those which are likely to be most effective for students with varying backgrounds and abilities.

### *The Organization of Learning Experiences*

Important educational objectives involve patterns of behavior of such complexity that they can be developed only gradually over considerable periods of time. For example, the ability to read critically and to make comprehensive interpretations of what one reads is not acquired in a few brief lessons. To understand the basic principles of science and to use these principles in explaining the biological and physical phenomena round about us require a variety of related experiences extending over many hours. If the development of such complex behavior patterns as these is left to isolated or unrelated periods of learning, adequate achievement is impossible. Hence, a major phase in building a curriculum is to

work out an organization of the many, many learning experiences required so that the student develops these complex behavior patterns gradually, day by day, and relates them to others so as to have an increasingly unified understanding, and a well integrated command of essential skills.

The purpose of organizing learning experiences is to maximize the cumulative effect of the large number of learning experiences required to develop complex behavior patterns. Three criteria are commonly considered as standards to be met by a well organized curriculum, namely, continuity, sequence and integration. Continuity refers to the reiteration of the desired behavior through the many learning experiences used. Sequence refers to the gradation of the learning so that each subsequent experience not only builds on previous ones but goes beyond in order to require a higher level of skill, or a broader or deeper degree of understanding. Integration refers to the relation of what the student is learning in one field to what he is learning at about the same time in other fields. A broader and deeper understanding is facilitated by comprehending the relation among the various concepts, facts and principles being studied, and a more adequate command of basic skills is achieved as the relation of these skills to one another is seen.

One surprising fact about curriculum development in the last fifty years has been the limited attention given to the theory of curriculum organization. Other than the common-sense notions of these criteria and of such rule-of-thumb principles as "learning experiences should proceed from that which is known to that which is unknown, from the simple to the complex, from the easy to the difficult," no new formulations have been made since the time of Herbart and of James. This is an area crying for substantial theory to be tested in practice and to provide a guide for practice.

At the more specific level, developments in reading and in the foreign languages have been most marked. In reading, continuity and sequence are commonly achieved through carefully controlled vocabulary development, adding new words gradually and systematically, and through the control of sentence structure in the reading materials, beginning with simple declarative sentences and moving gradually to compound and complex ones. Integration is sought both by relating the reading material to the common activities of the children and by introducing work-type reading in the other subjects on a gradual basis. A similar scheme of organization is commonly followed in the foreign languages.

In arithmetic, the development of skills is usually facilitated through an organization which begins with learning experiences involving addition and subtraction, then multiplication and division, then common

fractions and decimal fractions. No explicit scheme of organization for concept development in arithmetic can be found in the current courses of study. The content of arithmetic problems has changed greatly since 1900. Beginning about 1920, studies were made of the kinds of problems commonly encountered by children and adults. Typically, arithmetic courses now order the problem content in terms of frequency of occurrence of the problems outside of school and in terms of the age level at which this kind of problem is commonly encountered by children.

The typical high school curriculum in mathematics has changed little in the past fifty years so far as organization is concerned. Tenth grade geometry builds little, if at all, upon algebra. Advanced algebra and solid geometry in the eleventh grade have little sequential relationship to tenth grade geometry, and trigonometry in the twelfth grade does not provide a clear sequence for the eleventh grade work. The so-called "modern mathematics program" which is now getting under way with the sponsorship of the mathematical organizations should provide a much better organized curriculum for high school mathematics.

In organizing the so-called content fields, like the sciences and the social studies, major attention has been given to the ordering of content rather than behavior. At the beginning of this century science was not commonly taught in the elementary school, while in the high school botany was most frequently offered in the tenth grade, physics in the eleventh and chemistry in the twelfth. By 1920, general science was offered as the introductory science course in more than one-fourth of the high schools and now it is taught in almost all schools in the eighth or ninth grade, with biology in the tenth and physics and chemistry, where offered, being placed in the eleventh or twelfth grades. The content of general science is usually selected to relate to the scientific phenomena most commonly observed by children. The content of biology is usually chosen to explain the human body, the maintenance of health and the conservation of natural resources. The organizing notion here is to begin with phenomena which are common in the student's environment and in which he is likely to be interested. The advanced science courses, physics and chemistry, deal with the more abstract principles, which are thought to be less common and more difficult. The organization of these two courses has not greatly changed in the past fifty years. These illustrations in the field of science indicate the attention given to organizing the content dealt with in the learning experiences, but no similar effort has been made to organize the behavior, that is, the skills and abilities to be developed.

This is also true for the social studies. The changes taking place in their organization have been changes in the ordering of content. The

most common sequence of content in the social studies is to begin with the community, then the state, then the nation, and finally the world. There is little evidence to indicate that this is sequential in terms of difficulty in learning.

Thus far, we have been reviewing the continuity and sequence of learning experiences in the content fields. The problem of integration, that is, how to relate learning experiences so as to aid the student in seeing the relations between what he is learning in one field with what he is learning in another, has been attacked most commonly through changes in the structure of the curriculum. In 1900, the elementary school curriculum was composed of ten or more specific subjects like reading, writing, spelling, arithmetic, geography, history, nature study, hygiene, music, drawing. Now, the typical course of study includes reading and the language arts, arithmetic, science, fine arts, health. This reduction in the number of subjects has been accomplished by building a more closely related series of learning experiences in language, in which reading, writing and spelling are involved; in social studies, where geography and history are interrelated; and in the fine arts, where music, drawing, and painting are brought together.

In the high school, the broad fields of English, mathematics, science, social studies, foreign language and fine arts have frequently replaced more specific subjects, and in some cases the core curriculum has been developed which provides a large structure for learning experiences that occupy from one-third to one-half of the high school student's day. Since these larger structures are usually planned as courses rather than several separate sub-courses, there is opportunity for better integration. Typically, however, the only principle of integration which has been explored is to bring together the content and skills needed to deal with each of the student "problems" which provide basic units of the course. This principle does not always provide for the necessary continuity and sequence nor for all of the more helpful relationships among the fields which are involved. In many cases, a particular problem involves knowledge and skills from certain fields in only a minor degree and does not suggest the more significant ways in which these fields are related.

It is clear after reading the works of curriculum theorists and examining courses of study that the past fifty years have not been a time of great development in the organization of learning experiences. In this respect, curriculum changes have been relatively slow. The careful, systematic work done in the field of reading is a shining exception. The arousal of interest and stimulation to thought among secondary school teachers who have worked on the construction of core curricula suggest

the great intellectual resources available under effective leadership to attack fundamentally and systematically the problem of developing a better organized curriculum.

*Summary*

This review of changes in the curriculum of the American schools during the past fifty years has touched several high spots, but it has not presented possible explanations for the kinds of changes noted. It is probable that many of these developments can be understood in terms of the tasks which the American schools were facing at these different periods and the ideas prevailing in the field of psychology which school leaders found when they sought from scholars assistance in attacking critical school problems.

In the period prior to World War I, the elementary school was steadily growing to include a larger per cent of the children of age 6 to 14. The critical task was to teach the 3 R's to children of immigrant parents and those from the working classes. This required a re-examination of the psychology of reading and arithmetic, for the usual background of interest in and experience with language and children's stories could not be taken for granted, nor could early experiences with number concepts in the home be assumed for all children. It is not surprising that in this period long lists of specific objectives for these subjects were worked out and that special attention was given to children's interests and abilities in devising learning experiences.

World War I ushered in a fifteen-year period when the economy developed rapidly and workers were in demand. No wonder that job analyses and other forms of analyses of social life were used to identify more definitely the demands which the individual would be expected to meet and to use this source in setting objectives. From 1930 until World War II, the great depression increased rather than decreased the number of youth in school because of the limited opportunities for employment. But the same limitations in jobs made the analysis of social demands a less relevant source for educational objectives than a study of youth themselves to find needs and potentialities that might justify educational effort when no great social demand was apparent. This period also was a difficult time for many secondary schools because youth saw no future and had no interest in deferred educational values. In their eyes school work had to be justified immediately in terms of its interest and meaning to them. Hence, the devising of learning experiences of immediate interest and relevance to students and their organization around student problems helped to meet this pressing problem.

Since World War II, the insatiable demand for technically trained people has focused attention on the opinions of subject experts as a source of objectives and has given greater emphasis to knowledge and skills. The current demand for highly competent professionals and technicians has increased the interest of the schools in the education of gifted children and in devising a better organized curriculum to reach high levels of achievement. The shortage of teachers has furthered the demand for technological devices that increase the range of learning experiences, such as television and the sound motion picture.

Throughout the fifty years, the schools have been pressed by continuing conditions which create critical problems that cannot be solved without further curriculum developments. The first of these is the rapid change in technological development and social life which requires a continually increasing level of education on the part of our people. The second is the increasing proportion of children and youth who are sent to the schools for education. The third is the dislocation in other educational institutions, the home, the church, the neighborhood which rapid social change has engendered. The educational needs of today and the immediate future are greater than ever before. American education has done an amazing job in getting almost all children and youth in school and providing schools for this immense number. The schools have been astoundingly successful in building confidence on the part of the public in the capabilities of education in building our civilization. The time has come, however, to recognize realistically the magnitude of the job, to identify the objectives which the schools can best attain, to encourage the home, the church and other institutions to undertake the tasks appropriate to them, to devise learning experiences clearly relevant to the school's proper objectives and to work out an organization of the curriculum which aids the students in attaining a high level of educational competence. These steps still lie ahead of us.

# New Kinds of Students and New Ways of Testing Achievement

LOUIS M. HACKER

I

I should prefer to talk largely of some of the impressive problems that are to confront us in the immediate future rather than our shortcomings in the past. I can view both from the vantage point I occupy as an educator, associated as I am with a somewhat unique institution. The School of General Studies of Columbia University is an outgrowth of its Extension Division, which was established more than 50 years ago. Seeking to reach adults with a program of broadly cultural and narrowly vocational and technical subjects—and patterned after the Workers Education Association and University extramural tutorial schemes of Great Britain—up to less than 10 years ago, this supplement to formal education at Columbia University reached a large audience in an urban community. Given almost exclusively in the evening, its lecturers for the larger part from the University's own staff, the program mainly was a non-credit one, although a few hardy individuals by persevering through its miscellany of courses could put together a coherent curriculum to obtain a degree.

In 1948, the School of General Studies was established to grant an undergraduate degree but it did not surrender its association with its particular student community. That is to say, adult men and women who could attend formal studies only as part-time students either because they were fully employed or because they were housewives. Those who were seeking matriculation were required to take an aptitude examination generally similar to that of the College Entrance Examination Board. Those who had no high school diploma or its equivalent, following the aptitude examination, were permitted to matriculate—without conditions or deficiencies—upon the successful completion of a so-called Validation Program consisting of the following 5 subjects, each for one semester: mathematics, history, English, a foreign language, and chemistry or physics or zoology. In order to protect the program, non-credit courses of a vocational nature were dropped and every course offered had to be under the auspices of a department of the University.

To encourage these students many of them removed from classroom experiences and therefore formal learning for as many as 20 years or more, the status of non-matriculated student was created. Any person —

with a high school diploma or its equivalent—could take up to 9 points a semester without the Aptitude Test; in short, adults seeking a serious education were permitted to test their capacities in the classroom. If the courses for which they were enrolled were elementary ones, they could register without further ado; if they were advanced, prerequisites had to be satisfied or the instructor convinced of their ability to go ahead on the basis of training in or out of the classroom.

This was a bold step, and to it the Faculty of General Studies continues committed. It is based on the assumption that older persons should somewhere have the right—in the classroom itself—to find out whether they can train their minds for formal learning. It also is prepared to recognize that such people frequently are timid about exposing themselves immediately to entrance examinations or embarking upon an orderly program of studies of the kind we find in the characteristic first-year college program, that is, English, history, mathematics, a foreign language, a laboratory science. We have found that many older persons would like to try first the subjects they are already familiar with as a result of reading, or other experiences; in consequence, they are likely to concentrate in literature or history or mathematics or a foreign language. As non-matriculated students they can do so, so that confidence is built up, exposure to classroom procedures takes place, and opportunities for counselling—by teachers or advisers—present themselves. A student can continue under this non-matriculated status for as many as 90 points; a few of them—older women in particular—wait that long before formally entering for degree candidacy. Hundreds of such people each year are slowly led from non-matriculated to matriculated status. In fact, of the 1500 matriculated students in the School enrolled this winter semester, almost 50 percent started their educations in the School itself as non-matriculated students.

Increasingly, therefore, the School's 5500 non-matriculated students are the source from which its matriculated students are obtained. It should not be assumed that all of these have had no previous collegiate experience. At least 20 per cent are already undergraduate degree holders, taking courses or programs to complete their preparations for graduate study or because they wish to make up subjects for entrance to professional schools. Some are on probation because of previous failure or academic suspension. But the "casual" student, just taking a course, is in process of disappearing. The non-matriculated student in the School has become a pre-matriculant looking forward ultimately to a regular degree program.

There are risks, naturally, in these procedures: for non-matriculated and matriculated students are not segregated in the classrooms and—

with such a heterogeneous student body—there is the danger of lowering standards. These can be protected only by the most rigorous classroom controls: small classes, frequent examinations, highly competent instruction. Those who, it is perceived within the first few weeks, simply cannot make the grade are requested to drop out with full tuition refunds; those who fail are denied re-registration. Matriculated students are closely tested at the end of their degree programs, for each student must take a major and that cannot be completed or graduation achieved without a departmental comprehensive examination which is written or oral or both.

I have said that the School's Faculty is opposed to the general use of preliminary examination for course admission—in fact, only in one subject is a placement test employed and that is for elementary English composition. This winter, some 400 persons took this test and about 10 per cent failed it. These latter were put in English cA1, a remedial course in composition, or English c1, a remedial course in reading habits and study skills. Even English cA1 keeps the door open: for those receiving a grade of B − or better obtain credit and are exempt from English A1.

The School's Faculty is persuaded in its course as a result of its experiences. Not only are larger and larger numbers of persons successfully completing courses of their own selection—sometimes over as many as 5 or 6 years—and then becoming matriculated. In addition, as many as 60 to 70 per cent of the graduates each year enter at once further studies in professional and graduate schools. Here, before admission, most of them are formally tested by admitting offices; and their entrance and subsequent successes support the conviction that the School's educational standards have not been lowered or harmed by its open-door admissions policy.

This does not imply that the School is not constantly seeking objective and comparative measurements of performance. Columbia University has two other undergraduate institutions—Columbia College for boys and Barnard College for girls. They, like the School of General Studies, have the major system. Recently, I invited the deans of the other two schools to establish with us a common comprehensive examination in each department granting a major for the testing—and graduation—of the three student bodies under similar auspices. I was prompted to do so as a result of my personal experiences with the Oxford Colleges. There are at Oxford University, some 20-odd undergraduate colleges, each of which has its own admissions policy and its own instruction. But each student must enroll in an Honors School—the equivalent of our major—and at the end take a series of examinations set not by his

own instructors but by the Faculty in which he is enrolled. Graduation depends upon the successful completion of these examinations; consideration is given to no other factors.

I am saying, in consequence, that as far as older students at any rate are concerned, one of the significant ways of measuring achievement has very little to do with the customary testing devices. Sympathy for the problems of such students as they embark upon—or resume after long periods of discontinuity—the formal processes of learning; a willingness to disregard, from time to time, initial test scores; the refusal to insist upon adherence to a rigorous pattern of sequential studies—these attitudes have led to the ultimate salvaging of what turn out to be very good and not infrequently excellent minds. For, in the School of General Studies, we have seen thousands of men and women—who never had the chance earlier to begin a collegiate education or who failed somewhere else; many of whom begin with us falteringly—not only graduate well but go ahead in professional and graduate schools to careers of great usefulness. These attitudes on our part have added a new resource to the trained manpower of the nation. Particularly in the case of women, who with early marriage and early families cannot continue education after high school or the first two years of college, there is made possible training for a career at age 35 or later. When I remind you that the drop-out rate for women's colleges runs between 25 and 50 per cent, and that the life expectancy of women at 35 years is more than 40 years, then the need and possibility of resuming their education—given the proper auspices—is one of the important challenges confronting us.

II

But let us turn to the immediate future. There is no need to repeat to this audience the character of the educational problem that will confront the United States within the next decade. Not only has our birth rate doubled since 1940, and thus automatically doubled the age cohort from which our undergraduate college body is drawn. The efforts to move into college that significant group of high-score high school seniors who for one reason or another currently are not attending; the steadily mounting real family income, making it possible to finance privately the college education of more than one child in the family; the growing interest of companies and even trade unions in the establishment of scholarships—these inevitably are to have large effects on the size of our college population. At a minimum there will be a doubling, at a maximum at least a tripling, within the next 15 years of the residential full-time student body.

If to these, we add the part-time older students taking formal studies and many of whom will also be degree candidates, then the problem takes on staggering proportions.

Public authority—municipalities, the States—under constant pressures, will seek to discharge their obligations in the easiest and cheapest ways, and that is through the creation of large numbers of local institutions offering two-year programs. The pattern already is a familiar one. In part supported by the local community, in part by the state, these community colleges or institutes of the arts and applied sciences or junior colleges, are offering both terminal and lower-college programs. The terminal programs are vocational and end there. The lower-college programs are in the liberal arts and are also pre-professional, e.g., electrical engineering and mechanical engineering.

Two problems will emerge as a result of their establishment on a general scale. Because they are being locally supported, it will be difficult to resist the demand that entrance be general and not restrictive on the basis of admission tests. (These have been the experiences of the state universities, as we know.) Even if an effort is made to restrict entrance in the purely academic division, students will be able to enroll in the vocational division and seek transfer later, on the basis of good performance (as they will have every right to do).

In the second place, what to do with those students who, having completed a two-year academic or pre-engineering program, have the desire and the capacity to go ahead for, at first, an A.B. or a B.S. degree and later, for graduate or professional studies? Current experiences with these two-year programs already indicate that at least 25 per cent and perhaps as many as 50 per cent of such students want to and have the capacity to continue their education.

How to test the achievement of students in these two-year programs in order to make possible immediate full-time continuation, part-time continuation (for those who will go to work), or deferred continuation at a later date (the so-called "delayed transfers" as we refer to them in the School of General Studies)?

Let us look again at the older, part-time students in the School of General Studies. Many of them have lost a good deal of time as a result of discontinuity in education. (The so-called "delayed starters" and the so-called "delayed transfers," who constitute 66 per cent of the matriculated students, have a median age of 30 years. They are, on an average, 13 years behind in the completion of their education.) Every effort is made, as a result, to encourage acceleration: by professional option, by double summer sessions, by heavy programs (if they are full-time), and by credit by examination. (This last, approving 30 credits by examina-

tion in elementary subjects, has been adopted by the Faculty but as yet has not been accepted by the University.) With so much time lost and with so large a proportion of these older students heading toward advanced studies, acceleration is desirable and, in fact, necessary.

I am, in consequence, joining the problem of the older, part-time student and that of the accreditation of the student from the two-year community college who immediately, on a part-time basis, or subsequently, will seek continuation of his education.

For these I propose the establishment, for the first time in America, of a group of universities—regionally located—which will take on the original function of the university as an examining body. I do not mean that these universities should be separated from the teaching function and the residential students. I am suggesting that they add to their customary activities the examination of—and therefore the granting of degrees to—students who are taking the academic and pre-professional sides of the two-year community-college programs and the older students, largely seeking acceleration, who want college degrees.

My prototype, obviously, is London University. Initially, when it was founded more than 100 years ago, it was wholly an examining body and all its students were so-called "external" students. With this pattern (excluding Oxford and Cambridge, naturally), England found it possbile to establish its large network of university colleges through the country. This was the origin of the "Redbrick University:" locally set up, whether at Birmingham, Leeds, Nottingham, or elsewhere, the university college admitted its own students and gave them instruction. But the degree was that of London University, for the examinations were drawn up and graded by the professors and senior lecturers of London.

Within recent years, all but one of these university colleges have obtained charters and been set up as independent universities. At the same time, within recent years, London has been taking its own residential or "internal" students.

But London continues as an examining university for "external" students, notably in these three groups:

1. Students who cannot establish residence, either because they are part-time or because they live elsewhere. A student still can obtain a degree—undergraduate, graduate, or professional—simply by standing for the examinations offered by London University. No formal attendance, anywhere, is required.

2. Students who attend university college in the crown colonies. A group of such colleges—in Africa, the West Indies—has been set up: like the earlier English university colleges, they are residential. But the degree is that of London University.

3. Students attending the polytechnical schools of Greater London and elsewhere. These have been established by local authorities and are financed by these and not nationally, through the University Grants Committee. They have their own admissions policies and instruction programs. They started out, like our own institutes of arts and applied sciences, to give vocational programs; increasingly they are offering academic and pre-professional and engineering curriculums. Their degrees, however, are London University degrees, for the final examination of the students is the work of the London Faculties.

These "external" students of London University may take instruction from London through correspondence courses, whose papers are carefully read and graded. But not for degree purposes; here, again, as at the Oxford and Cambridge colleges, instruction is really coaching for the final examinations, which are the only bases for degree granting.

If, then, a group of American universities, located in great urban communities, take on the "external" degree examination function, in addition to the education of residential students (thereby assisting the development of London University), we shall be in a position to regularize and wholly accredit the work of older part-time students and that of those in the two-year schools on the academic and pre-professional sides.

For the older students, residence, formal curriculums, grades for each course, a minimum of 124 credits for graduation no longer need be required. A student may attend or not—and if he attends as an "external" student, it will be for coaching purposes only. Attendance, for "internal" students of course, will continue; and those who elect a formal residential program should have the opportunity of doing so.

Other coaching aids, for the "external" students—indeed for the community at large—might also be included. I refer to correspondence courses and to courses through television; these, of course, should be wholly formal, instead of broadly cultural (as under the old Extension movement), and should not carry credit.

For the two-year community colleges and their students, the advantages would be immense. The very difficult problem of accreditation—bound to plague these institutions as soon as the question of transfer arises—will be disposed of. Local control over admissions and instruction can be maintained at the same time that standards will be supported and degrees made meaningful, for employment or transfer to higher institutions.

This is a new kind of testing, obviously, but necessity and the appearance of new challenges force its consideration. I do not want to press the "novelty" of the proposal too far, of course; for certification