DOCUMENT RESUME

ED 173 439 TM 009 578

TITLE Proceedings of the Invitational Conference on Testing Problems (25th, New York, New York, October 31, 1964).
INSTITUTION Educational Testing Service, Princeton, N.J.
PUB DATE 31 Oct 64
NOTE 147p.

EDRS PRICE MF01/PC06 Plus Postage.
DESCRIPTORS Asian History; *Culture Free Tests; *Factor Analysis; Foreign Countries; Intelligence Factors; *Intelligence Tests; Mathematical Models; Psychological Testing; *Psychometrics; *Public Opinion; Social Differences; Test Construction; Testing; *Testing Problems; Test Interpretation
IDENTIFIERS China

ABSTRACT

This conference focuses upon various aspects of testing within a society, and upon technical advances in measurement. Philip H. DuBois's paper deals with China, a society dominated by tests, from 1115 BC--1905 AD. Donald W. Fiske discusses the effect of testing on the individual in modern America in his paper, "The Subject Looks at Psychological Tests." Three psychometricians discuss several new mathematical approaches to the analysis of test data: Louis Guttman, "The Structure of Interrelations Among Intelligence Tests;" Henry F. Kaiser, "Psychometric Approaches to Factor Analysis;" and Ledyard R Tucker, "Experiments in Multi-Mode Factor Analysis." Omar Khayyam Moore speaks on "Technology and Behavior." The four final papers dealt with various issues related to the testing of the culturally different: "Measurement Issues in the Counseling of the Culturally Disadvantaged," Martin Hamburger; "Principles of Developing Tests for the Culturally Different," William E. Coffman; "The Measurement of Environments," Richard Wolf; and "Social Class and Cultural Group Differences in Diverse Mental Abilities," Gordon Fifer. (BH)

TM009 578

ED173439

[redacted]

Library of Congress Catalogue Number: 47-11220
Printed in the United States of America

Invitational Conference on Testing Problems

October 31, 1964
Hotel Roosevelt
New York City

CHESTER W. HARRIS
CHAIRMAN

EDUCATIONAL TESTING SERVICE
Princeton, New Jersey
Berkeley, California

4

ETS Board of Trust
1964-65

ETS Officers

Henry Chauncey, *President*
William W. Turnbull, *Executive Vice President*
C. Russell de Burlo, Jr., *Vice President*
Henry S. Dyer, *Vice President*
John S. Helmick, *Vice President*
Robert J. Solomon, *Vice President*
G. Dykeman Sterling, *Vice President and Treasurer*
Catherine G. Sharp, *Secretary*
Joseph E. Terral, *Assistant Vice President*
David J. Brodsky, *Assistant Treasurer*
John Graham, *Assistant Treasurer*

Foreword

The themes of the 1964 Invitational Conference on Testing Problems focused upon various aspects of testing within a society and upon technical advances in measurement. A report in Session I on the effect of testing on the Chinese empire provided an interesting contrast with a paper on the effect of testing on the individual in modern America. Session III was devoted to the problems of testing across cultures within a society and was explored from the viewpoint of the educator, the test-maker, the sociologist, and the psychologist. In contrast, to these rather broad aspects of measurement, Session II featured a discussion by three leading psychometricians of some new mathematical approaches to the analysis of test data. All three sessions resulted in a program that was both lively and absorbing and altogether appropriate for the twenty-fifth anniversary of this annual event.

The success of this program was due largely to the imaginative planning of the chairman, Professor Chester Harris, to whom we are all most grateful. We owe our thanks also to Professor Omar Khayyam Moore for his luncheon address on "Technology and Behavior," and to the other distinguished speakers whose papers we are pleased to publish in these *Proceedings*.

Henry Chauncey
PRESIDENT

Preface

It was my pleasure to be invited to plan the twenty-fifth Invitational Conference on Testing Problems, sponsored by Educational Testing Service. To help me determine what might be appropriate papers and topics for discussion at this conference, I turned to a number of persons throughout the country for advice. Their responses were prompt and extremely helpful; in fact, I had little more to do than to select and then order into a reasonable sequence a limited number of topics from the wealth of ideas I received. (The suggestions for topics that we were not able to crowd into a one-day meeting have been preserved and passed on to future chairmen.) Thus, the character of this particular program in 1964 was shaped largely by a number of other persons, to whom I am indebted.

My second task was to secure the speakers. Again, my correspondents had been extremely helpful in identifying for me exactly the right person for each of a number of topics. I then went to work to secure acceptances from these "right" persons. Their participation gave the conference its strength; to them I am deeply indebted.

My third task was to preside at the two morning sessions and at the afternoon session. The interested audience responding to interesting and provocative speakers made these duties especially enjoyable. Thus, my final expression of indebtedness is to the many who came to hear.

The first morning session presented two quite different papers dealing with the effects of testing. The second morning session was devoted to three somewhat technical papers that seemed to me to represent three important, and fairly recent, lines of development that undoubtedly will influence our analyses of test data. Dr. Henry Chauncey presided at the luncheon meeting and introduced Dr. O. K. Moore, who spoke on "Technology and Behavior." The afternoon session included four papers dealing with various issues in the problem of testing the culturally "different."

Since the record of the twenty-fifth Invitational Conference on Testing Problems is printed here for all to read, it is unnecessary to say more. These men speak for themselves; I hope they will speak to you.

Chester W. Harris
CHAIRMAN

Contents

Contents continued

Session I

Theme:
Effects of Testing

A Test-Dominated Society: China, 1115 B.C.–1905 A.D.

PHILIP H. DUBOIS
Washington University

Our negative enthusiasm for the present government in Peking should not lead us to a lack of appreciation for great Chinese achievements of the past. They have been many.

It is often said that the Chinese invented gunpowder and, quite humanely, used it to frighten, rather than to kill, their enemies.

Certainly they solved the problem of diverse languages with the remarkable invention of a common written language — a code by which peoples who could not communicate with one another orally were able to communicate freely by means of writing. This invention was so successful that the Chinese came to regard themselves as a single people.

They invented paper, which the West did not know how to make until some Chinese papermakers were captured by Arabs at Samarkand in 751 A.D. They invented printing. They developed the arts. But, more importantly for our purposes this morning, they invented the psychological test, applying it to government, the very framework of their society, in such a manner that the test-makers, in effect, determined over many centuries much of the format of Chinese society.

The prolonged and intensive Chinese experience with testing seems to have been completely ignored by contemporary psychometricians. In none of the writings on psychometrics with which I am familiar is there any mention of some 3,000 years of examinations in the Chinese empire. This is rather surprising because in civil service procedures, it is easy to trace the continuity of Eastern and Western methods. Continuity between Western educational and psychological examining methods on the one hand and Chinese civil service testing on the other is more difficult to demonstrate, but some influence is probable.

Even if Western psychometrics had been completely independent of Chinese testing, the Eastern experience would have been of great interest to us. It affords the one historical example of a society in which examining methods introduced to attain certain restricted objectives actually began to determine many characteristics of the society itself.

It should be noted that through the ages the Chinese empire, unlike the West, did not have a numerous hereditary aristocracy to constitute its governing class. The chief way to a political career was through passing a series of examinations in which competition was very severe. Moreover, China lacked another invention of the West: the university. The learned Chinese was one who had been successful in passing competitive examinations, and whose success brought changes in his attire and in his title as well as public recognition of his abilities, and employment in government service.

For long periods of time the system worked very well indeed. Only occasionally were examinations suspended, one notable period being the time in which the Mongol emperors ruled in Peking. (The accounts of Marco Polo, who spent a number of years in China during their rule, make no reference to the Chinese civil service examining procedures.)

The Chinese scholar seems to have been a reasonably successful public administrator. Public office was often distributed by lot among the mandarins who had passed three successive sets of examinations.

Millions of men prepared for the tests, often for decades, and relatively few achieved final success. The selection ratio was so small that the tests themselves would not have had to be very valid in order to be useful. That they were useful is perhaps indicated by their long history and by the fact that for many centuries, with relatively few interruptions, the government of the Chinese empire preserved internal peace, provided security from many would-be invaders, and permitted a flowering of civilization that in many respects was far more advanced than that prevailing contemporaneously in the West.

The earliest development seems to have been a rudimentary form of proficiency testing. About the year 2200 B.C., the emperor of China is said to have examined his officials every third year. After three examinations, he either promoted or dismissed them from the service. There seems to be no record of the exact content nor of the methods of testing, but the precedent of periodic examinations was to continue for many generations.

A thousand years later in 1115 B.C., at the beginning of the Chan dynasty, formal examining procedures of candidates for office were established. Here the record is clear. Job sample tests were used requiring proficiency in the five basic arts: music, archery, horsemanship, writing and arithmetic. Of the five, at least two, writing and arithmetic, still have validity for public office. Knowledge of a sixth art was also required — skill in the rites and ceremonies of public and social life.

It should be pointed out that this examining system, which was later to be centered upon the Confucian classics, was actually in existence long before the time of Confucius (551-478 B.C.).

While the procedures changed from time to time, and the sources to which I have had access are somewhat contradictory, a few dates seem to be clear. In 165 B.C., by which time Confucian ethics had become current, moral standards were introduced in the selection of competitors. District magistrates were required to send to the capital candidates who had acquired a reputation for filial piety and integrity. Those whose moral character had been sufficiently attested were then examined with respect to their intellectual qualifications. At this time, the test included not only measures of the six arts, but also familiarity with the geography of the empire, civil law, military matters, agriculture, and the administration of revenue.

After 622 A.D., open competitive examinations took place at more or less regular intervals. By 1370 A.D. three levels of examinations were well established. The candidate who passed the examination in his district became eligible to take a test at the provincial capital, and those successful at the provincial capital were eligible for final examinations in Peking. For about 500 years the system was stable and a description by William A. P. Martin in 1870 is pertinent:

> . . . The candidates for office, — those who are acknowledged as such, in consequence of sustaining the initial trial, — are divided into the three grades of *siu-ts'ai*, *chu-jin*, and *tsin-shi*, — "Budding Geniuses," "Promoted Scholars," and those who are "Ready for Office." The trials for the first are held in the chief city of each district . . . They are conducted by a chancellor, whose jurisdiction extends over an entire province, containing, it may be, sixty or seventy such districts, each of which he is required to visit once a year, and each of which is provided with a resident sub-chancellor, whose duty it is to examine the scholars in the interval, and to have them in readiness on the chancellor's arrival.

About two thousand competitors enter the lists, ranging in age from the precocious youth just entering his teens up to the venerable grandsire of seventy winters. Shut up for a night and a day, each in his narrow cell, they produce each a poem and one or two essays on themes assigned by the chancellor, and then return to their homes to await the bulletin announcing their place in the scale of merit. The chancellor, assisted by his clerks, occupies several days in sifting the heap of manuscripts, from which he picks out some twenty or more that are distinguished by beauty of penmanship and grace of diction. The authors of these are honored with the degree of "Budding Genius," and are entitled to wear the decorations of the lowest grade in the corporation of mandarins. The successful student wins no purse of gold and obtains no office, but he has gained a prize, which he deems a sufficient compensation for years of patient toil. He is the best of a hundred scholars, exempted from liability to corporal punishment, and raised above the vulgar herd . . .

Once in three years these "Budding Geniuses," these picked men of the districts, repair to the provincial capital to engage in competition for the second degree, — that of *chu-jin*, or "Promoted Scholar." The number of competitors amounts to ten thousand, more or less, and of these only one in every hundred can be admitted to the coveted degree. The trial is conducted by special examiners sent down from Peking and this examination takes a wider range than the preceding. No fewer than three sessions of nearly three days each are occupied instead of the single day for the first degree. Compositions in prose and verse are required, and themes are assigned with a special view to testing the extent of reading and depth of scholarship of the candidates. Penmanship is left out of the account, — each production, marked with a cipher, being copied by an official scribe, that the examiners may have no clew to its author and no temptation to render a biased judgment.

The victor still receives neither office nor emolument; but the honor he achieves is scarcely less than that which was won by the victors in the Olympic games. Again, he is one of a hundred, each of whom was a picked man; and as a result of this second victory he goes forth an acknowledged superior among ten thousand contending scholars. He adorns his cap with the gilded button of a higher grade, erects a pair of lofty flag-staffs before the gate of his family residence, and places a tablet over his door to inform those who pass by that this is the abode of a literary prize-man. But our "Promoted Scholar" is not yet a mandarin, in the proper sense of the term. The distinction already attained only stimulates his desire for higher honors, — honors which bring at last the solid recompense of an income.

In the spring of the following year he proceeds to Peking to seek the next higher degree, the attainment of which will prove a passport to office. This contest is still with his peers, that is, with other "Promoted Scholars," who like himself have come up from all the provinces of the empire. But the chances are this time more in his favor, as the number of prizes is now tripled, and if the gods are propitious his fortune is made . . . If his name appears among the favored few, he not only wins himself a place in the front ranks of the lettered, but he plants his foot securely on the rounds of the official ladder by which, without the prestige of birth or the support of friends, it is possible to rise to a seat in the grand council of state or a place in the Imperial Cabinet. All this advancement presents itself in the distant prospect, while the office upon which he immediately enters is one of respectability, and it may be of profit. It is generally that of mayor or sub-mayor of a district city, or sub-chancellor in the district examinations, — the vacant posts being distributed by lot, and therefore impartially, among those who have proved themselves to be "ready for office."

Before the drawing of lots, however, for the post of a magistrate among the people, our ambitious student has a chance of winning the more distinguished honor of a place in the Imperial Academy. With this view, the two or three hundred survivors of so many contests appear in the palace, where themes are assigned them by the Emperor himself, and the highest honor is paid to the pursuit of letters by the exercises being presided over by his Majesty in person. Penmanship reappears as an element in determining the result, and a score or more of those whose style is the most finished, whose scholarship the ripest, and whose handwriting the most elegant, are drafted into the college of Hanlin, the "forest of pencils," a kind of Imperial Institute, the members of which are recognized as standing at the head of the literary profession. These are constituted poets and historians to the Celestial Court, or deputed to act as chancellors and examiners in the several provinces.

But the diminishing series in this ascending scale has not yet reached its final term. The long succession of contests culminates in the designation by the Emperor of some individual whom he regards as the Chuang-Yuen or model scholar of the empire . . . Provinces contend for the shining prize, and the town that gives the victor birth becomes noted forever. Swift heralds bear the tidings of his triumph, and the hearts of the people leap at their approach. We have seen them enter a humble cottage, and amid the flaunting of banners and the blare of trumpets announce to its startled inmates that one of their relations had been crowned by the Emperor as the laureate of the year. And so high was the estimation in which the people held the success of their fellow-townsman, that his wife was requested to visit the six gates of the city, and to scatter before each a handful of rice, that the whole population might share in the good fortune of her household . . .

It is obvious that which excites so profoundly the interest of a whole nation must be productive of very decided results. That it leads to the selection of the best talents for the service of the public we have already seen; but beyond this — its primary object — it exercises a profound influence upon the education of the people and the stability of the government. It is all, in fact, that China has to show in the way of an educational system. She has no colleges or universities, — if we except one that is yet in embryo, — and no national system of common schools; yet it may be confidently asserted that China gives to learning a more effective patronage than she could have done if each of her emperors were an Augustus and every premier a Maecenas. She says to all her sons, "Prosecute your studies by such means as you may be able to command, whether in public or in private, and when you are prepared, present yourselves in the examination hall. The government will judge of your proficiency and reward your attainments."

Nothing can exceed the ardor which this standing offer infuses into the minds of all who have the remotest prospect of sharing in the prizes. They study not merely while they have teachers to incite them to diligence, but continue their studies with unabated zeal long after they have left the schools; they study in solitude and poverty; they study amidst the cares of a family and the turmoil of business; and the shining goal is kept steadily in view until the eye grows dim. Some of the aspirants impose on themselves the task of writing a fresh essay every day; and they do not hesitate to enter the lists as often as the public examinations recur, resolved, if they fail, to continue trying, believing that perseverance has power to command success and encouraged by the legend of the man who, needing a sewing-needle, made one by grinding a crowbar on a piece of granite.

This quotation from Martin, describing and praising the Chinese testing system, is by no means unique. The use of competitive examinations for the selection of state officials was praised by many Western observers and writers, including Voltaire. In fact, it is clear that initially all civil service examining in Europe and in the United States used the Chinese system, directly or indirectly, as a model. Civil service testing was introduced in France as a 1791 reform, only to be abolished by Napoleon. In England the first competitive examinations in connection with public office were instituted for the selection of trainees for the civil service in India by men familiar with the Chinese system. Later, when the question of civil service examinations for Great Britain as a whole was debated in Parliament, the Chinese model was discussed with both favorable and unfavorable comments. As a part of an extensive study, Congressman Thomas A. Jenckes, one of the fathers of the United States Civil Service, wrote 12 pages on the civil service of China.

Westerners seem to have been particularly impressed with the fact that competition was open, that distinction came from merit, and that a highly literate and urbane group of public officials resulted from the examination system.

The great crisis in Chinese affairs came, of course, when the Chinese realized that they were militarily inferior to the West. They quickly discovered that equality in military power could not be achieved without modern science and technology. Accordingly, technological schools and universities were set up, but as long as the civil service examinations, which were largely literary in character, continued to be the way for an ambitious man to have a career, modern education was not sufficiently attractive. Consequently, in 1905, the Chinese examination system was abolished as a reform measure.

So much for a description and a bit of the history of an ancient Chinese venture in psychological examining as a tool of government. What can be said about their testing techniques from the point of view of the modern psychometrician? In the first place, I find no evidence to indicate that they invented either the multiple-choice format, the test-scoring machine, or item analysis. They did, however, recognize that a relatively short performance under carefully controlled conditions could yield an estimate of the ability to perform under less rigorously controlled conditions and for a longer period of time. I think there is no doubt that the procedure selected capable public servants.

They recognized the problem of objectivity, concealing candidates' names and sometimes using a bureau of copyists to copy examination material before it was graded. In some cases, tests were read by two independent examiners who handed their sealed evaluations to a third examiner who reconciled any differences. Scores seem to have been in terms of rank order.

The need for uniformity in testing conditions was well recognized. Considerable attention was given to proctoring the examination halls, which were large and permanent installations consisting of hundreds of small cells. Sometimes candidates died during the rigors of the examinations, which went on day and night.

In this year 1964, when psychological tests are being used more and more extensively at critical points in the careers of all our citizens, we will do well to consider their effects on individuals, on specific institutions, and on society. The long Chinese experience is a pertinent case history. It is a plausible hypothesis that much of the great strength of the Chinese empire came from the intellectual vigor of men who were bright enough to compete in examinations requiring the writing of poems and "eight-legged essays."

Certainly the opportunities that were opened up by success in the examinations stimulated millions of individuals to long years of scholarship. Perhaps the greatest drawback was that the scholarship was not always pertinent. In the nineteenth century, China suddenly found herself surpassed in technology by the West. While Chinese civilization had been relatively static, Westerners invented the steam engine, the power loom, and the ironclad. It was then that the Chinese, in order to preserve their country and their institutions, began to desire progress according to the Western model. At that time the age-old examining system was discovered to be a hindrance.

So far, with 60 years of experience, we Westerners have not found our psychological examining a hindrance. But it is becoming increasingly apparent that our test-makers, like those of ancient China, establish goals for individuals and influence the shape of social institutions. Item writers as well as song writers mold the patterns of a culture.

SOURCES

1. Etienne Zi (Siu), S. S., *Pratique des Examens Littéraires en Chine*. Chang-Hai: Imprimerie de la Mission Catholique, 1894.
2. Franke, Wolfgang, *The Reform and Abolition of the Traditional Chinese Examination System*. Cambridge: Harvard University Press, 1960.
3. Hsin T'ang Shu, *Le Traité des Examens*. Paris: Librairie Ernest Leroux, 1932.
4. Jenckes, Thomas A., "Civil Service of the United States" Report No. 47, 40th Congress, 2nd Session, May 25, 1868.

5. Kracke, E. A., Jr., *Civil Service in Early Sung China, 960-1067*. Cambridge: Harvard University Press, 1953.
6. Martin, William A. P., "Competitive Examinations in China," *North American Review*, 1870, 111, 62-77.
7. Ssu-rü Têng, "Chinese Influence on the Western Examination System," *Harvard Journal of Asiatic Studies*, 1943, 7, 267-312.
8. Ssu-rü Têng and Fairbank, John K., *China's Response to the West*. Cambridge: Harvard University Press, 1954.

ACKNOWLEDGMENT

Dr. Stanley E. Spector, chairman of the Department of Chinese and Japanese at Washington University, suggested certain of these references. His helpfulness is gratefully acknowledged.

The Subject Looks at Psychological Tests*

Donald W. Fiske
University of Chicago

This study originated in a methodological question. Operations for measuring personality are less satisfactory than those for assessing abilities and aptitudes. Does this relative inadequacy stem from the heterogeneity of subjects' reactions to tests? More specifically, are reactions to personality tests more diverse than those to cognitive tests? If so, does such greater diversity help to account for this inadequacy?

The initial purpose was to explore reactions to tests by means of the simulated subject technique which Orne (1962) developed to evaluate the adequacy of the experimental conditions used in some areas of psychological research. To permit greater generalization of the findings, the survey method was selected. During the planning stage, it became obvious that other significant information could also be obtained at little additional cost. The actual study sought answers to five basic questions:

1. What experiences have people in this country had with psychological tests?
2. What views about psychological tests have they encountered?
3. How do laymen evaluate psychological tests of various kinds?
4. How do people feel about taking tests? (How do they feel about being given a test? How do they feel while taking it?)

* This study was supported by Grant MH-06582 from the National Institute of Mental Health, United States Public Health Service.

I am indebted to Naomi Berne, Jere Brophy, Castellano Turner, and Thomas Tyler for their assistance in the statistical analyses. I am also indebted to the staff of the National Opinon Research Center, especially Carol Bowman, Galen Gockel, Harold Levy, and Patrick Page.

5. What are the methodological implications of the answers to these four questions, especially the latter two?

Method

The Sample

The study of these questions was conducted with the assistance of the National Opinion Research Center whose excellent organization conducted the interviews and did the coding and tabulating. A national sample of 589 respondents was obtained. It was limited to persons 21 through 64 years of age who had completed at least the sixth grade. The sampling was intentionally biased to obtain about one-third more respondents who had been to college than would be found in the national population: there were 27 per cent in that category rather than 20 per cent.

The sample was quite representative of the country at large with respect to sex, race, and age. The correspondence on place of residence and on occupational group was close enough to permit confident approximate generalization to the national population.

The Procedure

The interview generally took about an hour. The questions asked will be brought out in the presentation of the findings. However, the simulated subject procedure requires detailed description. This part took place during the middle portion of the interview.

Each respondent was assigned in advance to one context and to two tests. The Job context had instructions asking the subject to suppose that he was applying for a job and that he was given a test. This supposed test started with the instructions and the several items that were actually presented to him; he was asked to respond to these items. In the Research context, the instructions asked the subject to suppose that a college professor had asked him to take the test for some research he was doing, and to answer anonymously. The context was referred to again when the second test was given; it was also mentioned twice during the exploration of his reactions to each test.

The six incomplete, simulated tests had titles and instructions closely resembling those from which they were borrowed. The six included the following: an intelligence test based on the ACE; an interest test derived from the Strong; an innocuous personality questionnaire adapted from the Thurstone Temperament Schedule; a less innocuous Multiphasic Inventory based on the MMPI; an Incomplete Sentences Test; and an instrument with two ink-blots. The first five had seven to eleven items each.

(The standard test for the significance of differences in proportions was used when the comparison was between two parts which together comprised the total sample. For comparisons of pairs within a set of several proportions, the studentized range test provided the criterion for each comparison, with recognition of the relative severity of this standard.)

The method appears to have worked fairly well. Since it seemed to make little difference whether a test came first or second, all respondents given each test within the same context were grouped together. Thus we had about one hundred subjects per test per context. Notwithstanding the arbitrary assignment to these experimental conditions, the six groups within each context were reasonably comparable on demographic variables, as were the two main context groups themselves.

The experimental technique appears to have been fairly effective. This impression is supported by internal evidence, such as the direction of group differences. Also, the subjects reported that they maintained the instructional set for the specific context fairly well. As might be expected, however, a handful of subjects did object to their being asked to take the tests. The level of affect aroused in the subjects seems to have been below that which the subjects recalled having experienced during their last actual test.

Findings

Experiences with Tests and Evaluations of Tests

Almost two-thirds of this sample of adults reported having taken at least one test, other than tests taken while in some school. Half of the total group had had a test when applying for a job. Only 15 per cent had taken tests for admission to college, these comprising a little over half of those who went to college. Of those going to other schools, smaller proportions took tests for admission.

Exposure to views about tests was reported by 62 per cent. Many remembered having read something, some had heard discussions of tests, and a smaller number had taken courses in which tests were presumably mentioned. The views recalled were more favorable than unfavorable.

When asked to select the best way to find out whether a person is qualified for something (such as a job or admission to a school), the interview method received most first choices, especially among those with the most college education! Tests came second, receiving one-third of the first choices and one-quarter of the second choices. Application blanks and references were preferred by much smaller proportions of the sample.

Respondents were given four choices, from very good to very poor, to indicate their evaluation of tests for finding out about a person's abilities and skills, his likes and dislikes, how he gets along with other people, and what his personal problems are. The results are given in Table 1. More than 75 per cent of the sample evaluated tests as good

Table 1

Evaluations of Tests for Different Characteristics (In Per Cents)

How good or poor do you think tests would be for finding out:	Very Good	Fairly Good	Fairly Poor	Very Poor
. . . . what a person's *aptitudes, skills, and abilities are?*	35	51	10	5
. . . . what a person *likes and dislikes?*	35	44	10	11
. . . . how a person *gets along with other people?*	23	30	23	24
. . . . what someone's *personal problems are?*	20	28	22	30

for assessing abilities and for measuring likes and dislikes. In contrast, only about half rated tests as good for finding out about a person's social relations or personal problems. I suppose personality testers can be pleased that the proportions are even that large.

Interpretations of Tests

After these general opinions had been explored, each subject was given two simulated tests. The final question asked about each was: "What do you think that test is supposed to find out?" The replies are classified in Table 2.

Table 2

Replies to the Question:
"What do you think that test is supposed to find out?"
(In Per Cents)

Simulated Test

Classification of Responses	"Psychological Examination" (Intelligence)	"Preference Inventory" (Interests)	"Likes and Dislikes" (Temperament)	"Multiphasic Personality Inventory"	"Incomplete Sentences"	"Ink-Blots"
Intelligence	31	2	5	4	12	12
Education and knowledge	9	1	0	0	2	3
A specific ability	7	2	2	2	4	3
Verbal skills	3	0	0	0	4	0
Personality	4	6	25	31	24	15
Stability	0	1	4	14	1	5
Getting along with others	0	1	4	0	1	0
A specific trait	0	0	3	0	1	0
Interests	0	51	8	5	2	2
Imagination	0	0	0	0	0	17
Multiple answer	21	19	27	29	24	14
Irrelevant and miscellaneous	7	6	6	4	10	6
Don't know	17	10	15	10	15	23

The majority hazarded specific guesses and these are the ones with which we shall be concerned. The intelligence test was interpreted as such by a majority of those giving specific replies. However, a few thought that it measured personality! The interests test received an even clearer consensus of correct judgments. The so-called "Inventory of Likes and Dislikes" (borrowed from the Thurstone Temperament Schedule) was given more varied interpretations, with personality being the modal specific response. A similar pattern of perceptions was obtained for the Multiphasic Inventory, but stability was more frequently mentioned for this test. The same general pattern recurred for the Incomplete Sentences, although intelligence and verbal skills received more mentions and stability received almost none. While the modal response for the Ink-Blots was imagination, almost as many saw it as measuring personality and, like the other projective test, a number thought it measured intelligence.

It is obvious that the majority of those giving specific perceptions of these tests made interpretations which those who developed the tests would consider highly appropriate and accurate. Not even the general purpose of projective tests could be considered as disguised. Examination of the simulated tests indicates that the respondents based their interpretations primarily on the content of the items, rather than on the names of the tests or the instructions.

Reactions to Simulated Tests

After each subject completed the brief simulated test, he was asked how he would feel about taking it in the Job or Research context which had been portrayed to him before the testing. This open-ended question elicited a diversity of reactions. A quarter of the sample expressed indifference. In addition, almost half of the Job group and more than a quarter of the Research group expressed some kind of negative reaction, with criticisms of the test predominating. Very few expressed uneasiness or discomfort; very few said tests are prying or can be faked. (For each context group, less than 5 per cent of the replies fell in each of these latter four categories.)

In the attempt to uncover hidden negative reactions, respondents were also asked whether other people would feel differently about taking the test, and if so, how they would feel. While close to 10 per cent projected some kind of negative reaction, the majority of those saying that others might feel differently had the good psychological sense to observe that it would depend upon the person. Although the layman's psychological understanding is naive and untutored, we once again find that it is rather sound.

Nomothetic data on reactions were also obtained by asking the subjects to indicate which of 13 phrases described and which did not describe how they felt while taking the test. (See the first two columns of Table 3.) "Didn't mind," "Interested," and "Curious" were selected by majorities. In addition, denials of most of the phrases indicating negative feelings were made by majorities.

Of the many comparisons between the Job and the Research context groups, only a few were significant. These are judged to be real differences in most instances since they tend to be consistent and in the expected direction. For example, the Research group expressed more willingness to cooperate. More of the Job group commented that the test had a poor relation to job performance and gave negative comments of other kinds. But also, more of the Job group said tests were good. (Once again, the relative frequencies are small and these findings show diversity within the group rather than inconsistency within the individual.)

Table 3

Subjects' Feelings while Taking Tests (In Per Cents)

		Feelings While Taking Simulated Test		Recollection of Feelings While Taking Last Actual Test	
				Purpose	
	Describes	Job Context	Research Context	Application for Job	Application for Admission to School or College
I was *interested*	Does	59	64	69	74
	Does not	14	10	7	4
I was *curious*	Does	55	61	35	52
	Does not	14	10	17	13
I felt *uncomfortable*	Does	12	10	28	24
	Does not	53	60	34	37
I *didn't mind*	Does	69	75	42	44
	Does not	12	8	19	19
I was *annoyed*	Does	6	3	11	9
	Does not	61	67	53	56
I felt *like laughing*	Does	17	16	2	4
	Does not	49	54	65	61
I was *anxious*	Does	12	10	55	54
	Does not	44	46	16	19
I was *irritated*	Does	4	2	4	4
	Does not	68	70	58	54
I was *bored*	Does	5	2	2	4
	Does not	63	70	63	63
I felt *it was silly*	Does	25	14	6	11
	Does not	44	56	68	54
I was *frustrated*	Does	5	4	8	9
	Does not	55	62	46	44
I was *tense*	Does	8	10	56	48
	Does not	57	60	19	28
I felt *pleased*	Does	23	27	27	24
	Does not	28	24	25	20
N		(297)	(292)	(259)	(54)

Reactions to the several pairs of tests were examined separately for each context. Few significant differences were found and in no case was the same pairing significant in both contexts. Since the statistical criterion may have been too stringent, and since there may have been interaction between test and context, some of the significant differences will be mentioned. Most involved the Ink-Blots Test. More often than for other tests, it was seen as ridiculous and it made subjects feel embarrassed. The intelligence test was more likely to make subjects feel uneasy and the Multiphasic Inventory elicited more comments that tests can be faked.[7]

Reactions to Last Actual Test

After the simulated tests, the subjects were asked about their experiences with actual tests and their most recent actual test was identified. For many, this was a test taken in school many years ago. Even for those taking tests when applying for a job, the median time interval was six years. Therefore, the reported feelings about these last actual tests must be seen as filtered through years of memory. Reports for tests taken when applying for a job and for admission to a school or college are given in the last two columns of Table 3.

Granted this limitation stemming from remoteness in time, the descriptions of feelings on the thirteen items show a number of striking differences from the feelings about the simulated tests. With one minor exception, for each of the several real contexts, a majority of the subjects reported feeling anxious and tense; these proportions are several times those for the simulated tests. Reactions to the actual tests more frequently included feeling uncomfortable and frustrated while the feelings about the simulated test were more often "I didn't mind," "I felt like laughing," and "it was silly."

Thus, much more negative affect was reported, in long retrospection, for *actual tests*. In comparison, while the free responses to the question about how they *would feel* about taking the simulated tests in the indicated context did produce more negative reactions than positive, these were typically criticisms of tests rather than reports of negative feelings. Finally, we have the reports of how the subjects *felt while taking* the simulated tests, reports which are more positive than negative. Thus, we find increasing amounts of negative reactions with increasing levels of potential consequences, starting with how the subjects felt about taking short, obviously incomplete mock tests in the survey interview, going on to how they thought they would feel about taking a complete test in real life, and ending with how they recalled having actually felt while taking a test which often could affect their lives. (Data from another group given a similar, but briefer, procedure are consistent with this trend: for the reports of these subjects on how they would feel about taking the simulated test in a given context, the frequencies for "anxious" and "tense" are intermediate between those for the actual test and those for the simulated test in the survey group.)

On the one hand, it is clear that the simulation of context worked to only a limited degree: the simulated tests were probably not sufficiently realistic and the subjects were not given enough help in the role-taking we tried to have them perform. On the other hand, the findings clearly suggest that being tested, even in an artificial situation, has some impact on people, and that the degree of impact increases with the potential consequences for the subjects.

Methodological Implications

Our findings indicate that a person who is given a psychological test will usually have a reasonably good idea about the general characteristic which is being tested. However, his feelings and reactions will be determined only in small part by the perceived nature of the test: rather, he will be reacting primarily to the fact of being tested. If a great deal depends upon his performance, he is likely to feel anxious and tense, and quite understandably so. But even if the test results will not affect him one way or the other, there is still a good chance that he will feel uneasy, perhaps even resentful.

While such feelings do occur commonly, other feelings and perceptions may also be present. The subject may have great faith in tests, or he may see them as silly and invalid; he may even feel that testing is wrong. Then again, he may enjoy taking the test. The diversity of subjects' perceptions of tests and feelings about being tested must be considered in designing the procedures for administering a test, and particularly in formulating the instructions.

In planning this research, it was anticipated that the diversity of perceptions and reactions would be least for the intelligence test, and would increase for the other tests, through the test of interests and the other questionnaires to a maximum for the projective tests. These differences were not found. Hence the differences in psychometric adequacy between intelligence tests and personality tests cannot be explained on this basis. It now seems that the critical difference is in the tasks involved.

Tests of ability present the subject with a clearly defined task, one which he understands and which he is able and willing to perform. In addition, they present a set of stimuli, each of which is interpreted by the several subjects in the same way, at least with respect to how their responses to it should be determined. By these means, tests of ability minimize the potential intrusion of such irrelevant sources of variance as the diversity among subjects' evaluations of tests and the variation among their subjective feelings while taking the test.

On the other side, the task posed by the typical personality test is much less precisely defined. The instructions, though intelligible, do not furnish the subject with explicit criteria for selecting his response: if, after making his response, he were to review it, he could usually not be confident that he had given the response called for in the instructions. In addition, if the test is verbal, there is vagueness and ambiguity in the items themselves and in the permitted response alternatives (e.g., see Benton, 1935; Eisenberg, 1941).

This lack of structuring and of consequent restraints in personality tests permits the subject to perceive the task as a whole, and each item in particular, in a way determined in part by his reactions to the total testing situation. Hence the differential reactions found in this research can affect the way in which a person responds to a test of this kind. We can expect that different subjects will approach the test differently. If so, the scores derived from their responses will not be comparable because the experimental conditions have not produced similar constraints in the several subjects (cf. Fiske & Butler, 1963).

Summary

A majority of the adults in this country have taken a psychological test outside of their schoolwork. A majority have also been exposed to views about tests. Although tests are not of great concern to the layman, he makes highly appropriate interpretations of the purposes of tests. He typically evaluates tests favorably, especially for determining a person's qualifications, for assessing abilities and skills, and for determining likes and dislikes.

This study found a considerable range of reactions to the condition of being tested. The particular test is unimportant, in contrast to the effects associated with being tested. Negative reactions, and especially reports of negative affect, increase in frequency with the significance of the situation to the subject.

It is argued that tests with explicit tasks and definite criteria for the subject to use in selecting his responses can provide sufficient structure and constraints to minimize the unwanted variance which differential reactions would otherwise contribute to obtained test scores.

REFERENCES

1. Benton, A. L., "The Interpretation of Questionnaire Items in a Personality Schedule." *Archives of Psychology,* (N.Y.), 1935, No. 190.
2. Eisenberg, P., "Individual Interpretation of Psychoneurotic Inventory Items." *Journal of General Psychology,* 1941, 25, 19-40.
3. Fiske, D. W., and Butler, J. M., "The Experimental Conditions for Measuring Individual Differences." *Educational and Psychological Measurement,* 1963, 249-266.
4. Orne, M. T., "On the Social Psychology of the Psychological Experiment: With Particular Reference to Demand Characteristics and their Implications." *American Psychologist,* 1962, 17, 776-783.

Session II

Theme: Technical Advances

The Structure of Interrelations among Intelligence Tests

Louis Guttman
The Hebrew University and the Israel Institute of Applied Social Research

Recently I have been working on two different kinds of problems which are not unrelated. One is substantive, dealing with certain data of intelligence tests; the other is computational. The substantive problem has concerned me over a number of years. A few months ago, it finally seemed tractable, although incompletely solved. The data were now organized in a certain fashion, and my plan was to present them in this manner. But something happened after I arrived in Ann Arbor (to serve as Visiting Professor for this academic year).

Just before leaving Jerusalem, I had finally succeeded in working out the basic equations for the computing problem, providing a general nonmetric technique for defining Euclidean spaces for data. This sounds very forbidding, but is actually very simple in its outcome. Small empirical examples were worked out on a desk calculator; these and the equations will appear shortly in *Psychometrika* (6). Larger problems, like the substantive data of this paper, are more conveniently done on an electronic computer. Dr. James Lingoes and I have been collaborating recently on developing an appropriate IBM 7090 program. The week before this conference, the computer was fed the data, using the new nonmetric program. The results were so startling, I tore up the intended speech and must present a different one!

The general problem is that of the structure of interrelations among mental tests. It has been tackled from several points of view in the past, one of which is the definitional aspect. For several years some colleagues and I have been working on a definition of the universe of mental abilities. Recently we arrived at a proposal, which is in press in Jerusalem (4). It reads as follows:

A Faceted Definition of Intelligence

An act of a subject is *intelligent* to the (extent) to which it is classified by a (tester) as (demonstrating) a *correct* perception of an unexhibited *logical* (aspect) of a (relation) intended by the tester, on the basis of another (exhibited) *logical* (aspect) of that relation that is correctly perceived by the subject.

Israel is not as old as China; so it is not yet a "test-dominated society" so interestingly described by Dr. Philip DuBois earlier this morning. But we are getting more involved in testing. Two testing programs with which I have been associated have created for me the need to clarify a number of the issues and have led to this proposal. A full explanation and exploration of the above definition could take several weeks or months. But we need grasp only a few aspects of it for our purposes this morning.

This definition is called "faceted" because it is in terms of what I call "facets," namely, the parenthesized variables. In essence, it says that testing is a process of communication between the tester and the subject, and that to characterize an intelligence test item we have to characterize both the stimulus and the response.

In working with pencil-and-paper tests (as testers have for these many years), if we ask how a tester can communicate with a subject, the answer appears to be that there are three basic kinds of "languages" at his disposal. He can use *words* — the mother tongue of the subject and of the tester perhaps. He can use *symbols of a more formalized language* — such as arithmetic or algebra — where not just the symbols are involved but the subject has been educated in the syntax for use of these symbols (e.g., the subject knows what algebra is). Or he can use *pictures* — whether objective or non-objective. These three languages can be used separately or in combination. It is hard to think of any other way of communicating by paper and pencil.

These languages of communication constitute one important facet of the means of "exhibiting logical aspects of relations," that is, means of constructing items. Our point of view is that facets of this sort are *guidelines for item construction*. If you want to construct items, one thing to consider is: what language of communication should be used? Should it be verbal, numerical, or pictorial or some combination of these?

Apart from the language of communication, one has to be concerned about the *substance* being communicated. This again is an important facet of what is being exhibited, or of item construction. We have found it possible and useful to distinguish between "analytical ability" items and "achievement" items just by making a crude dichotomy of the kinds of things which can be exhibited about a relation. There are two kinds of things: a relation has a *rule* and a relation has *elements*. Interpret the word "relation" here in the sense of the theory of sets, of which we need no technical knowledge beyond recognizing the two aspects: a rule and the things which conform to the rule.

An *analytical ability* item is an item in which the subject is asked to deduce the rule. You show him some of the elements which conform to the (unexhibited) rule, and he has to infer the rule from these. For example, analogies are analytical ability items. You exhibit paired elements, such as (dog, puppy), (cow, ?). The subject, in answering this properly by setting "calf" in place of the question mark, shows that he has deduced the rule. Elements were exhibited here which obey a rule, but it was not said explicitly what the rule is; the tester infers from the response that the subject has a correct perception of the rule.

In an *achievement* item, the rule itself is exhibited and assumed to be understood by the subject; the subject has to operate with the elements properly according to that rule. If the subject is asked: "What is the sum of 7 and 5?" the assumption is he knows what addition is. The intended problem is: can he work properly according to the rule of addition, that is, can he do what he knows he ought to do? The question "Who was the first president of the United States?" is an achievement item, assuming the subject knows what "president" means and what the "United States" means. He knows what the relation is — he knows the rule; the question is: can he come up with the proper element which obeys this rule?

I have recently been exploring these ideas by looking over previously published results of others on various batteries of tests, since we have not yet been able to design and administer large batteries in Israel. There are many rich treasures to be found in our literature of correlation matrices for large batteries of tests, but among the richest still are the excellent series of the Thurstones, the very ones from which "primary mental abilities" were first sought by simple-structure multiple-factor analysis (7,8). I had recently been looking at these large batteries again, and had thought to give you today some pictures of portions of their correlation matrices, selecting out *a priori* the tests which clearly obey — or better, are defined by — the two facets just discussed: the facet of language of communication (numerical, verbal, geometrical) and the facet of what is being done with the substance being communicated (analysis or achievement). This gives a six-fold classification of the tests, since each of the three elements of the first facet can go with each of the two elements of the second facet.

The tests can vary on other facets. For example, the subject can be asked either to *recall* or to *recognize* an answer. An attempt was made to hold this particular facet constant by picking out only tests in which checklists were exhibited. A few exceptions were allowed when there were not enough tests to fill in an area. Tests calling only for memorizing were completely omitted. Among verbal tests, the rule in some cases referred to the *sound* of the words rather than their meaning; such tests were also culled.

There are many possible facets. For example, within arithmetic — or even within addition alone — the number of digits per number is a facet. The number of numbers to be added is also a facet. Thus, addition by itself is an infinite domain; one can construct an infinite variety of addition tests which vary among themselves on at least two facets (the number of addends and the number of digits per addend, respectively).

Not all of such further facets could be held constant because existing batteries have not been constructed by taking these systematically into account. Nevertheless, guided by the facet-culling strategy, I had arrived at some pictures for portions of the Thurstone batteries. Seventeen tests from the *Primary Mental Abilities* monograph seemed to fit the culling rules rather well; their intercorrelations are given here in Table 2. Twenty-one tests were selected similarly from the *Factorial Studies of Intelligence* monograph; their intercorrelations are in Table 1.

My hand analyses of these tables revealed certain interesting circumplexes, taking two languages of communication at a time against the achievement-analysis facet. This essentially duplicated, verified, and expanded the findings previously obtained from Guilford's data in reference 4. It was not clear how to unite all these interrelated circumplexes into one picture. The total picture appeared multidimensional and too difficult to handle by my primitive (but effective for one or two dimensions when there is little error) hand manipulations. It seemed best to show you just the subanalyses — the matrices being arranged differently from the present Tables 1 and 2 — and leave open the question of the over-all picture.

Subsequently, Tables 1 and 2 were fed into the University of Michigan computer with the new nonmetric smallest-space-analysis program, and diagrams were drawn from actual Euclidean coordinates that resulted from the calculations. One big surprise was that only *two* dimensions sufficed to describe each of the configurations. The nonmetric technique in general will yield a smaller space than will conventional factor analysis; even so, two dimensions were less than had been anticipated from the previous attempt to organize these data. A second surprise was that, for the first time, we had produced actual empirical pictures of what I have called a *radex*!

Now the concept of the radex was introduced over ten years ago (1). It didn't catch hold for several reasons, one of which was the lack of a good computing technique for plotting a radex. Many empirical examples of parts of radexes, especially those called simplexes and circumplexes, have been published. A simplex is a simple ordering of tests, or of any other set of interrelated variables. Consider, for example: addition, subtraction, multiplication and division. Holding constant the facet of the number of digits involved, there is evidence that addition correlates more with subtraction than with multiplication, and least with division; two tests which are closer in their *a priori* order also correlate higher with each other. In a circumplex, there's a circular ordering of the tests; tests correlate more closely with their immediate neighbors and less with more distant tests around the circle.

In the radex, both orderings happen simultaneously. Tests are in a two, or even higher, dimensional space, radiating out from some center. But still the contiguity relation remains. Tests closer together in the space intercorrelate more highly, and tests farther apart intercorrelate less. The computing problem is: can we find an analytical way of plotting these relations of nearness and distance? This problem has been solved, and the resulting technique applies to a wide variety of problems, not just to radexes nor just to correlation matrices. It is appropriate for a great variety of psychometric problems in which only qualitative notions of distance occur, as well as to data given originally in some metric form.

We can see immediately now a great virtue of this nonmetric approach: you can compare your original data directly with the final diagram without looking at the intermediate calculations. You do not

Table 1

Correlations from Thurstone and Thurstone

FACTORIAL STUDIES OF INTELLIGENCE†

N = 710 Eighth Grade Children

Test		Analytical			Analytical or Complex					Achievement								Achievement				
Name	Code Number	N* 29	V* 33	G* 18	V* 34	N* 56	N* 1	N* 30	N* 38	N (V) 5	V 10	N 37	G (N) 14	G (N) 16	G (N) 15	G 44	G 31	G 28	V 43	V 45	V 47	N 3
High Numbers	29	—	33	41	32	50	32	28	26	33	32	22	22	19	30	26	31	32	20	18	17	22
Letter Grouping	33	33	—	37	44	33	31	29	29	24	34	30	24	20	24	14	27	17	25	23	20	23
Faces	18	41	37	—	32	31	32	39	31	13	26	24	27	31	35	40	52	27	16	16	13	15
Letter Series	34	32	44	32	—	38	31	15	32	40	40	20	17	17	24	20	28	30	34	34	31	21
Three Higher	56	50	33	31	38	—	33	29	27	45	38	44	28	21	27	18	23	26	27	28	26	39
ABC	1	32	31	32	31	33	—	28	20	29	31	33	28	24	27	23	25	22	20	24	21	29
Identical Numbers	30	28	29	39	15	29	28	—	30	03	18	47	36	32	39	31	40	10	00	-01	02	28
Number Patterns	38	26	29	31	32	27	20	30	—	09	20	24	17	17	28	22	29	22	07	05	03	16
Arithmetic	5	33	24	13	40	45	29	03	09	—	46	21	08	-03	09	14	09	25	39	43	36	26
Completion	10	32	34	26	40	38	31	18	20	46	—	29	08	03	07	14	19	24	52	72	57	23
Multiplication	37	22	30	24	20	44	33	47	24	21	29	—	42	33	30	12	15	05	06	13	16	50
Dot Counting I	14	22	24	27	17	28	28	36	17	08	08	42	—	70	50	20	14	03	00	02	01	38
Dot Counting III	16	19	20	31	17	21	24	32	17	-03	03	33	70	—	60	27	19	04	-03	-02	-02	33
Dot Counting II	15	30	24	35	24	27	27	39	28	09	07	30	50	60	—	46	29	17	-01	-03	-02	27
Pursuit	44	26	14	40	20	18	23	31	22	14	14	12	20	27	46	—	45	35	08	06	02	08
Identical Pictures	31	31	27	52	28	23	25	40	29	09	19	15	14	19	29	45	—	31	12	08	08	05
Geometrical Forms	28	32	17	27	30	26	22	10	22	25	24	05	03	04	17	35	31	—	23	17	10	03
Proverbs	43	20	25	16	34	27	20	00	07	39	52	06	00	-03	-01	08	12	23	—	55	44	09
Vocabulary	45	18	23	16	34	28	24	-01	05	43	72	13	02	-02	-03	06	08	17	55	—	67	16
Paragraphs	47	17	20	13	31	26	21	02	03	36	57	16	01	-02	-02	02	08	10	44	67	—	19
Addition	3	22	23	15	21	39	29	28	16	26	23	50	38	33	27	08	05	03	09	16	19	—

†Based on Table 1 (pp. 84-87)

*N, V, and G represent numerical, verbal, and geometric tests. Starred symbols represent analytical tests; unstarred symbols represent achievement tests.

Table 2

Correlations from L. L. Thurstone
PRIMARY MENTAL ABILITIES†
N = 240 University Students

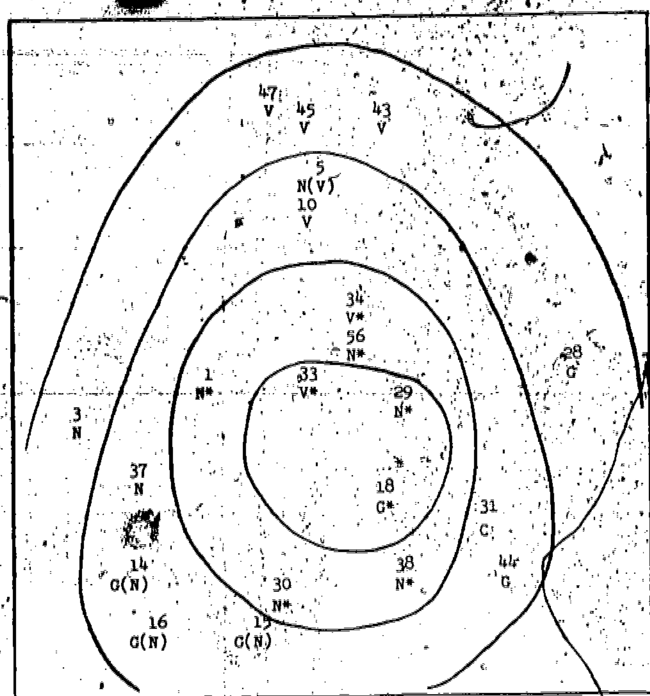
Test Name	Code Number	Analysis V* 41	Analysis N* 37	N* 39	V* 6	G* 24	G* 8	Achievement N 35	G 22	G 19	V 40	Achievement N 34	G 23	G 20	V 12	V 57	Achievement N 33	G 26
Verbal Analogies	41	—	54	48	64	49	46	40	40	49	54	28	46	42	40	46	16	37
Number Series	37	54	—	72	45	48	52	68	41	42	55	41	33	17	29	39	30	16
Arithmetical Reasoning	39	48	72	—	50	53	30	58	52	34	50	62	42	34	34	40	48	16
Verbal Classification	6	64	45	50	—	47	46	41	56	38	42	40	42	37	29	48	28	71
Punched Holes	24	49	48	53	47	—	51	35	62	63	48	24	55	54	37	28	21	30
Figure Classification	8	46	52	30	46	51	—	37	54	40	28	12	42	42	30	30	10	22
Tabular Completion	35	40	68	58	41	35	37	—	33	36	45	56	30	20	29	33	47	16
Lozenges B	22	40	41	52	56	62	54	33	—	64	30	30	57	61	31	28	22	25
Lozenges A	19	49	42	34	38	63	40	36	64	—	46	22	54	62	37	37	15	29
Reasoning	40	54	55	50	42	48	28	45	30	46	—	27	33	30	34	62	10	10
Division	34	28	41	62	40	24	12	56	30	22	27	—	27	27	[illegible]	43	62	19
Surface Development	23	46	33	42	42	55	42	30	57	54	33	27	—	63	[illegible]	28	11	43
Flags	20	42	17	34	37	54	42	20	[illegible]	62	30	27	63	—	22	21	23	21
Disarranged Words	12	40	29	34	29	37	30	29	31	37	34	25	35	22	—	46	35	25
Grammar	57	46	39	40	48	28	30	33	28	37	62	43	28	21	46	—	26	12
Multiplication	33	16	30	48	28	21	10	47	22	15	10	62	11	23	35	26	—	03
Identical Forms	26	37	16	16	71	30	22	16	25	29	10	19	43	21	25	12	03	—

†Based on Table 2 (pp. 110-111)

*N, V, and G represent numerical, verbal, and geometric tests. Starred symbols represent analytical tests; unstarred symbols represent achievement tests.

FIGURE 1

Data from Thurstone and Thurstone
Factorial Studies of Intelligence (Table 1)



N = numerical G = geometric V = verbal * = analytical unstarred = achievement

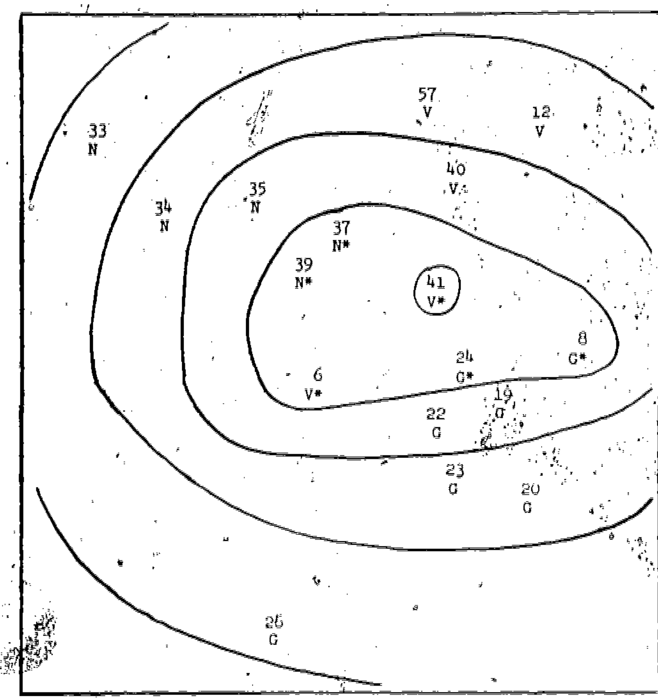
have to know how you get from one to the other; the goodness of fit can be checked directly and visually.

Figures 1 and 2 are the graphic equivalents of the large correlation submatrices borrowed from the Thurstones, namely, Tables 1 and 2 respectively. I hope you will agree that these are the simplest faithful portrayals yet made of such large empirical matrices, along with those of references 4 and 5. Two tests tend to be closer in a figure as their correlations tend to be larger in the corresponding table. The rank correlation between figure distances and table correlations is about .9 in each case, indicating quite satisfactory fits to the two submatrices.

The tests classified in advance as being numerical, verbal or geometrical, are symbolized by "N," "V," or "G" respectively. Analytical tests are indicated by stars, and the achievement tests are unstarred.

FIGURE 2

Data from Thurstone
Primary Mental Abilities (Table 2)



N = numerical G = geometric V = verbal * = analytical unstarred = achievement

In one battery, there were not enough analytical tests. I took tests of achievement which were *a priori* more complex than other tests of their same language and classified them with the analytical tests to build up more points in that region. This assumed complexity should behave similarly to analyticity in the present context, and raises definitional and other problems which require deeper consideration at a later date.

Empirically now, the computer gave the plots of Figures 1 and 2. The relative distances between the points are as they came from the computer. I added the circular contours by considering the definitional system. The blind empirical results of the computer cannot by themselves show a substantive law of formation; they give only dimensionality and distance. Without the faceted definitional system, it would be very difficult to interpret a plot.

One of the most striking things about these plots, again to my surprise, is that all the analytical tests appear in the middle. In each diagram the starred tests group around the center, and the achievement tests go out towards the periphery.

When first discussing the radex for intelligence tests some dozen years ago, I hypothesized it would express a radial expansion of complexity; simplicity would be in the center and expand outwardly into complexity. Complex tests in different areas would tend to be less correlated with each other because they go off in different directions of complexity. The present data show that quite the opposite may be true, although complexity is not the same thing as analyticity (the distinction between being complex or simple is not the same as the distinction between analysis and achievement). In Tables 1 and 2, it is the analytical tests that tend to correlate more with each other, as is shown by their greater mutual proximities in Figures 1 and 2.

A second striking thing about these plots is that the plane seems to divide into three distinct sectors, one for each language of communication. In Figure 1, the verbal tests (the V's) all fall in the upper region, the geometrical tests (the G's) fall lower to the right, and the numerical tests (the N's) fall towards the left half of the plane. Exceptions are three G tests, coded as "G(N)," which fall into the numerical region. These are all tests of *dot-counting*, a combination of numerical and geometrical concepts. They use dots, and hence are geometrical by language of communication; but the exhibited rule — counting — is numerical.

A similar exception is Test 5, coded "N(V)"; it gives arithmetical problems stated in verbal form. Considering our *a priori* culling rules for the present study, we had debated whether or not to leave it out, because of possible confusion on language of communication; some numbers, of course, occur in the midst of all the words. However, Test 5 was retained; it turns out to fall into the verbal region of Figure 1. It perhaps is better classified as V(N), since its language is most predominantly verbal, despite the fact that the exhibited rule is arithmetical. It is interesting that the dot-counting tests fell more into a region corresponding to their *rule* rather than their language, while the verbal-arithmetic test behaved mostly according to its language.

The same general kind of radex appears in Figure 2 as in Figure 1. The verbal tests are off in one region; the geometrical tests tend to be in another region; and the numerical tests tend to be in a third region.

These figures are from two different batteries. These are different sets of tests, given at different times and to different populations. Figure 1 is for eighth grade children in Chicago and Figure 2 is for university students in Chicago. The correlations of Table 1 were calculated directly as product-moment correlations; the correlations in Table 2 are tetrachoric coefficients. Sampling errors are also involved here. But the over-all picture in both cases is the same. Similar results can now be seen by looking again at the nonmetric analyses I had made of data of Guilford (4) and Coombs (5) before the computer technique was available, and before the present radex became "obvious." The next step is to study even larger batteries, allowing more facets to vary systematically, and to see if the multidimensional structure predicted by the corresponding facet design does in fact occur.

What are the basic implications of this kind of structural analysis of interrelations? First of all, for those who have been interested in factor analysis, this may throw a different light on the problem. The purpose of factor analysis *has* basically been to study configurations of points. It just so happens that it is difficult to do this without calculating coordinate systems, but perhaps we have placed too much emphasis on such coordinate systems. The emphasis should be on the configuration itself and not on the coordinates. By emphasizing the coordinate systems, we have been led to try to call them "factors" and to name them. Considering diagrams such as Figures 1 and 2, I think we can realize that perhaps one should not look at concepts like "numerical," "geometrical," "verbal," and so forth, as names of factors, but rather as *elements of rules for item construction*. We can give detailed instructions to item constructors on how to make up types of items we want by using these and elements of other facets implied in the definition of intelligence above. We are exploring and using such facet designs in our test construction program in Israel.

A second important implication is for the problem of prediction. If there are external criteria we wish to predict — such as success in college or success on the job — having a simple map such as in Figures 1 and 2 enables us to deduce what the most parsimonious way may be for accomplishing this. The ideal thing would be to have a map of the set of criteria as well as of the predictors and to see what the joint map is. The combined space will usually be larger than that for each set alone, since the criteria and the predictors will differ usually on some facet which is constant within each set separately. That point in the predictor space closest to a criterion in the combined space can be called the *image* of that criterion. In general, a good way of predicting a criterion is to use the tests which most closely surround its image in the predictor space. For example, if a criterion's image happens to be amidst the four points in the lower right hand corner of Figure 1, then a good choice of a small set of predictors is the corresponding four tests (namely, tests 18, 31, 44, 38). Adding more tests from outside this region generally won't help the prediction much. Perhaps this helps explain why, in the past, people have kept reporting it doesn't pay to use more than three or four variables in a multiple regression. Adding further variables adds largely to sampling error and decreases cross-validity. But should the prediction battery be not much more than two-dimensional, this would help explain why there is no point in using more than three or four variables as predictors. Knowing where a criterion's image is in the predictor space will tell how best to pick an optimal subset of three or four predictors. Each criterion will, of course, have its own image, and hence its own particular optimal sub-set of predictors.

REFERENCES

Papers by Louis Guttman

1. "A New Approach to Factor Analysis: The Radex," in P. F. Lazarsfeld (editor), *Mathematical Thinking in the Social Sciences*, The Free Press, 1954; pp. 258-348.
2. "Empirical Verification of the Radex Structure of Mental Abilities and Personality Traits," *Educational and Psychological Measurement*, 1957, 17:391-407.
3. "What Lies Ahead for Factor Analysis?" *Educational and Psychological Measurement*, 1958, 18:497-515.
4. "A Faceted Definition of Intelligence," in *Scripta Hierosolymitana*, The Hebrew University, Jerusalem, Israel.
5. "Order Analysis of Correlation Matrices," in R. B. Cattell (editor), *Handbook of Multivariate Experimental Psychology*, Rand McNally (in press).
6. "A General Nonmetric Technique for Finding the Smallest Euclidean Space for a Configuration of Points," *Psychometrika*, 1965 (in press).

Monographs by the Thurstones

7. *Factorial Studies of Intelligence*, Psychometric Monographs Number 2, Univ. Chicago Press, 1941.
8. *Primary Mental Abilities*, Psychometric Monographs Number 1, Univ. Chicago Press, 1938.

Psychometric Approaches to Factor Analysis*

Henry F. Kaiser
The University of Wisconsin

The title of this paper is seemingly delimiting: apparently I am only to talk about the "psychometric" side of factor analysis. This restriction gives substantial pause for thought, for what indeed does "psychometric" mean? Is it unduly restrictive?

First, the term can be used in a negative sense, so that I am not to be concerned with recent developments on the statistical (inferential) side of factor analysis; consequently, I shall not talk about sampling distributions, tests of significance, and all the rest of the jargon associated with the problem of making probabilistic inferences from samples of individuals to populations of individuals; we shall consider that populations of individuals are available — or, more realistically — that our sample-population inferences are non-probabilistic. Therefore, let us first consider the psychometric side of factor analysis as a concern for what indeed *is* the mathematical model of factor analysis, and how, algebraically, one thinks about it. My predecessor on this program, Professor Guttman, once said informally, "We have to clean up the algebra of factor analysis before we can, appropriately, hand it over to the statisticians." My successor on this program, Professor Tucker, has many times said essentially the same thing. Since these two gentlemen are undoubtedly the two heaviest of the heavyweights in the theory of factor analysis, restricting consideration to the non-statistical side is, in all likelihood, not entirely a sterile enterprise.

The term "psychometric," though, has an important, and somewhat more pretentious, meaning — with connotations that are dis-

* This paper was prepared while the author was the L. L. Thurstone Distinguished Fellow, Psychometric Laboratory, University of North Carolina.

tinctly of a scientific, inferential nature. Often hidden in the maze of algebra and computations of factor analysis is the fact that these procedures are ultimately rather explicitly concerned with inferences about domains of scientific content — i.e., that we are faced with the important problem, beyond merely statistical description, of attempting to infer the nature of the structure of a large — usually arbitrarily large — battery or universe or domain of variables or tests on the basis of a limited sample (*selection* might be a better word) of variables. In this sense, factor analysis might be called "psychometric inferential," and an explicit concern for this kind of scientific inference is indeed the first problem of factor analysis.

Before I get down to brass tacks, allow me to delimit myself in a way not suggested by my title. I shall restrict myself to considerations of what might be called traditional — or exploratory — factor analysis. I can illustrate this by giving two extreme kinds of situations in which factor analysis is applied. Often, a consultant is confronted with a seemingly sincere "investigator" who presents a large table of numbers and wants to "do" a factor analysis, and for which the apparent principal justification for proposing to do this deed is that the large table just referred to is neatly written out. At the other extreme is the proposed factor-analytic investigation for which every variable to be included has been subjected to the most searching rationalization and, with this case, the hope is that very specific hypotheses can be adjudicated definitively. The former extreme is, of course, to be eschewed, while for the latter, there are perhaps more powerful techniques available (some of which are factor-analytic in nature, thanks to Professors Tucker and Guttman). Thus, I would like to put in a plug for steering a middle course and suggest that factor analysis as it is usually known enjoys its greatest justification as an exploratory technique, by which the variables under consideration all enjoy a reasonably well rationalized, but not necessarily certain, probability of belonging to the scientific domain of interest, and the structure of which is essentially unknown.

From this viewpoint, factor analysis is an unpretentious "bringer of preliminary order out of well-perceived chaos," a technique which is capable of generating ideas rather than providing final answers. Thus, be reasonably — not compulsively — careful in planning a factor-analytic investigation. But don't take your results with such seriousness as to preclude further — probably more penetrating — analysis.

It is convenient to divide the theoretical problems of factor analysis into two groups: (a) those concerned with finding an interesting factor space, the so-called communality problem, and (b) those concerned with finding an interesting basis (coordinate system) for the interesting factor space, the so-called rotation problem. It is almost accurate to say that these two problems, conceptually, are independent of one another. First, then, the communality problem.

D. N. Lawley in 1940 gave what we now recognize as one definitive solution to the problem of factoring. This is his famous maximum likelihood solution. Subsequently, much work has been done with his procedures, particularly by Lawley himself, and by Rao, who, in clarifying and specializing the method, renamed it "canonical factor analysis." It is perhaps best described in Rao's terms as the solution of a certain problem in canonical correlation, and operationally as a principal axes solution of a *rescaled*, reduced correlation matrix. That is, the correlation matrix, with communalities in the diagonal, then has the variances of the variables changed in a way to solve a problem in canonical correlation. The result of this rescaling, or change of metric, has the most important property of leading to solutions which are invariant under *any* rescaling: the solution is scale-free, different from many traditional factoring methods — e.g., that anachronism, the centroid method.

Now the Lawley-Rao method, in particular, rescales in the metric of the unique parts of the observable variables, but to obtain this profoundly important property of securing scale-free solutions, it is sufficient (but not necessary) to rescale in the metric of the unique parts; one can go to the other extreme and rescale in the metric of the common, or communality, parts of the observable variables. John Caffrey and I noticed this a couple of years ago; more accurately, we noticed nothing more than that one could write down an eigenequation paralleling Rao's — and wondered if it meant anything. Apparently it does, for in groping for a rationalization for our writing Rao with backwards rescaling, we found that we were, in solving this affair, determining common factors which successively have maximum reliability in the generalized Kuder-Richardson sense — i.e., factors with maximum "alpha," to use Cronbach's term.

For those of you who are interested in more detail in these two scale-free procedures for finding arbitrary factors, let me refer you to our forthcoming paper, "Alpha Factor Analysis," in *Psychometrika*. I like to think that this paper is quite readable; I would be the first to admit that our very psychometrically oriented procedure undoubtedly enjoys its principal justification today because it was provoked by attempting to imitate the work of Lawley and Rao. I would hope that those who read our paper will be able to provide Caffrey and me with more incisive understanding of what is going on.

Both attacks described above for the communality problem were developed most fundamentally under what might be called the "pure" factor-analytic model of L. L. Thurstone. As many are aware, there are other models, factor-analytic in flavor, giving factor-analytic-like results in practice, but which are not strictly factor analysis. I refer here to the monumental work of Hotelling in his component analysis, and to the monumental work of Guttman in his image analysis. My reference above to the problem of finding an "interesting factor space" rather than a "common factor space" was deliberate: I most certainly do want to talk about both component analysis and image analysis, because practically these methods will yield results which differ only minutely from those obtained under the pure factor-analytic model, while they enjoy many elegant technical simplifications, and in practice have the tremendous advantage of being capable simply of solving the factor score problem (in "pure" factor analysis, the factor score problem is insolvable*).

I do not have the time to describe Guttman's image analysis; I will only assert — and refuse to stand for any arguments — that his basic 1953 *Psychometrika* paper is required reading for the serious student, for all who aspire to be lifted out of the primordial mire which preceded it. Now, Professor Guttman did not explicitly develop his ideas with direct reference to work-a-day factor analysis; this link was provided by our chairman, Professor Harris, in the most important paper to appear in the last ten years on the communality problem. It is in *Psychometrika*, 1962 and is also required reading.

What Harris accomplished was to bring together many of the provocative psychometric properties of image analysis and show their intimate relationship to Rao's version of maximum likelihood factor analysis — again probably better termed "canonical factor analysis." The crucial insight here was brought about again by rescaling: in order to obtain a scale-free solution, Harris rescaled in the metric of the anti-images (the analog of unique parts in factor analysis) and provided us, right down the line, with the image-analytic version of

* More accurately, [illegible] solution does not exist.

canonical factor analysis. Most important practically, in this context, Harris solved the "number of factors" question, but I shall defer talking about this until I summarize the communality problem below.

The three solutions described above, 1. Rao-Lawley canonical factor analysis, 2. Kaiser-Caffrey alpha factor analysis, and 3. Guttman-Harris image analysis, let me reiterate, all have the elegant property of yielding factors which are invariant under changes of scale — one no longer need be concerned about units of measurement, as the same results are obtained regardless. Initial factoring procedures which do not have this property deserve only one treatment: the ash can. What then about the popular use of Hotelling's principal component analysis, which does *not* have this property? The rule of thumb of taking only such components of the *correlation* matrix with latent roots or eigenvalues greater than one — proposed years ago by me originally for no really good theoretical reason — has disdainfully been referred to as "Little Jiffy" by some of my thoughtful colleagues. It turns out though, in a component analysis setting (no unique factors postulated), that if we maximize Cronbach's coefficient alpha rather than Hotelling's variance accounted for, Little Jiffy can be rescued from the ash can of non-scale-free solutions; more exactly, if we maximize alpha, we automatically rescale into the correlation matrix associated with any covariance matrix and thus, are able to determine scale-free components. So, more or less after the fact, we can add a fourth scale-free procedure — the popular Little Jiffy — for determining an interesting factor space.

These four methods can be summarized in the following two-by-two table:

Canonical Factor Analysis	Image Analysis (á la Harris)
Alpha Factor Analysis	Principal Components (Little Jiffy)

All four procedures involve principal axes solutions, with rescaling; all four are scale-free. The two on the left are "pure" factor analysis and may be obtained operationally by iterating on the solution immediately to its right. The two on the right in this sense are first approximations to those on the left, and while they are not factor analysis in the strict sense, they do yield very similar results to their counterparts. They have the substantial advantage of giving factors which are linear combinations of the observable variables and thus, of yielding scores on the derived variables or "factors" with no fuss — in stark contrast to the absence of a unique solution for factor scores in the purely factor-analytic solutions.

Probably the most important side of the communality problem in practice is the hoary old question of the "number of factors." In the foregoing two-by-two table, the number of factors which come out "naturally" in the first row, for canonical factor analysis and for image analysis à la Harris, is Guttman's classic strongest lower bound, empirically "too many" — about half the number of variables observed. For the second row — alpha factor analysis and principal component analysis of the correlation matrix (Little Jiffy) — the number of factors which come out "naturally" is Guttman's classic weaker lower bound, the number of eigenvalues greater than one of the correlation matrix, usually about the "right" number.

I might say that Guttman's weaker lower bound, in giving *about* the "right" number, can sometimes give too few, and underfactoring is a catastrophe. His stronger lower bound, which invariably gives too many, can never lead to this disaster of underfactoring; overfactoring is usually only a mild irritant, which may be cured by careful thought. I wish that I were able more successfully to promote this idea, but generally practitioners want exactly the "right" answer for the number of factors to be given without careful thought, and that, gentle readers, is as unattainable as Eldorado.

I'm not going to reveal which of the four solutions discussed above is the right one; the question really is without meaning. I will say that four is a somewhat smaller number than the myriad of techniques one is confused by in the earlier literature.

The result is an interesting factor space, preferably found by one of the four methods just reviewed. Its basis, however, is not interesting — being some variety of (rescaled) principal axes — so this original basis should be changed by a linear transformation — i.e., it should be "rotated." What is the current state of this general problem? For objectivity's sake, I shall restrict myself to analytic methods; consistent with considering factor analysis in its traditional exploratory vein, I shall also consider that I am going in "blind," searching for a structure, rather than testing for the existence of a postulated structure.

In this venture, I first distinguish the so-called "orthogonal" from the so-called "oblique" case, according to the kind of restrictions I place on the transformation matrix. This distinction usually — but not always — bears a one-to-one relation to whether I restrict the transformed factors to being uncorrelated or not.

In the orthogonal case, present criteria are almost invariably of the form:

$$Q - wK = \text{maximum}$$

where Q is the quartimax function, and K is the function, which in conjunction with a weight w of one turns the criterion into varimax. More generally this weight w controls the variability in the variance contributions (column sums of squares) of the factors; at one extreme, for $w = -\infty$, we have principal factors; for $w = 0$, quartimax; for $w = 1$, varimax; for $w =$ one-half the number of factors, Saunders' "equamax;" and for $w = \infty$, all factors have exactly equal column sums of squares. Recently, I have looked at this class of criteria with some care, particularly at values of w greater than one, in the hope of obtaining orthogonally rotated factors with more nearly equal variance contributions — a hope generated by the concern of some investigators who have expressed little doubts here and there as to the ubiquity of varimax. Unfortunately, I have been able to prove that it is always possible, for any trans-varimax solution where w is chosen greater than one, even by a smidgen, to produce examples in which utter catastrophe occurs and two pure independent uncorrelated clusters will be subjected to a 45° rotation. I have also looked at criteria in this class where w is generated by the data; here I have been able to prove that the factorial invariance (in simple cases) which obtains for varimax cannot, in general, occur. I must conclude that within this class of solutions, varimax best does the deed. But here "best" has mostly a negative connotation: the old reliable work-horse varimax will never result in catastrophe, while other criteria in this class, perhaps better in specific problems, can be disastrous in others.

It appears, then, that further progress in the orthogonal case must come from an entirely fresh start. The most promising development that I know of occurs tangentially in Peter Schönemann's recent doctoral dissertation at the University of Illinois. He may have something very good for us in the near future.*

Analytically, in the "pure" sense of writing down some to-be-optimized function of the loadings to be obtained via a non-orthogonal transformation, the oblique case has been quiet for about five years. The last serious effort in this line of thought initiated by John Carroll and followed up by Dave Saunders, Kern Dickman and me was something dubbed "binormamin" by Professor Dickman and presented by the two of us at the '59 APA meetings. We discarded it, more or less, because of the simply awful computational problems involved and because of its apparent sensitivity to the number of factors retained (it would seem to collapse the factor space if you gave it a dirty look). There has been a minor revival of interest now that we have the super-speedy computers to do the job; it appears that some of our initial concern was overly pessimistic. The one positive statement I can make about it is that when it works, it works fine — better, say, than the earlier oblimax and oblimin criteria. In general, though, *any* analytic oblique criterion is not to be trusted routinely to give adequate results because of the apparent lack of our ability to apply the constraint of non-singularity on the transformation matrix in the traditional approach to oblique "rotation."

The terrible terror of singular transformations in the oblique case was resolved in the summer of '63 when Professor Harris suggested to me that we apply orthogonal transformations to deliberately rescaled and then "mis-scaled" principal axes solutions. Patting him gently on the head, I pointed out that this was of course illegal. Smiling serenely, he decimated my conservatism by pointing out that after such transformations, all the preliminary scaling can be taken out very easily with non-singular diagonal matrices. The result? All possible transformations, and thus all possible factor analytic solutions — involving uncorrelated or correlated factors — for a given factor space can be obtained with orthogonal transformations. This is a Big Breakthrough, for being able to reduce the entire rotation problem simply to one of applying orthogonal transformations makes the problem really tractable for the first time. Already Harris has brought this entirely new framework to bear in solving definitively the problem of cluster analysis. Further applications will undoubtedly be forthcoming soon. Be prepared: read "Oblique factor analytic solutions by orthogonal transformations," soon to appear in *Psychometrika.*

* Doctor Schönemann is now at the University of North Carolina.

I shall stop now. Let me close on an optimistic note. I don't think I'm becoming senile when I say that, psychometrically, the period of the algebraic explication of traditional exploratory factor analysis is drawing to a close. We are near to having things well enough in hand to give the model to the statisticians. Perhaps pretentiously, I think I can say that we psychometricians are like Galton, who, when having the correlation coefficient sufficiently under control conceptually, turned the thing over to Karl Pearson.

Experiments in Multi-Mode Factor Analysis*

LEDYARD R TUCKER
University of Illinois

Multi-mode factor analysis is being developed to help meet the needs for a procedure to search for and to represent relations existing in data from observations which may be classified according to several identifying attributes. Initial discussions of this development appear in the monographs: *Problems in Measuring Change* edited by Chester Harris (8), and *Contributions to Mathematical Psychology* edited by Norman Frederiksen and Harold Gulliksen (9).

An example of the simplest case for an identifying attribute of observations would be when a single test is administered to a group of individuals. The scores on this test constitute our observations and each observation is identifiable by the individual. Thus, the individuals constitute an identifying classification of these observations, or a mode of identifying classification. An alternative example is when a single individual is given a large number of tests. Then each score can be identified by the test. The group of tests now constitutes a mode of identifying classification. A step up in complexity of classification occurs when, say, a group of people is given a group of tests. Now each score is identified by the person and by the test, thus involving two modes of identifying classification. These data may be recorded in a table with a row for each individual and a column for each test. Each score is recorded in one cell of such a table. Traditional factor analysis was developed to systematize and represent relations in such data.

A next step upward in complexity occurs when, for example, a

* The research reported in this paper was jointly supported by the University of Illinois and the Office of Naval Research under contract Nonr 1834 (39).

group of individuals is given a battery of tests on a number of different occasions. This means that each score must be identified by the individual, the test, and the occasion, thus constituting three modes of identifying classifications. These data may be thought of as being recorded in a three-dimensional table such as the data box, illustrated in Figure 1 where there is a row in mode 1 for each occasion, a column in mode 2 for each test, and a layer in mode 3 for each individual. Each cubicle of this box contains one test score. Multi-mode factor analysis is being devised to treat data such as those in this data box.

An extensive variety of situations can be conceived which involve three or more modes of identifying classifications. First is the example just discussed. A second example occurs when a number of individuals are rated on a number of traits by a number of other individuals; an analysis of such data will be presented. An alternative to the preceding example is when each of a number of individuals is rated on a number of personality traits in different kinds of situations. An extended example would involve four modes: persons, traits, occasions, and situations. In this example, occasions might be interpreted as particular periods of life such as entering first grade, in the middle of first grade, toward the end of first grade, etc. The situations might be: in the classroom, on the play yard, at home, etc. Such a study might have extended relevance to developmental psychology. A third example could be when individuals are learning a complex task on which a number of different scores can be taken on a number of trials. Such data constitute our second example to be presented. The third example of analysis to be presented involves individuals imagining that they are in a number of different situations and indicating the extent of response for different modes of responding while in these situations.

Figure 2 presents a schematic representation of the three-mode factor model. There is a factor matrix for each mode of identifying classification, mode 1 on the left, mode 2 at the right, and mode 3 as if in depth. Each of these factor matrices relates the observed variables of the mode of identifying classification or factor variables for the mode. Thus, if method of rating constitutes mode 1, the rows of the mode 1 factor matrix would be the various methods of rating, and the columns would be conceptual methods of rating. The observed methods would be described in terms of their relation to these conceptual methods of rating. For the three examples to be described, mode 3 will be considered to be the individuals on whom the observations were made so that columns of the mode 3 factor matrix in Figure 2 represent the actual individuals. The rows of this matrix represent conceptualized or idealized individuals.

One of the developments beyond the procedures described in the monograph *Contributions to Mathematical Psychology* (9) is a procedure by which allowance can be made for errors of measurement and other influences that affect the measures for each particular combination of variables in mode 1 and mode 2. This raises a problem analagous to the communality problem in traditional factor analysis and results in an indeterminancy of the entries in the mode 3 factor matrix. This is analagous to the factor score problem. The three factor matrices are tied together by a small core box which gives the relations among the three types of idealized entities. A point of interest is that the new procedure utilizes a matrix of mean cross products between variables analagous to the Campbell and Fiske multitrait-multimethod matrix (1). The examples will present the mode 1 and mode 2 factor matrices and the core box for each case.

Consider Table 1. This presents the results of an analysis of multitrait-multimethod data. These data were collected by Professors Lowell Kelly and Donald Fiske (5) as part of a research project on the selection of clinical psychologists conducted at the University of Michigan. Data were supplied by Professor Donald Fiske. The analysis was conducted by Mr. Edward Hoffman (4) while he was a graduate student at the University of Illinois. The methods of rating constitute mode 1; these methods are ratings by the staff, by the teammates, and by the trainee himself. The mode 1 factor matrix is given at the upper left. Two factors were extracted, a general factor and a factor related to the self-ratings. These factors are unrotated principal axes. Note that the staff- and teammate-observed methods of rating are very similar whereas the self-rating depends in part on the general and in part on the specific idealized rating method for the self.

Five factors were found for mode 2 having to do with the different traits that were rated. Fifteen traits were selected from the 22 that actually were rated. These 15 were to represent the five recurring factors identified by Fiske (3). The first factor strongly apparent was the social adaptability factor, the second was the factor seriousness or conscientiousness, the third trait factor was an inquiring intellect, whereas the fourth appeared to be emotional control. The fifth factor was less well defined, but appears to be related to self-assertiveness.

FIGURE 1

The Data Box

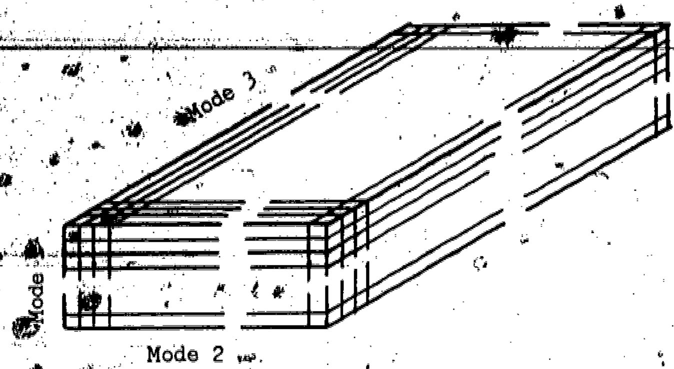
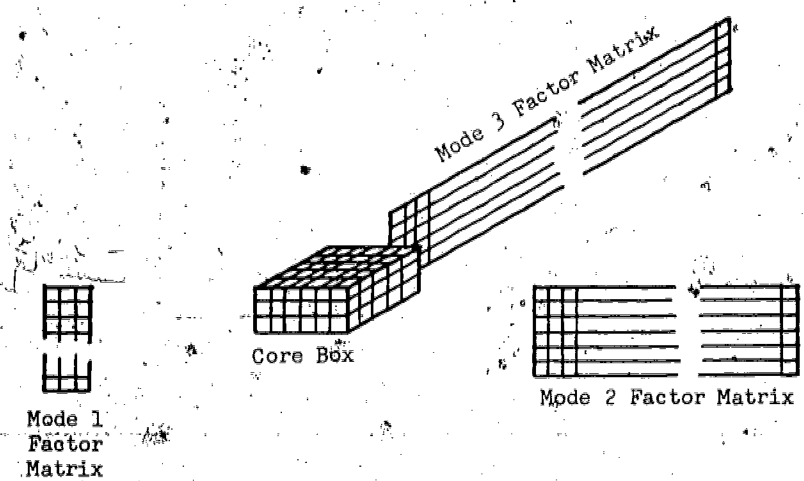


FIGURE 2

Schematic Representation of the Three-Mode Factor Model



Entries for the core box are given in the lower middle section of the table. The first person factor is characterized by high values for social adaptability and self-assertiveness as rated on the general rating method factor. A person who was describable as high on the first person factor and near average on the remaining person factors would be rated highly social adaptable and highly self-assertive. All other ratings would be near average. Another person characterized by a low value in the first person factor would tend to be low in social adaptability and low in self-assertiveness when rated by the general rating factor. All self-ratings would be near average. The second person factor is characterized by a high positive loading on seriousness and conscientiousness and a high negative loading on self-assertiveness when rated by the general rating factor. The third person factor also involves the general rating factor and positive values for an inquiring intellect and self-assertiveness. The fourth person factor brings in the self-rating method factor to a greater extent where the individuals high on this factor rated themselves high on inquiring intellect and emotional control. The fifth and sixth person factors involve the emotional control as rated on the general rating factor with different combinations of seriousness-conscientiousness and self-assertiveness. It is of considerable interest that the self-rating factor is involved to a marked extent only in one of the person factors, factor four. This person factor may represent the difference between the way a person views himself and the way others view him. We may summarize the results, then, as indicating a fairly strong communality of ratings by the different rating methods in the general factor with the self-ratings diverging in part only toward the self-rating factor. The trait factors are relatively clear and identifiable. The person factors indicate an area within which the self-ratings diverge from the ratings by the staff and teammates.

Results for our second example are given in Table 2. The data for this analysis were taken from the Psychological Monograph by James Parker and Edwin Fleishman on the *Ability Factors and Component Performance Measures as Predictors of Complex Tracking Behavior* (7). The tracking device employed was constructed so as to simulate roughly the display characteristics and control requirements of an airborne radar intercept mission. Time of the tracking performance of each subject was divided into a number of trials. On each trial four measures were taken: a measure of horizontal error, a measure of vertical error, a sideslip error, and a time-on-target measure. Data for

Table 1

Analysis of Multitrait-Multimethod Data

Mode 1: Methods Factor Matrix

Rating Method	General	Self
Staff	.76	−.18
Teammate	.65	−.13
Self	.46	.48

Mode 2: Trait Factor Matrix (Rotated)

Variable		Social Adaptable	Serious Conscientious	Inquiring Intellect	Emotional Control	Self-Assertive
4	Cheerful—Depressed	.75	−.10	−.06	.34	−.07
6	Attentive to people—Cool, aloof	.68	.15	.00	.09	−.08
20	Frank, expressive—Secretive, reserved	.61	.10	.02	−.07	.05
10	Good natured, easy going—Self-centered, selfish	.55	.24	−.03	.19	−.38
11	Talkative—Silent, introspective	.50	−.13	−.06	.05	.32
12	Adventurous—Cautious	.41	−.07	.10	.24	.13
5	Serious—Frivolous	−.17	.68	.01	.03	.10
17	Conscientious—Not conscientious	.14	.60	.00	.01	−.10
8	Broad interests—Narrow interests	.02	.13	.64	.05	.10
18	Imaginative—Unimaginative	.24	−.04	.60	−.07	−.03
21	Independent minded—Dependent minded	−.14	.22	.38	.25	.32
7	Unshakable poise—Easily upset	.05	.08	.06	.75	.11
16	Placid—Worrying, anxious	.29	−.09	.00	.70	−.16
15	Self-sufficient—Dependent	−.08	.26	.16	.53	.18
3	Assertive—Submissive	.00	.04	.04	.04	.61

Core Box

Person Factor		Social Adaptable	Serious Conscientious	Inquiring Intellect	Emotional Control	Self-Assertive
1	General	1.56	−.34	.25	−.11	1.31
	Self	.01	.06	−.03	−.14	.12
2	General	.12	1.25	−.05	−.05	−1.01
	Self	−.09	−.26	−.30	.05	−.07
3	General	.21	.17	1.45	.05	1.01
	Self	−.10	−.11	−.12	−.19	−.20
4	General	.16	.18	.16	.50	.30
	Self	.33	.37	.48	.60	.29
5	General	−.05	.41	.27	.75	−.43
	Self	−.10	.04	−.12	−.26	.29
6	General	.17	−.48	−.04	.91	.04
	Self	−.20	−.37	−.10	.05	−.17

ten of the trials were presented in the form of correlations in a matrix analagous to the Campbell and Fiske multitrait-multimethod correlation matrix. A consequence of using correlations in this situation is that the analysis involves consistencies in deviations from the mean learning curve for each measure.

Mode 1 was concerned with the scoring measures. Again there were two factors. This time one of the factors was related to direction errors involving the horizontal errors, the vertical errors, and the time-on-target. The second measure factor was related to sideslip control involving the sideslip error and time-on-target. These two factors indicate a marked distinction between the two aspects of control in this tracking test.

The mode 2 factor matrix is for trials and appears at the upper right of Table 2. Four factors were apparent, one for very early learning, second for middle early, a third for middle late, and a fourth for very late trials. These trial factors constitute standard patterns for discrepancies from the mean learning curves.

The core box is at the bottom portion of the table. There are seven person factors, representing seven dimensions of individual differences in performance on this complex tracking task. The first four person factors involve the direction error factor. The first person is for very early performance on direction errors, the second is somewhere between middle early and middle late on direction errors, the third emphasizes the middle late direction errors, and the fourth factor tends to emphasize the very late direction errors factor. The fifth, sixth, and seventh factors are concerned mainly with the sideslip control: the fifth factor involving very early performance on sideslip control, the sixth factor being related mostly to middle early and middle late performance on sideslip control, and the seventh factor being related to the very late performance in sideslip control.

A major point of interest is the almost complete disassociation in the person factors of effects of the direction errors factor and the sideslip control factor. Only during the middle late trials factor does there seem to be an interaction between the direction errors and the sideslip in discrepancies from the mean learning curves for the measures. This independence should be important to our theoretical considerations of such learning tasks and to our attempts to measure performance on them. If similar independence of measure factors were observed for other learning tasks, serious doubt would exist as to the

Table 2
Analysis of Complex Tracking Task Measures

Mode 1: Scoring Measures Factor Matrix

Measure	Direction Errors	Sideslip Control
Horizontal Error	2.51	−.10
Vertical Error	2.50	.01
Sideslip Error	−.13	2.39
Time-on-Target	2.58	1.08

Mode 2: Trials Factor Matrix

Trial	Very Early	Middle Early	Middle Late	Very Late
1	1.09	−.03	−.01	−.02
2	1.39	.02	.01	.00
3	.30	1.16	.00	.07
4	−.03	1.26	.51	−.05
5	.06	.83	.85	−.06
6	−.02	.39	1.13	.06
7	.00	−.02	.83	.65
8	−.03	.04	.36	1.10
9	−.03	.01	.03	1.49
10	.06	−.03	−.03	1.38

Core Box

Person Factors	Measure Factors	Very Early	Middle Early	Middle Late	Very Late
1	Direction	1.60	.10	.02	−.07
	Sideslip	.37	.07	.07	−.10
2	Direction	.14	1.22	1.21	−.02
	Sideslip	.10	.11	.35	.02
3	Direction	.04	.72	3.05	.34
	Sideslip	.03	.17	.58	.10
4	Direction	.00	.01	.95	1.13
	Sideslip	.03	.00	.11	.15
5	Direction	.19	.01	.04	.01
	Sideslip	1.18	.19	−.03	−.01
6	Direction	.06	.11	.51	.06
	Sideslip	.05	.67	1.97	.27
7	Direction	−.10	.03	.12	.29
	Sideslip	−.03	.00	.77	.85

equivalence of a number of learning experiments when different measures of performance were employed in these experiments.

The results in Table 3 are taken from an analysis by Dr. Joseph Levin (6) of situation versus mode-of-response rating data collected by Norman Endler, J. McV. Hunt, and A. J. Rosenstein (2). Mode 1 involves the mode of response. There are 14 different listed modes of response given in the upper portion of the table. Eleven situations are shown in the middle section of the table. The task of each subject was to imagine that he was in each of these listed situations and to rate the extent that he would respond for each of the modes of response. The rating scales were from 1 to 7. The ratings were standardized over individuals for each mode of response within each situation.

The mode of response factor matrix for mode 1 is given in the upper portion of the table. The four modes of response listed first have high loadings on a factor called Distress or General Distress. The next four modes of response have high loadings on a factor that could be called Exhilaration. And the last six modes of response are grouped under a factor that might be called Autonomic responses.

The factor matrix for situation is given for mode 2 in the center of the table. Five of the situations tend to group under a factor that could be called Interpersonal situations. Three more of the situations have high loadings on the second situation factor which might be called Inanimate situations. The last three situations load highly on the third factor, which puzzled us for some time but seemed to have a common characteristic in that the subject is, as it were, going into a situation whose exact nature is unknown to him.

The results in the core box at the bottom of the table are especially interesting. There appear to be three person factors which could be characterized as idealized persons. The first person seems to react with both General Distress and Autonomic reactions to all three types of idealized situations. A negative person in this person factor would not be distressed nor have Autonomic reactions to these idealized situation factors. Persons characterized by either high positive loadings on the first person factor or high negative loadings on this first person factor would be about average as to Exhilaration responses to the conceptual situation factors. The second idealized person would be characterized by Exhilaration to all three types of situations, especially the Inanimate, and a negative amount of Distress to this Inanimate situation, while he would show Distress in Autonomic reaction to the Unknown to Subject factor. The person with a high negative loading on the second person factor would have the reverse effects. For the third person factor there is an interesting differentiation drawn between the Interpersonal and the Inanimate situation factors. A person high on this third person factor would be exhilarated by the Interpersonal and have a negative Exhilaration for the Inanimate. This person would have negative distress for the Interpersonal situations and about average Autonomic reaction. In contrast, for the Inanimate situations, he would be positively distressed and have positive Autonomic reactions. The Unknown to Subject situations are more like the Interpersonal situations to this individual, but the reactions are not quite as extreme.

These results appear to me to reflect rather commonplace obser-

Table 3

Analysis of Situation-Mode of Response Rating Data

Mode 1: Mode of Response Factor Matrix

Mode of Response	Distress	Exhilaration	Autonomic
2 Get "uneasy feeling"	.68	−.02	−.05
4 Heart beats faster	.63	.10	−.04
3 Emotions disrupt actions	.46	−.04	.10
6 Perspire	.34	−.03	.18
8 Enjoy the challenge	−.04	.77	.00
12 Seek experiences like this	−.05	.73	.05
4 Feel exhilarated and thrilled	.11	.60	−.02
5 Want to avoid situation	.34	−.38	.03
13 Have loose bowels	−.15	.01	.55
7 Need to urinate frequently	.07	.09	.49
11 Get full feeling in stomach	.05	.05	.42
14 Experience nausea	.02	−.09	.41
9 Mouth gets dry	.21	−.03	.31
10 Become immobilized	.23	−.05	.25

Mode 2: Situation Factor Matrix

Situation	Interpersonal	Inanimate	Unknown to Subject
5 Speech before large group	.55	−.10	.03
10 Interview for important job	.52	−.03	.04
8 Competitive contest	.46	.09	−.14
11 Final exam in important course	.33	.09	.05
2 New date	.31	−.12	.30
4 Ledge high on mountain side	−.06	.56	.01
9 Alone in woods at night	.07	.47	−.06
7 Sailboat on rough sea	−.01	.42	.08
3 Psychological experiment	.00	.03	.54
1 Auto trip	.01	.04	.38
6 Counseling Bureau for personal problems	.20	.04	.36

Core Box

Person Factors	Response Factors	Interpersonal	Inanimate	Unknown to Subject
		Situation Factors		
1	Distress	.40	.39	.22
	Exhilaration	−.05	−.11	.03
	Autonomic	.41	.37	.23
2	Distress	−.07	−.29	.20
	Exhilaration	.29	.44	.32
	Autonomic	.19	.05	.33
3	Distress	−.22	.30	−.16
	Exhilaration	.37	−.19	.19
	Autonomic	−.01	.37	−.15

vations as to different types of people: one, a very shy person; two, a person who is generally exhilarated; and three, a kind of person contrasting Exhilaration between Interpersonal and Inanimate types of situations. I submit that the existence of these types of people raises real problems in the development of psychological theory as related to the situation and the responses of individuals to these situations. The psychological theory must be sufficiently complex to take into account the various kinds of individuals involved. It does not appear that a simplified general theory will be adequate. A postscript that must be made for this study is that the data are the ratings by the subjects when they imagine themselves in the particular situations. A much more extensive study in which the individuals were actually placed in these situations and measures of their responses made would be desirable.

In the three studies reported, the factor matrices for modes 1 and 2 tend to confirm previous results and observations. The values in the core box, however, tend to bring out newer statements of relation. One might take the view that the mode 1 and 2 factor matrices are dealing with relations of a more surface type and that the core box is dealing with deeper and more subtle relations. These relations in the core box should have more general effects on the phenomena being observed. One might argue, then, that the relations indicated in the core boxes should be of greater general interest in understanding the phenomena. It is my belief that the core boxes for the three analyses presented do indicate general effects of considerable interest in the content area of the observations made.

REFERENCES

1. Campbell, D. T., and Fiske, D. W. "Convergent and Discriminant Validation by the Multitrait-Multimethod Matrix." *Psychological Bulletin*, 1959, *56*, 81-105.
2. Endler, N. S., Hunt, J. McV. and Rosenstein, A. J. "An S-R Inventory of Anxiousness." *Psychological Monographs*, No. 536, 1962, *76*, No. 17.
3. Fiske, D. W. "Consistency of the Factorial Structures of Personality Ratings from Different Sources." *Journal of Abnormal Psychology*, 1949, *44*, 329-344.
4. Hoffman, Edward L. and Tucker, Ledyard R *Three-Way Factor Analysis of a Multitrait-Multimethod Matrix*. Technical Report, Department of Psychology, University of Illinois, 1964. (Mimeographed)

5. Kelly, E. L. and Fiske, D. W. *The Prediction of Performances in Clinical Psychology.* Ann Arbor: University of Michigan Press, 1951.
6. Levin, Joseph. *Three-Mode Factor Analysis.* Technical Report, Department of Psychology, University of Illinois, 1963. (Mimeographed)
7. Parker, James F., Jr., and Fleishman, Edwin A. "Ability Factors and Component Performance Measures as Predictors of Complex Tracking Behavior." *Psychological Monographs,* No. 503, 1960, *74,* No. 16.
8. Tucker, Ledyard R "Implications of Factor Analysis of Three-Way Matrices for Measurement of Change." In C. W. Harris (Ed.), *Problems in Measuring Change.* Madison: University of Wisconsin Press, 1963.
9. Tucker, Ledyard R "The Extension of Factor Analysis to Three-dimensional Matrices." In N. Frederiksen and H. Gulliksen (eds.), *Contributions to Mathematical Psychology.* New York: Holt, Rinehart & Winston, 1964.

Technology and Behavior*

OMAR KHAYYAM MOORE
Rutgers—The State University

I

Occasionally, while having lunch in the faculty club, some friend in the physical sciences wants to talk about the prospects of the behavioral sciences. The questions which are raised often have sharp edges. For instance, a physicist once asked me whether I thought it was likely that the behavioral sciences will have caught up with seventeenth century physics by the end of the twentieth century. He went on to sympathize with our difficulties, adding that it must be trying to sit around waiting and waiting for our "Galileos" and "Newtons," to say nothing about our "Einsteins."

I attempt to parry such thrusts (jocular, we hope) in a variety of ways. One line of defense which works reasonably well goes as follows: Success in behavioral science research presupposes a high level of competence in the physical sciences — our work requires the engineering applications of basic discoveries in the physical sciences. These discoveries and resulting applications have been very slow in coming. More specifically, the behavioral sciences, for their proper development, must have subtle, sophisticated scientific instruments. Entities as complex as human beings cannot very well be studied with the help of the unaided senses and intellect *only*.

As a case in point, I sometimes mention digital computers: research instruments which are now playing an important part in the behavioral sciences. Such instruments are vital to the processing of complex data, and are indispensable for the rapid simulation of human

* I am indebted to Alan Ross Anderson and Richard Kobler for helpful criticisms of earlier drafts of this paper.

functions such as higher-order problem solving. I claim that there has been an unconscionable delay in the development of digital computers — after all, the basic idea is an old one. It was fully worked out between 1828 and 1839 by Charles Babbage, a professor of mathematics at the University of Cambridge; he called his machine an "analytic engine." However, as a mechanical contrivance it was a failure. The clumsy wheels and cards which he sought to use were too slow and unreliable. Had the physical sciences done their work so that there was an electronics industry in the early 1800's, we would have had more than 100 years of simulation studies. And think of what the analysis of all those data might have meant to the development of empirical research in the behavioral sciences! But, alas, we have had access to high-speed computers only since World War II, and most of the time the physical scientists themselves were using them. As can be readily imagined, it is not difficult to furnish examples other than that of the digital computer so as to fill out any conversation and exhaust the patience of the interrogator: there are many other technological innovations, for instance, carbon-14 dating, now being used in the social and behavioral sciences, which were only recently derived from fundamental research in the physical sciences.

Nevertheless, even if one succeeds in putting off one's colleagues, the nagging question remains, "When *will* our 'Galileos' and 'Newtons' appear?" What is more disquieting, if serious thought is given to such questions, Galileo's experiments, which were important for Newton, did not require complex instruments. As Alfred North Whitehead noted, "So far as experimental skill, and delicacy of apparatus were concerned, this experiment [Galileo's experiment showing that bodies of different weights, if released simultaneously, would reach the earth together] could have been made at any time within the preceding 5,000 years" (1). Galileo's and Newton's triumphs depended on clear reasoning and mathematical imagination, not on scientific equipment. We should also consider the distressing possibility that these giants might have been thrown off stride had they had more precise instruments — with greater precision of measurement, they might have discovered that their theoretical model and empirical data did not quite match. They were saved this embarrassment by the crudity of their scientific instruments.

It is certainly true that as the physical sciences developed from their early striking successes, they came to rely more and more heavily upon a complex technology in the conduct of their research, a technology which they helped create.[1] Experimentally developed machines have often been a stimulus to important theoretical work. As J. R. Pierce remarks, ". . . the existence of the steam engine, in which phenomena involving heat, pressure, vaporization, and condensation occur in a simple and orderly fashion, gave tremendous impetus to the very powerful and general science of thermodynamics" (2). As another example, Pierce traces the development of information theory as a response to the problems of efficiently encoding messages in the arts of telegraphy and telephony.

In other words, sometimes theoretical formulations in science have arisen out of our attempt to understand and to control man-made devices—devices which originally were created with little understanding of their principles of operation. This latter point bears further emphasis: Benjamin Franklin's kite experiment contributed to our understanding of the connections between lightning and electricity, but *did not presuppose a theory which applied to both*; only a hunch.

It seems reasonable to conclude from the history of science that there is a partial interdependency between theoretical formulations, suitable scientific instruments, and technology. Sometimes pure, abstract reasoning goes a long way even in the absence of elaborate equipment; sometimes mechanical inventions themselves are the subject matter of theoretical investigations; and sometimes clever inventors create useful things which they do not understand fully. Nonetheless, three things stand out as far as the behavioral sciences are concerned: 1) they have been marked by the absence of mathematically rigorous theoretical breakthroughs, 2) they have not produced much in the way of practical inventions, and 3) it is only in recent times that they have benefited substantially from developments in mathematics, and from engineering applications deriving either from other scientific disciplines or from the contributions of free-lance inventors.

II

At this point I should like to turn to a discussion of model construction in the behavioral sciences, and some of the relations between

[1] I say "helped create" rather than "created": there have been thousands of maverick inventors who stood outside the scientific establishment while making important contributions to an evolving technology—witness Thomas Alva Edison.

model construction and technology. I will suggest that the kinds of questions my colleagues in the physical sciences are asking, and which I have asked myself in the past, are misleading. These questions take it for granted that model construction in the behavioral sciences will take the same general path that has been taken in the physical sciences and that these models (though there will be substantive differences) will first be formulated in the *usual* axiomatic manner: theorems will be deduced from them in the *usual* way.

I will argue, on the other hand, that we have been looking in the wrong place for theoretical leads, and that we have not recognized such theoretical structures as exist even though many of them have been in plain view for most of human history. To put the matter another way, we are very much like the fish in the story about fish becoming scientists. Someone once conjectured that if fish were to ever develop a scientific culture, the last thing that the scientists among them would discover would be water. In a similar way, it is my position that our theoretical ideas about human behavior are so largely ingredients of the medium in which we live, ourselves included — we are, so to speak, made out of models of ourselves — that we have great difficulty in distinguishing between ourselves and our theoretical formulations. We not only swim in a medium of abstraction, but unlike fish, we are in part made out of the same stuff.[2]

Let me turn to the task of specifying the major outlines of the point of view about model construction which I have in mind. In a number of papers (7)(8)(9), Alan Ross Anderson and I have urged that we cease looking at the ordinary human being as an *a*theoretical or a *non*theoretical or even an *anti*theoretical creature. Behavioral scientists generally take the stance that the ordinary man, a citizen in good standing in whatever community he may live, has very little in the way of intellectual resources to guide him in managing his affairs. He is credited with having some folk sayings and aphorisms. Behavioral scientists are quick to point out that even these blunt conceptual

[2] Admittedly, the thesis which I am advancing is counterintuitive. It also may be dead wrong. I invite you to consider it because it is a perspective from which my colleagues and I have been able to make a number of inventions which are useful in the behavioral sciences, for example, the talking typewriter (3), and it has been productive of some formal work in mathematical logic (4) (5) (6). No claim is made that this perspective yields these products as strict deductions from an axiomatic scheme nor is the claim made that it is the only perspective which would have yielded these same results.

tools are generally contradictory. For instance, "Absence makes the heart grow fonder" and "Out of sight, out of mind." In addition to these saws, "the average man" has some practical knowledge, some skill at rule-of-thumb reasoning, and if he is really hard pressed about something, he can fall back on various tradition-based explanations.

We would like to suggest a different view of "Man." We believe that early in human history, probably at about the same time man developed natural languages, he also created models of the most important features of his environment. These were abstract models, which collectively covered relations holding between 1) man and nature — insofar as nature is not random — 2) man and the random, or chancy elements in experience, 3) man in his interactional relations with others like himself, and finally, 4) man and the normative aspects of group living. Structures falling within these four classes of models were created by unsung Newtons, so that there does not exist a society, however primitive, that does not have cultural objects falling in these four categories of models. [For an explication of the concept "cultural object" see (10).]

Every society, as far back as we have any evidence, has *puzzles* which stand in an abstract way for man-nature relations. Every society has some *games of chance*. (According to our view of the matter, games of chance are abstract models of the aleatory aspects of existence.) Every society has *games of strategy* in the sense of von Neumann (11). These games capture some of the peculiar features of interactional relations among men, relations in which no party to an encounter controls all of the relevant variables upon which the final outcome depends, though each controls some of these variables and each participant can take account of the potential actions of others involved in the same situation. Every society has *aesthetic* entities: art forms, which we claim give people the opportunity to learn to make normative judgments about their experiences. All societies make use of these cultural objects in the socialization of the young and for the re-creation or recreational enjoyment of those who are older. Simple forms of these models are internalized in childhood, and more complex versions of them sustain us in adulthood.[3]

[3] It should be pointed out that until mathematicians created formal analyses of the structure of some of these models, their depth and subtlety were not appreciated fully. Of the four classes of models, two have received adequate formal analyses, namely, theories of probability have all games of chance as models, and the various theories of games of strategy have all games of

III

From the point of view presented here, the ordinary man in any society should be thought of as having in his mind (if you are of a school of thought which grants him one) at least four classes of models which he can use in a highly abstract way to guide his behavior.[4] From this standpoint, it probably would not have been possible for mankind to develop complex civilizations without these abstract folk models. We make the assumption that having these models was a necessary condition for the development of complex societies — without these abstractions, man would have been imprisoned in small groups, not unlike bands of especially intelligent baboons.

But historically speaking, man not only developed these fascinating conceptual structures; he also devised suitable techniques for seeing to it that they were mastered. I would like to point out that, for the most part, one learns, but is not necessarily taught, to play with folk models. What is taught are the "rules of the game." Once the rules are understood, each participant is largely on his own except when the models are perverted by professionalism.

In every society there are social norms which distinguish between serious matters, on the one hand, and fun and games on the other. Generally, specific times and places are set aside for the enjoyment of these folk models. Also, the stakes for winning or losing are kept

strategy as models. The formal treatment of puzzles is not in as satisfactory a condition as are games of chance and games of strategy; however, we have suggested that the theories of natural deduction may be appropriate here (12). When it comes to aesthetic objects everyone is at sea and it is not known whether formal analyses of aesthetic objects, should such analyses prove possible, will result in one or more distinct classes of theories. We should bear in mind that until the work of von Neumann, no one was in a position to draw a theoretically cogent distinction between games of chance and games of strategy.

[4] It certainly is true that the ordinary citizen is the last person you should ask about how he sizes up situations and makes decisions. He has had little training in theoretical discourse, and it required the highest order of genius to carry through sound formal analyses of these familiar cultural products. It is interesting to note too that none of the four classes of models calls for very much in the way of technological expertise in fashioning the equipment used in connection with them—for example, bits of wood will do for checker pieces. It is clear that the physical side of these folk models is trivial, well within the technological competence of early man.

at some nominal value insofar as profit and loss enter in. In addition, there are norms which regulate expressions of feeling and emotion with reference to using these folk models. During the course of playing with a model, one is permitted to experience a fairly wide range of feelings and emotions, but extremes are excluded. These models, as it were, serve as a school for emotional expressions — this is the kind of school in which boredom is unlikely and uncontrolled emotional frenzy is forbidden. All in all, the set of norms governing the use of the models and the models themselves have proved so successful that people have to be prohibited from playing too much, despite the conceptual depth of the materials with which they deal.

IV

Now, someone might be tempted to ask the following question: "If you think so highly of these models as guides to action, and if they indeed represent abstractly so many salient features of human existence, why should we not rest content with them?" My answer is that these models did indeed serve man well during most of his history. However, there is something radically wrong with them with respect to their present theoretical relevance. So long as the ordinary man lived out his life within the context of a static social framework these models matched his world — the models themselves are essentially static entities. For instance, in any play of the game of chess the rules — that is, the boundary conditions — remain constant. There may be plenty of lively action going on within this stable frame of reference and the participants may feel a wide range of emotions, but the rules are both fixed and inviolable, in a normative sense. If you are working a puzzle, say a jig-saw puzzle, the picture to be completed does not change as you work on the puzzle, and the pieces preserve size and shape constancy. If you go to see a play two nights in a row, with trivial variations, it remains the same play — the actors do not change their lines because you have seen it before, though you may appreciate it more thoroughly on seeing it the second time (assuming it is an interesting play).

V

The basic point I am attempting to make is that the folk models mirror the static quality of unchanging or imperceptibly changing societies. The folk models in this respect are like the Newtonian conceptions of space and time — both presuppose a frame of reference which is invariant with respect to all that goes on within them.

I scarcely need to remind you that we no longer live in such an absolutistic universe. Since World War II there has been a massive acceleration in the rate of social change — an acceleration so great that it is difficult to graph it. This fact is now appreciated by most of us, although we are not prepared as yet to deal with its consequences. From this standpoint, it is not inappropriate to divide human history into two major periods, before and after this rapid acceleration got under way. The primitive period of human history ends in the 1940's and the modern begins then. It really does not make much difference what scientific, technological or social function you care to consider — each has gone off the graph and is shooting upward. We now live in a truly dynamic era and the most serious question which faces us is whether we can maintain a stable dynamic equilibrium.

Turning again to folk models, it is my point that they tend to inculcate an abstract conception of the world which is incompatible with a civilization in acceleration. The challenge is to create new models appropriate for these changed and changing circumstances — we need dynamic models — and these models must pass the stringent test of being enjoyable to their users. It is essential that people have to be admonished to put them aside.

I have given considerable thought to the problem of devising suitable scientific models to replace the traditional folk models. [See (13) as a first effort in this direction.] The "rules of the game" must be changing ones, yet it will not do if the rules simply change arbitrarily or irrelevantly. The successive sets of rules must bear some plausible relation to one another and to what goes on in the course of play. If you give serious thought to the problem of devising dynamic models, and I hope that you will, I believe that you will come to the conclusion that it will be very difficult to create such models without extensive employment of the latest developments in technology. We cannot interrupt the players over and over again to explain new sets of rules — it is no solution to substitute one fixed frame of reference for many fixed frames of reference. Instead, the player will have to discover that the old rules are, for some reason or another, not quite working — he will have to find out for himself that something has gone wrong and induce a new modus operandi that will suffice to meet the new conditions. To choose a trivial example, the jig-saw puzzle again, I do not see any way of making a dynamic jig-saw puzzle without employing some fairly complicated electronic equipment.

The final point which I wish to make has to do with testing. We not only want to create new and more appropriate models, but it is important to make certain that they have the desired consequence of helping to fashion a highly abstract, open and creative approach to the solving of dynamic problems. If we look at the development of tests from the standpoint that has been explained here, we can see that, generally speaking, they fall in the category of the puzzle model. Various I.Q. tests are cases in point. These tests consist almost entirely of sequences of short static puzzles. For this reason, assuming the folk-model analysis is correct, we should not be too surprised if the assessment of peoples' puzzle-solving abilities should have some predictive value for their behavior in everyday life. However, since most tests in current use do not measure skill in handling the other three classes of models, we also should expect a good deal of unexplained variance — and we are not disappointed.

Even if those who are interested in test construction were to build tests for the other traditional folk models — a demanding task in its own right — the opportunity still would exist for building tests for the projected dynamic models. Since these models are, for the most part, simply speculative notions at the present time, there is an opportunity for model construction and test construction to work hand in hand. And since the new models would have no traditional warrant, it is doubly important that they be tested as they are developed. It is my opinion that such new tests, like the models themselves, will require a full employment of our technology.

To recapitulate, let me say that there is a sense in which our conceptual grasp of what is now the subject matter of the behavioral sciences exceeded our grasp of what is now the subject matter of the physical sciences for most of human history. The folk models created by now-forgotten geniuses represent theoretical schemes which match the modern constructions of the physical sciences in subtlety, relevance, and applicability. These folk models are, strictly speaking, prescientific creations which were essential for man so long as he lived within the framework of fundamentally static social orders. They are antithetical, however, to the needs of societies in rapid acceleration. It is suggested here that what is wanted are models that are dynamic and will help in the creation of new approaches to problem solving. It is difficult to see how to make such models without the help of advanced technology. These models also must be tested and this requirement has many implications for test construction.

As a final thought, it may be well to recall the words of the perceptive Henry Adams. In 1905 he stated, in the context of a discussion of social acceleration, that this acceleration —

> prolonged one generation longer — would require a new social mind. As though thought were common salt in indefinite solution it must enter a new phase subject to new laws. Thus far, since five or ten thousand years, the mind has successfully reacted, and nothing yet proved that it would fail to react — but it would need to jump (14).

REFERENCES

1. Whitehead, Alfred North, *Science and the Modern World*, New York: Mentor Books, 2nd printing, p. 117, 1949.
2. Pierce, J. R., *Symbols, Signals and Noise*, New York: Harper Modern Science Series, p. 19, 1961.
3. Moore, O. K. and Kobler, Richard, *Educational Apparatus for Children*, Patent No. 3,112,569, United States Patent Office, filed May 29, 1961, granted December 3, 1963. Also, State of Israel, Patent No. 17,376, granted May 28, 1962.
4. Anderson, Alan Ross, "A Reduction of Deontic Logic to Alethic Modal Logic," *Mind*, 67: 100-103, 1958.
5. Anderson, Alan Ross and Belnap, Nuel D., Jr., "Tautological Entailments," *Philosophical Studies*, 13: 9-24, 1962.
6. Anderson, Alan Ross and Moore, Omar Khayyam, "The Formal Analysis of Normative Concepts," *American Sociological Review*, 22: 9-17, 1957.
7. Moore, Omar Khayyam and Anderson, Alan Ross, "Some Puzzling Aspects of Social Interaction" in *Mathematical Methods in Small Group Processes* (ed. by Joan Criswell, et al.) Stanford University Press, 232-249, 1962.
8. Moore, Omar Khayyam and Anderson, Alan Ross, "The Structure of Personality," in *Motivation and Social Interaction* (ed. by O. J. Harvey), New York: The Ronald Press, 167-186, 1963.
9. Anderson, Alan Ross and Moore, Omar Khayyam, "Autotelic Folk Models," *Sociological Quarterly*, 1: 203-216, 1960.
10. Anderson, Alan Ross and Moore, Omar Khayyam, "Modal Logics I: Toward a Formal Analysis of Cultural Objects," in *Boston Studies in the Philosophy of Science* (ed. by Marx W. Wartofsky), Dordrecht, Holland: D. Reidel Publishing Company, 117-143, 1963.
11. Neumann, J. von, "Zur Theorie der Gesellschaftsspiele," *Mathematische Annalen*, C 295-320, 1928.

page 67

12. Fitch, Frederic B., *Symbolic Logic*, New York: The Ronald Press Co., 1952.
13. Moore, Omar Khayyam, "Autotelic Responsive Environments and Exceptional Children" in *The Special Child in Century 21* (ed. by Jerrome Hellmuth), Seattle: Special Child Publications, 87-138, 1964. Paper delivered at the Second National Conference, summer 1962. Also issued as a report from the Responsive Environments Foundation, Inc., 20 Augur Street, Hamden, Conn.
14. Adams, Henry, *The Education of Henry Adams*, Sentry Edition, 1961, p. 498.

Session III

Theme: Testing the Culturally "Different"

Measurement Issues in the Counseling of the Culturally Disadvantaged*

MARTIN HAMBURGER
New York University

There are several vantage points for viewing application of measurement to problems concerning the culturally disadvantaged. An obvious one would be from within the science and technology of measurement itself, which is the context of this conference, the bias and the avenue of contribution of its sponsors, and the vested interests and professional legacy of its participants. Here, one might rush pell-mell but bravely into such matters as culture-free tests or the growing inventory of specific abilities and the search for their functional utility.

Another vantage point would be from the frame of reference of the culturally disadvantaged as a significant social force. This would involve, of course, our broader societal concerns with such problems as urban poverty, minority youth, civil rights, integration, unemployment of marginal labor groups, or manpower training. The urgent social and economic forces and the concomitant national concern lie most certainly at the core of the recent surge of interest among social scientists; the same forces dictate solutions much more frequently than the efforts of social scientists. A social-historical perspective may therefore provide a rapidly shifting definition of "disadvantage" which is sometimes a minority group problem or sometimes a function of economic and population trends that are shaking our social order.

No discussion of cultural disadvantage among psychologists and educators could ignore either the social-historical domain or the measurement technology which we employ to identify and describe disadvantaged individuals. Yet there is even another vantage point which

* The writer is indebted to J. A. Davis for critical review and editorial suggestions.

would seem particularly useful, for it brings to light certain crucial issues which have pervaded the rapidly burgeoning literature on cultural deprivation and at the same time reveals how effective resolution of those issues has been conspicuously by-passed. This vantage point is that of the counselor.

The term "counselor" is used here in the generic sense of the professional in schools and in public or private agencies who is concerned with the educational or vocational development of individuals. As such, the term subsumes many change agents who are dealing with the problems of human deprivation and learning deficit as well as with the compensatory programs designed to reverse the process of deprivation. But the unique aspect of the counselor's role, for our purposes, is that he must be dedicated to optimal development for all. The *all* is now beginning to include, in practice, the culturally disadvantaged. And the counselor *must* now view the disadvantaged as individuals who are underdeveloped in one way or another.

It is not the purpose of the present argument to suggest that there may be something magic in the counseling process which yields a brave and competent new self where fear and impoverishment existed previously. It would also be patently absurd to attempt to show that counselors have learned more about their measurement tools than the psychometrists who have developed and researched them. Nor can it be maintained that the typical counselor has a rich background of experience with the deprived, or that he is free from negative biases which result in a devaluing of members of groups who seem unable to achieve or to produce in standard ways. Counselors, like other helping professionals, have been employed by, and addicted to the views of, the "power structure." In fact they now find their job more difficult precisely because it more frequently includes the disadvantaged.

Today, as a nation in the midst of an ideological crisis, we are revising our premises about human potential as a result of social, economic, and moral revolution rather than new discoveries of unusual abilities or of new outlets for traits heretofore lacking in application. While the general public has other options, it is the counselor's function to guide his counselee through rebirth and regeneration, if necessary, and to help him find a place in the social system where his developed capabilities may be utilized. Face-to-face with disadvantage in the most personal sense, the counselor fulfilling his function must ask what measurement offers beyond the somewhat painful confirmation that disadvantage exists in measurable units and that the going may be rough.

Measurement is essentially a matter of science and technology. It is good science to describe and to attempt prediction from samples of behavior, permitting the objective findings and probabilities to guide further action. We may measure air pollution, for example, or frequency of smoking and predict the incidence of cancer, or we may observe the effects of harmful food additives or pesticides. In the application of measurement to the disadvantaged, however, we go beyond scientific issues. Although disadvantaging physical environments are clearly revealed through scientific research, the social impact of the proof provided by science can best be summarized as provocative of *some* moral indignation and very *little*, if any, action. Moreover, the science and technology of measurement of human traits are not yet at the level to enable us to make individual predictive decisions for most people except with wide degrees of latitude and error. Thus, with less than reasonable certainty of prediction, and with little comfort to be gained from labeling alone, the issues are evidently as much moral as scientific. In the current atmosphere of developing the great society, of combating poverty, of fighting inequity, there are salient issues of a political, social, and economic nature. It seems to me that these issues are scientific-technological, philosophical-ideological-moral, and political-economic-social. Let us consider each of these.

The Scientific-Technological Issues

It should be assumed at the outset that theory is powerful and that, the oft-proclaimed "lack of theoretical underpinnings" notwithstanding, certain theoretical formulations about human ability *have* influenced practice. The considerable body of data about human abilities and differences reviewed by Anastasi, Tyler, and Super and Crites provides good evidence of the existence and impact of theory. The more recent works of Hunt and Bloom reveal that implicit or explicit theories of human development, especially the development of intelligence, have influenced measurement, research, technology, and consequently counseling practice. In relation to problems of cultural disadvantage, the vast literature on the intelligence of Negroes demonstrates how seemingly neutral "scientific" data and "objective" theory are imbricated in everyday practice.

Even the most basic theoretical formulations have sometimes been responsible for widespread fallacies in the profession. One example is with the normal curve, so tremendously useful in describing majority populations. We expect, seek, and find those individuals to fill the tails of the distribution. Yet Tyler points out that "It is quite possible to change a skewed distribution into a normal one simply by making the test on which it is based a little harder or a little easier, depending upon the direction of the skewness." This is defensible for mathematical and practical reasons, but Tyler continues: "We know now that test scores can be manipulated to give us any sort of distribution we want."

The difficulty with the normal curve from the counselor's viewpoint is not with the test construction and the norming, however, but with the misconceptions that follow, or are fortified by, the presumably, immutable distribution of traits and talents. Thus, intelligence is seen as (1) a *dimension* on (2) a *scale*. Both notions have been attacked as scientifically fallacious and practically deplorable, especially by Hunt who argues that drawing samples of behavior for evidence of such organizational structures as schemata, operations, and concepts is more scientific than dimensionalizing and scaling within a vast domain of human functioning. If persons are open systems in which change occurs as a function of unspecifiable future conditions, then the use of *dimension* and *scale* in counseling is particularly serious.

Despite the evidence of Tuddenham and others that World War II soldiers performed considerably better than World War I soldiers on intelligence tests, the normal curve as an eternal, immutable verity continues to guide our perception of the particular distribution of traits in fixed and inevitable proportions. The counselor of the disadvantaged, however, can only wonder if the "normal distribution" has become a modern version of the Procrustean bed on which individuals have their standard deviations stretched or lopped off as desired. Thus, for the counselor of the culturally disadvantaged, the normal distribution, despite its utility in other areas, can only reinforce the notion that intelligence is like an oil reserve: there is just so much of it, and our job is to dig wells and retrieve oil, using our science to make the operation maximally profitable. If I may inject another metaphor of great relevance to counseling, we believe that we can't make a silk purse out of a sow's ear, but we ignore the fact that with modern chemistry we can! We cannot allow such notions to distract us from renewed and modern efforts to develop and transform people through education and other institutions.

Another problem results from confusing test performance with genotype or innate capacity. Hunt points this up most clearly. His proposal is that we think rather in terms of *phenotype*, that is, already actualized capacity. The research studies that deal with the proportions of intellectual development that are attributable to genes are most fascinating, but they may serve to inhibit efforts with deprived youth if the phenotype is confused at any point with the genotype. I feel that this is precisely what happens. This leads us to consider the most psychometrically respectful use of tests in counseling, that is, for purposes of prediction.

To follow Hunt once again:

> One implication of (the) idea that development . . . is an open-ended process puts a logical limitation on the predictive validity of tests of intelligence, or on measures of any personal characteristic. Prediction in science, however, is always a matter of stating what will happen to given objects in a closed system for which the relevant characteristics can be and are fully specified. . . . It may become possible to formulate laws which predict the characteristics that organisms with specified genetic constitutions will develop under specified programs of encounters with the environment. It would appear to be outside the realm of scientific possibility, however, to predict with precision the future characteristics or phenotypic fate of any organism from knowing merely its present characteristics, without being able to specify the future conditions under which it will live.

It seems to me that herein is to be found the central issue of all counseling, but especially counseling of the disadvantaged. The essence of the counseling transaction is the opening of the system — the phenomenal system — as well as the opportunity structure comprised of educational and vocational resources. This view of the counselor, the only view tenable under our working definition of his role, places great importance on testing for diagnosis rather than prediction. Measurement should be used "to improve status," as Ebel suggests, rather than "to determine status."

The emphasis on prediction, so well-founded for the purpose of test construction, validation, and especially selection, leads the counselor to an illusion about the nature of the "permanent" endowment that individuals bring to counseling. This chimera blocks the entrance to that mysterious realm which may hold the secrets of the developmental and transformational activity which successful work with the disadvantaged entails. Lest we conclude that we are discarding too quickly a rigor we have treasured in the past, we can also note the rather hard-headed conclusions of Tyler and Ghiselli that aptitude tests have very limited predictive value. That the success of traditional tests of scholastic aptitude for predicting achievement in the popular educational systems does not carry over to other tests and other criteria is supported by Goldman's recent review of vocational assessment. The single test on the one hand and the elaborate test battery on the other certainly merit the same criticism if either is geared to brief "counseling" interviews which do not actively engage the client as a participant in the process. Too frequently in the past the fusion of counseling and testing has amounted to a pattern of four or five hours of testing and an hour of "counseling." With the seriously disadvantaged, diagnosis and treatment are infinitely more important than the misplaced emphasis on testing and prediction. And, in this view, the more modern emphasis among leaders in test theory and construction on *construct* validity (e.g., Messick's work at ETS) seems more promising. We may fear, however, that it will be some time before the use of tests *to predict* and, in the case of the disadvantaged, the use of prediction *to select out*, is displaced by the use of tests for diagnosis toward understanding and treatment.

No consideration of the scientific and technological issues would seem complete without some examination of the culture-fair or culture-free tests. In the framework of the foregoing discussion, this approach promises to be a will o' the wisp insofar as it pursues the unproductive search for genotype. The culture-fair or culture-free test is dedicated to the proposition of genotype. If we think in terms of diagnosis, in terms of point-to-point phenotypical sampling and then of appropriate intervention, the usefulness for such instruments is minimal. To strain toward their development is not only to ignore largely that the majority culture will prevail in the environments in which the disadvantaged must operate, but also seems most clearly a plaintive attempt to demand proof of an adequate genotype before investing in any developmental program.

The Philosophical, Ideological, and Moral Issues in Measurement

Missing in many of the current policies and programs that are concerned with the disadvantaged is a fresh view of human potential. The ideology of *great expectations* must replace the seemingly objective but quite clearly ideological commitment to hurdles and cutoff described above. The economics of scarcity induced the psychology of scarcity. The economics of abundance has unfortunately given us very little more than the psychology of affluence; the psychology of abundant *ability* — of concern for creating development opportunities for all levels along the spectra of abilities as currently assessed — is not yet with us. Plainly, it is easy to attribute the wealth we have to a finite reservoir of both natural and human resources, and in ascribing the transformation of the former to the gifted portion of the latter. But we all know, too, that selection is a short-run concept. We achieve temporary success, as did Binet, in screening out of the regular system those who would deter the program's impact on the "appropriately talented." Such efficiency becomes a culprit when it becomes costly, when it results in the depositing of large groups who can only be a burden to the by-passing society.

The essence of the philosophical, ideological, and moral issues in the application of measurement to the disadvantaged lies in the supposition that our basic concepts of human ability have been drawn from instrumentation which has been developed to conform to a particular view of human nature. This view considers the essential development capabilities to be fixed, or genetically determined, with only relatively little change through environmental influences possible after the second or third year of life.

Thus, the counselor engages in a self-fulfilling prophecy: Starting with certain assumptions about human nature, about racial groups, about the etiology of disadvantage, he may use the genotype as the model of individual development. The great temptation in counseling is to inventory the sort of pragmatic solutions that society has provided, rather than to enter into the real effort that is involved in altering a process which has always seemed irreversible.

Another related issue is that we are so committed to traditional concepts and formulations and are so persuaded that change is slow that we fail to note those events or scientific researches which show radically different concepts and formulations to be vital, dynamic, and useful. For example, many people in this country harbored the illusion that the Soviets were mechanically inept; we know, because it is a part of our own recent past, what happened to this attitude when they put a Sputnik up before us. The evidence provided by Margaret Mead of the rapidity with which cultural change in the South Seas moved people from the Stone Age to the jeep are even more germane. More to our point today are the Iowa studies, much maligned for years, on the great accelerations that could be induced by conscious, planned processes in child development; or, to the same point, Kirk's studies with mentally retarded children. In a yet unpublished study which I have been conducting, we have found real and significant increase in IQ for 18-year-old experimentals (disadvantaged youth in a work-study program) over similar controls in a seven month period. There is indeed mounting evidence that intelligence is far from immutable.

Beyond these considerations, there is still another difficulty. What isn't answered by increases in IQ is the fact that increased intelligence without ready opportunity to function is not much of an outcome. The genotypically perfect plant can be encased in an otherwise perfect environment, but without a window to admit the sun there can be no tropism. The counselee, however fortified with training or increased capability to deal with the environment, still needs a window before self-actualizing movement can take place. The heliotropic client is the kind of success story everyone is looking for. We revere our notion of the individual who, despite every handicap, despite every conceivable kind of buffeting by fate, emerges at the end as invincible. The windowless room would be poor botany. This gross distortion of Darwinism is poor psychology. Yet this is the dominant ideology of limited expectations.

We know the jokes about the clerks who were made cooks and the cooks who were made clerks through the aptitude tests administered to inductees in World War II. But in real life, these people could, and did, function in the new roles. Super's basic premise over a decade ago, that "although each occupation requires a characteristic pattern of abilities, interests, and personality traits, there are tolerances wide enough to allow some variety of individuals in each occupation and some diversity of occupations for each individual," may be an understatement. We may recall Jones' classic work which exposed the fact that the presumed decline in intelligence of the aging was an artifact of the tests employed. Our commitment must be to the individuals and the job; we need to look for capabilities to occupy productive roles without the restrictiveness which our conventional notions of testing and fitting may impose. While such examples may seem remote, we have the evidence from Rosenthal's studies of experimenter bias that even the most careful researcher tends to get the kind of results that he wants. The influence of ideology makes the counselor susceptible to the same kind of biases, only in his case he perpetuates a wide class of human behavior.

Political, Economic, and Social Issues in Measurement

The problems of disadvantage are frequently matters of power and power structure, with the disadvantaged being powerless as well as deprived and poor. Nominal or real disenfranchisement is often correlated with minority group status and poor education with nominal or real segregation. These economic, political, and social issues are involved in measurement when, through pupil placement laws, tests are used to fortify a social system. The *de facto* segregation of students by school and by ability groupings provide significantly lower levels of expectation, experience, and expression for teachers and students. The teacher turnover is greater, the level of staff competency is lower, and the quality of instruction in the very situation that demands the most skill and versatility is poorer. This becomes part of cultural deprivation for the students and serves to close rather than open the system.

The fact is that we can now afford to train and to develop people, and we cannot afford not to train and develop people. That political leader and friend of education in the South, O. Max Gardiner, is reputed to have said that "Good education is expensive — in fact, the only thing more expensive is ignorance." Our economics of abundance must be matched by a psychology of abundance. We are obtaining the financial means for the programs and for the research simultaneously; the research must not lag.

Implications for the Professional Role of Counselor

The three areas of concern about measurement issues in work with the disadvantaged fuse nicely into an indication of what the *scope* of the professional role of the counselor must be. As a pragmatic scientist, he cannot afford to divorce his focusing on a particular problem from the total context. The striking thing about the really great physicists who have made important contributions to modern theory is their general knowledge of total nature and their pervasive moral concern. This will be the crucial factor in determining how long it may be for the morality to become actuality.

Chauncey and Dobbin's discussion of guidance in the American public schools is a most eloquent statement of what *might* be a good model. The way in which testing is harnessed to guidance, however, confines the utility of this model to the majorities. It does not fit the conception or the practice of counseling with the culturally disadvantaged. Indeed, the role in which measurement in counseling has been cast by Chauncey and Dobbin is one of amelioration. Perhaps coming to grips with reality (and using tests as one source of information about important realities) is the first step in changing those realities that the individual does not want to accept. With the disadvantaged, however, we are concerned about the realities of poverty, of segregation, of street corner gangs. The security blanket that testing normally provides counselors wears thin simply because the problem is so vast.

The great biologist Dubos has pointed out that laboratory science in his area is insufficient unto itself to ease the ills toward which it is directed. He says "The most effective techniques to avoid disease came out of the attempts to correct by social measures and by sanitation and public health. The introduction of cotton undergarments easy to launder and of transparent glass that brought light into the most humble dwelling contributed more to the control of infection than did all drugs and medical practices." Modern microbiology is superb, but it is poor sanitary practices that are responsible for periodic outbreaks of staph. Even if the Chauncey-Dobbin model and the Mental Measurements Yearbook give us the illusion that we have a pharmacopeia, and though testing as ritual seems to cleanse the counselor and the client at once, infection still lurks just outside the door.

In this sense, testing would appear to be very much secondary to working with the individual and through the environment to improve the opportunity structure for the disadvantaged. One does not have to ignore totally the knowledge of human differential psychology that we have acquired. The assessment of, say, a nineteen-year-old high school dropout as having an IQ of 70, a 11.8 grade-level score on paragraph meaning on the Stanford Achievement Test, and a 7.8 grade-level word meaning score provides some very cogent suggestions about the way this individual functions, and may aid in the design of remediation activity, or better, of developmental activity. But still another factor is that few practitioners of counseling have the knowledge and experience required to use tests in this manner. The use of the most gross measures, the reliance, in administrative decisions, on the global score (and this is what actually happens in the long run), is contrary to the counselor's purpose of increasing human variability.

I am not impressed by the evidence, therefore, that statistical and actuarial methods are superior to clinical or instructional techniques in counseling activities with the disadvantaged. Here the issue is erroneous and simply not appropriate. Testing in this view is preliminary and incidental to helping, to modifying, to teaching, and to re-educating; the view of testing as fact-finding and prediction, I maintain, leads to the outclassing of instruction.

The assessment-appraisal model is therefore necessary but insufficient to the performance of the counseling function in work with the disadvantaged. The auto-education model is necessary and insufficient. The decision model is spurious in this instance because of the powerlessness of the clientele and because the counselor must become a decision surrogate.

On the other hand, the model of the physician, the social biologist, the engineer, the therapist, and particularly the teacher are all meaningful and afford us important reference points for the evolution of the professional role of the counselor. Yet in work with the disadvantaged particularly, counselors faced with large numbers of children must evolve a unique role. The use of measurement and science is fundamental to such an evolution, but it must be science, not scientism; profession, not technocracy; diagnosis, not prognosis; commitment, not impartiality; effort, not expediency; and the contrivance of new systems of development, not reliance on those already available. In this sense, the poet may have more to say than the so-called scientific psychologist: I recall Thomas Gray's

> Full many a gem of purest ray serene,
> The dark unfathom'd caves of ocean bear:
> Full many a flower is born to blush unseen,
> And waste its sweetness on the desert air.

Counseling, of course, cannot provide the idyllic state envisioned. Piaget is correct in stating "The mental functions of human beings develop in the process of learning . . . they develop into hierarchical structures, with the more advanced systems arising during the course of the child's experience only if earlier formations have provided the foundational elements." This is the nature of the beast, the kind of process in which we must achieve effective intervention. Testing can provide a useful, though incidental, tool, but we must know full well that there are political, social, and moral responsibilities inherent in each act of application of a measurement device, especially in dealing with the focus of our present discussion, the disadvantaged. The simple improvement of the science and technology of measurement cannot in itself provide the kind of results we must achieve.

Principles of Developing Tests for the Culturally Different

William E. Coffman
Educational Testing Service

This is not the first time that the problem of the measurement of the culturally different has engaged the attention of participants in the Invitational Conference on Testing Problems. The first panel at the 1949 Conference dealt with the topic, "Influence of Cultural Background on Test Performance," and much of what was said at that time is equally relevant today. The second panel of the 1952 Conference focused on the topic, "Techniques for the Development of Unbiased Tests," and again, what was said might be repeated in today's discussion. In fact, one may, with some justification, wonder why we're at it again today.

The fact of the matter is, of course, that the topic is a highly complex one and deals with a central issue in the field of testing, namely, since individual differences are a function of many factors, how does one insure that in any particular test relevant factors are being measured and irrelevant ones excluded? Like the issue of essay versus objective testing, it is one which needs to be examined periodically in the light of technical advances in the field and of social changes taking place in the context within which the issue is imbedded.

At the 1949 Conference, Turnbull defined one position on the issue in these words:

> It is my contention that on a predictive test any score difference between groups whose backgrounds differ should be judged, not good or bad, nor right or wrong, but useful or not useful, valid or invalid for the prediction of future behavior. (1)

On the same panel Haggard stated the position which has been expounded, perhaps with somewhat greater intensity but certainly with no greater clarity, by critics of testing during the past year. He said:

> The thesis here is that we cannot assume the various sub-cultures in America to be comparable, simply because of a common geographical boundary. Nor is it enough to say "We consider all that when we evaluate the I.Q. of a lower-class ethnic child." Why not? Because all too often the educational opportunities — from the early grades on — are determined by how well a child does on our present standardized tests, regardless of whether we intend them to be. And those who do poorly at first are often given inferior educational opportunities, so that a vicious circle is set up, and a great deal of potential ability is lost to our society. (2)

For Turnbull, the issue was whether or not differences in the performance of different groups on a test are significantly related to prediction. If they are, then groups should be treated separately when making predictions, or the testing procedure should be modified to eliminate distorting differences. But first, let's be sure that there is a problem, that is, that the regression lines for predicting a future performance are actually different. If they are the same, then how can one talk of bias? For Haggard — and for many critics of testing today — the issue is much broader than simply one of maximizing prediction. Test scores can be misused by uninformed people to perpetuate social injustices. Even where the evidence indicates that sub-groups should be separated and treated differently in making predictions, the differentiations are often not made; and even if there is no evidence that regression lines are different — well, a prediction of the usual academic criterion may not be so important if by making it we eliminate at the start those who might achieve in the long run. We may grant the predictor his point; the trouble is that his criterion is almost never an ultimate one.

Essentially, the issue is one involving values; that's why it is so persistent and so resistant to the impact of research. Proponents of different points of view tackle different kinds of problems and assemble different kinds of data. It takes a long time for the collections to begin to converge. On the other hand, as one looks at the issue today in the light of what can be learned from research completed and in progress, some converging is discernible. I believe we are ready to agree that a test which does not provide a valid prediction of some kind of significant behavior is not worth the time it takes to answer the questions. At the same time, however, I think we're ready to accept the proposition that pure empirical prediction, irrespective of possible, indirect social effects, is indefensible. Tests should be construed and used with indirect consequences as well as predictive effectiveness in mind. Finally, I think we've agreed that we have an educational problem on our hands. We can't concentrate all our efforts on making foolproof tests; we have to do something about improving the sophistication of test users. My specific responsibility this afternoon, however, is to discuss test construction, not the education of test users.

Historically, both the College Entrance Examination Board and Educational Testing Service have been aware of the significance of cultural differences in the interpretation of test scores and have mounted systematic programs of research designed to clarify such differences. The differences recognized in the *First Annual Report of the Commission on Scholastic Aptitude Tests, 1926* are obvious ones:

> At some time in the future, then, it may be expected that differential weights will be applied to the tests for varying purposes. Boys and girls, engineering applicants, and liberal arts applicants would all take the same examination under identical conditions and their scores would all be expressed on a scale in which the mean is 500 and sigma 100, but the various tests determining the total score would be differentially weighted depending on the sex of the applicant or the type of curriculum to be taken. (3)

Since that pioneer effort, literally thousands of studies have been carried out in which the Scholastic Aptitude Test or tests of the same type have been incorporated, along with other measures, in equations for predicting the academic performance of specific groups. Member colleges and colleges contemplating membership in the Board have been encouraged to conduct studies in which sub-groups that might be expected to show different predictive relationships are treated separately: men and women, liberal arts and engineering students, independent and public school applicants — logically, there is no limit to the breakdowns one might try. There are, however, limitations of a practical nature. It may be that students from that little high school over in the neighboring state, where all the teachers have Ph.D.'s and where there are a maximum of ten students in a class, might behave differently. The trouble is that a college may get one or two applicants every other year from that school. And as for the sons of immigrant farmers from Faraway County, where there is one high school with two teachers neither of whom has yet completed the bachelor's degree, the situation is even less promising; a college may have never had a candidate before, and this year it may have only one.

It doesn't take much imagination to realize that if we expect each college to conduct its own studies and develop separate predictive equations for every significant sub-group of the applicant population, we're being unrealistic. Often, there simply weren't enough cases available to establish a line of relationship. Inevitably, there comes a point where the test user is forced to make clinical judgments as he encounters individual cases which seem not to fit into any of his previously studied categories. This is not necessarily an undesirable practice; actually, it may be the only way to temper the impact of large-scale data processing methods on the unusual individual.

But isn't it possible that the problem may be reduced somewhat by a direct attempt to eliminate irrelevant cultural differences at the point of test construction? And if the attempt is not entirely successful, is it not possible to at least identify potential problems for the test user and provide him with some guidelines for his clinical judgments? The College Board and ETS have thought so and have acted accordingly.

With the development of large-scale data processing equipment, the question has naturally arisen whether or not it would be possible to pool data across a number of colleges and thus get around the problem of too few cases in a sub-group. To date, the explorations of this approach have not been particularly fruitful; furthermore, the refinements, which appear possible by separating for analysis special groups of candidates, are quite small in contrast to the high, over-all validity of the test scores. SAT scores appear to be equally valid for groups in which the average score is above 600 and for those in which the average score is below 400, in colleges enrolling youth from quite homogeneous backgrounds, and in those enrolling youth from a wide range of backgrounds. This is, to some extent at least, attributable to systematic efforts at the test construction stage to eliminate unnecessary and irrelevant cultural factors from the test.

To a considerable extent, efforts have been guided by rational considerations; but, particularly in recent years, systematic research has guided our efforts. Let me cite a few examples. The early forms of the SAT were made up of sub-tests, each containing so many questions that nobody could finish in the time allotted. We knew from studies reported in the general literature of psychology that a rural youth is likely to respond differently from the way an urban youth will to a speed task. We therefore set about systematically to reduce the weight of the speed factor in the test.

We knew that some parents could pay for special coaching in preparation for the test while others could not, and that certainly the availability of practice exercises would vary widely. Thus, we decided to design item types which were relatively unresponsive to special coaching experiences. And we collected data indicating that our efforts had been reasonably successful. (4) At the same time we prepared booklets describing the test and providing enough illustrative questions to insure candidates of the opportunity to know the kinds of questions to expect. (An order form for these free booklets is mailed to every secondary school in the country at the beginning of each school year.)

We knew that familiarity with the general content of a reading passage would improve a candidate's chances of understanding the passage and answering questions about it. We also knew that the reading interests of candidates would determine to a considerable extent the kinds of materials they would read. Some would favor literary reading, others scientific; some would read agricultural magazines, others would read magazines aimed at the homemaker. Our studies suggested, for example, that girls were more likely to do well with material in the biological and medical sciences while boys excelled with material in the physical sciences. (5) We therefore set up categories to guide our selection of reading passages which would insure a balance of content. We learned that women excelled in knowledge of words related to people while men tended to excel with vocabulary related to things. (6) And we reasoned that regional and cultural backgrounds would also be reflected in the vocabularies of different candidates. We therefore established categories to guide our selection of the verbal omnibus questions in the test.

In order to increase our understanding of the cultural factors which influence test scores, we have in recent years deliberately sought out samples from widely different backgrounds and studied their responses to questions. We have studied samples from Scotland and from Alberta, from Nigeria and from East Africa, from the rural Midwest and from New York City, and from segregated centers in the South.

As our questions became more focused, we realized that we had methodological problems to solve. The central problem had been anticipated by Tukey in the discussion of the papers at the 1953 Invitational Conference with this question: "Isn't the main interest of this operation the comparison of these groups, the interaction between status and other variables?" (7) At that time, Eells and his associates at the University of Chicago had completed studies in which the chief method was that of comparing the percentage of one group marking a correct answer to a question with the percentage of another group marking a correct answer to the same question. There were two limitations to this approach. In the first place, since a number of different questions had been answered by the same people, it was improper to make the tests of significance as if the successive differences were based on independent samples. The replies to the different questions are necessarily correlated. Each new analysis did not yield a new piece of information. In the second place, the effect one wished to study — the interaction between status and test questions — might possibly be confounded with real ability differences between the groups.

We have not yet solved all the problems of sorting out the various factors contributing to differences in test performance of different groups, but we have made some progress and are continuing our studies. During the summer of 1963, Cardall (8) applied the analysis of variance design for two factor experiments with repeated measures on one factor to data from three different groups and showed how mean effects and interaction effects might be separated. She demonstrated significant interaction effects for items in the SAT across the three groups. At the same time, she pointed out that the statistical analysis did not permit us to differentiate between balance and bias in a test. We are continuing our search for such a differentiating procedure.

In the summer of 1964, Fremer (9) examined Cardall's data subjectively for evidence of balance or bias and concluded that whatever factors were causing the significant interactions were not readily apparent from an examination of item content. In a few instances, item content could be related to relative difficulty. For example, we plotted the item difficulties for a sample of candidates from New York City against those for a sample of candidates from communities of less than 15,000 population in Illinois and Indiana and more than fifty miles from Chicago and Fort Wayne. Only two of the 42 verbal questions fell outside a clearly defined ellipse. Both were easier for the Midwestern sample and both involved content with a rural flavor. But by and large, the factors producing changes in the relative difficulty of questions from group to group appeared to be the result of complex factors such as those producing individual differences within groups. And since the interactions tend to be represented in the scatter plots by ellipses in which the points are distributed symmetrically, we have tended to favor interpretations in terms of balance rather than bias.

When we have compared samples from outside the United States with those from within, we have found a wide variety of patterns. English speaking samples from Scotland and Canada produce interactions no greater than those from different groups within the United States. On certain types of mathematical aptitude questions typically found in school textbooks it appears that the Scotch sample may have an advantage. When we examine samples for whom English is a second language and where differences in culture are immediately apparent, as with our African samples, however, both mean differences across all items and interaction effects indicating changes in the relative difficulty of items are dramatic. Details of our findings have been presented elsewhere. (10) Essentially, we found that there was a large difference in over-all performance and a large shift in relative difficulty of questions, both attributable to the fact that English was a second language. The shifts in relative difficulty were also attributable to the fact that the individuals in the sample had grown up in a rural, tropical society rather than in an industrialized country located in the North Temperate Zone. We reasoned, soundly, I believe, that many of these differences were irrelevant to our purpose in assessing aptitude for undergraduate college studies in the United States. We therefore set about constructing a special Scholastic Aptitude Test from which inappropriate items had been excluded. Our evaluations of this effort are still in progress. Meanwhile, we have recognized that no amount of tinkering would produce a valid test on which African and American students of equal ability would make equal scores. The cultural differences are so great that the Africans must be treated as a separate group throughout the selection and placement process.

For the most part, in assessing the validity of the test for African candidates, we had to be content with the testimony of African educators — that their better students tended to make higher scores on the test. We did locate one sample for whom both PSAT scores and Cambridge School Examination marks were available and demonstrated that the correlations between total Cambridge marks and PSAT-V and PSAT-M scores were similar to correlations between school grades and PSAT scores for samples in the United States. Also, we found sixty pairs of African students for whom grades in an American college were directly comparable since each member of a pair attended the same college and was following the same curriculum. A sign test indicated a positive relationship between grades and PSAT-M scores even though a systematic effort had been made to place students of like ability in the same college. Presumably, other things being equal, the higher the score, the more promising the candidate.

The data from our African studies led us to become concerned about other College Board candidates for whom English is not a native language. Studies by Howell (11) have provided evidence from which we have developed a frame of reference for interpreting scores of such candidates. Thus, the booklet describing the SAT this year contains a section headed, "If English is not your native language," which reads:

> The Scholastic Aptitude Test is based on the assumption that the test-taker is a native speaker of English. The verbal questions on the test have been designed specifically to probe the kind of understanding of the English language that is acquired by one who has spoken it all his life. The SAT, in other words, tests not so much your "academic vocabulary" as your total mastery of the language. This means, of course, that the SAT is not entirely suitable for testing all foreign students, many of whom will take the test without having acquired this kind of understanding of English. The extent of this handicap obviously will vary depending upon the degree of familiarity with, and mastery of, English.
>
> The College Board has taken steps to make admissions officers in American colleges more aware of the language handicap some students may have. If English is not your native language, and you are applying for admission to an American college that requires the SAT, you should inform the admissions officer of that college of the extent of your experience with English. Information of this kind will help him make a proper interpretation of your test scores. (12)

Our studies have shown generally that responses to the questions in the mathematical sections of the SAT are relatively unaffected by differences in cultural background. Unlike the verbal aptitude score, which reflects many and varied influences within a culture, the mathematical skills and abilities are generally developed in school. Such differences as do appear are likely to be reflections of curricular differences across schools rather than linguistic, geographic, or socioeconomic differences except as these may be indirectly reflected in school practices. We did find significant shifts in the relative difficulty of SAT-M questions when they were answered by our African samples. Apparently, there was much greater emphasis on routine computation in the African schools and relatively less emphasis on what we refer to as "novel" items. And our indices of interaction were statistically significant although relatively small when we compared samples from different groups within the United States. It is interesting, however, to note that Cardall reported interactions of the same magnitude for three strikingly different samples. Apparently, there is no greater difference in the mathematical curricula between Negro schools in the Southeast and non-segregated schools in the North than between schools in New York City and schools in the rural Midwest.

The curricular developments in secondary schools at the present time are generating groups of "culturally different" candidates so far as tests of academic achievement are concerned; and much of our research and development effort during the past several years has been devoted to dealing with problems arising from this source. These efforts are simply the extension of a policy adopted by the College Board in the early forties when the Board moved from published syllabi, and tests designed to cover them, to achievement tests designed to measure the outcome of good instruction. This policy left wide flexibility for variations in details of courses from school to school.

The problem of validation of an achievement test is somewhat different from that for an aptitude test. While it is true that the College Board Achievement Tests have proved useful in predicting performance in college, there are those who will argue, with justification, that the tests should be considered assessments rather than predictors. As an assessment, the validity of an achievement test rests on the judgments of competent authorities that it requires of the candidate the performance of a representative sample of the tasks toward which the instruction of the school is legitimately directed.

But how shall one decide who are competent authorities and what can be done about providing those authorities, once identified, with the data necessary to arrive at sound judgment? In the selection of committees of examiners, in the developing of specifications for each new test, in the detailed test analysis prepared following the administration of each new test, and in special research studies undertaken from time to time, both ETS and the College Board are seeking to answer this question. There is only time to mention a few examples of steps which are being taken to insure fairness to candidates with different curricular backgrounds.

One thing we have done is to provide committees of examiners with information about what is being taught in the schools. Such information often leads to changes in specifications for tests. Thus, a study of the preparation of mathematics candidates provided one basis for changes in the mathematics achievement tests during the current academic year. (13) A great many similar studies have guided committees in mathematics and in other fields over the years.

Another contribution to the construction of tests which sample fairly what is taught in the schools has been systematic comparisons of the performance of candidates with different curricular backgrounds. Several years ago, for example, we began offering a special examination for candidates who had studied the PSSC physics course after studies had indicated that such students were at a disadvantage on the regular Physics Test. At the same time, we set about designing a test which would be appropriate for all groups of physics students. Now we are again offering a single physics achievement test.

More recently, recognizing the primary place of expert judgment in determining the appropriateness of achievement tests, we have assembled and analyzed the judgments of teachers of various courses with respect to the appropriateness of questions in the College Board tests. (14)

Finally, we have undertaken a study of the relationship between curricular background and test performance which involved all of the College Board Achievement Tests. It is our hope that the data from this study will provide us with a vastly richer frame of reference for building specifications to insure that the tests reflect accurately the learning of all candidates.

I have wandered over a wide territory, and perhaps it would be a good idea for me to summarize what I think I've been saying about principles of developing tests for the culturally different. First, since test scores are often misinterpreted or used in ways which reflect unfavorably on the culturally different, whenever possible without serious reductions in validity, differences should be eliminated at the point of test construction. Second, if the test is a predictor and if sub-groups of sufficient size are available, the validity of a test should be determined separately for each sub-group. Third, if cultural differences remain after careful test construction, the fact should be communicated to test users with guides to proper interpretation of test scores for those who are culturally different. Finally, when assessments rather than predictions are involved, the basis of comparison should be expanded to include the judgments of different groups of experts as well as differences in responses to test items by different groups of test-takers.

REFERENCES

1. Turnbull, William W., "Influences of Cultural Background on Predictive Test Scores," *Proceedings, 1949 Invitational Conference on Testing Problems*, p. 32.
2. Haggard, Ernest A., "Influences of Cultural Background on Test Performance," *Proceedings, 1949 Invitational Conference on Testing Problems*, p. 19.
3. Brigham, Carl C., *A Study of Error*. New York: College Entrance Examination Board, 1932, p. 335.
4. French, J. W. and Dear, R. E., "Effects of Coaching on an Aptitude Test," *Educational and Psychological Measurement*, XIX, 319-329 (autumn 1959).
   Malcolm, Donald J., "Summary of Studies Pertaining to the Effects of Coaching on the Performance of Students on the Scholastic Aptitude Test." (Unpublished report, March 21, 1961.)
5. Coffman, William E., "The Effects of Lapse of Time on Item Statistics for the Contemporary Affairs Test," *The Tenth Yearbook*, National Council on Measurements Used in Education, 1953, pp. 58-65.
6. Coffman, William E., "Sex Differences in Responses to Items in an Aptitude Test," *The 18th Yearbook*, National Council on Measurements Used in Education, 1961, pp. 117-124.
7. *Proceedings, 1952 Invitational Conference on Testing Problems*, p. 130.
8. Cardall, Carolyn and Coffman, William E., "A Method for Comparing the Performance of Different Groups on the Items in a Test," *Research and Development Reports*, 64-5, No. 9, CEEB (November 1964).
9. Fremer, John (study in progress).
10. Coffman, William E., "Evidence of Cultural Factors in Responses of African Students to Items in an American Test of Scholastic Aptitude," *Research and Development Reports*, CEEB (June 1963).
11. Howell, John J., "College Board Scores of Candidates of Non-English Language Background Tested at Foreign Centers," *Research and Development Reports*, 64-5, No. 2, CEEB (August 1964).
12. *A Description of the College Board Scholastic Aptitude Test*, CEEB, c1964, pp. 7-8.
13. Malcolm, Donald J., "Preparation in Mathematics of Candidates Who Take the Intermediate or Advanced Mathematics Tests," *Research and Development Reports*, CEEB (June 1963).
14. Malcolm, Donald J., "Study of the Relevance of Items in CEEB Science Tests for Special Curricula," *Research and Development Reports*, CEEB (November 1961).

The Measurement of Environments

RICHARD WOLF
University of Chicago

In the course of preparing to address this august assemblage, I sought guidance from the *Proceedings of the Invitational Conference on Testing Problems* for the past fifteen years. This bibliographic search served two purposes. First, I was interested in obtaining some estimate of the level of complexity of the papers so that I would not grossly underestimate the sophistication of the audience. (I had already assumed that it was virtually impossible to overestimate the intelligence of this group.) Second, I was curious as to whether the subject of measuring environments had been previously treated and, if so, to what extent, since I didn't want to go over previously covered ground.

My search produced the following results: In the 1949 Invitational Conference, a panel composed of Anastasi, Haggard, Stephenson and Turnbull addressed themselves to the problem of the influence of cultural background on test performance (6). The panel discussed at length differences in test performance for various cultural groups and considered the problem of test items which seemed to differentiate various cultural groups. The reports of the panel members as well as the lively discussion that ensued have fortunately been preserved in the *Proceedings*. In 1956, Bayley delivered a paper on "A New Look at the Curve of Intelligence," which contained a section on some environmental determiners of intelligence (2). Unfortunately, a time limit prevented Bayley from developing this section of her paper more fully. Since 1956, the record indicates that the topic of environmental measurement has not been dealt with. This seems to me to be rather unfortunate, and I shall endeavor to draw your attention to this rather neglected area.

There is abundant evidence th[illegible]the behaviorial differences we find in test performance ar[illegible] result of differences in the environments in which indivi[illegible]d rather than of inherent differences in individuals the[illegible]finding has been demonstrated rather spectacularly i[illegible] (13) and Kirk (12), and most recently in the work o[illegible]uglass in England (5). Hunt (9) and Bloom (3) have [illegible]arized the results of a number of such studies. However, d[illegible]s accumulation of evidence, we have persisted in our view [illegible]ing in essentially individualistic terms. That is, the individual is still considered an "X" to be solved, and the environment in which the individual has lived is considered as but one source of error which can be handled by appropriate weighting of test results or some other procedure. This view of testing is reflected in Buros' *Mental Measurements Yearbook*, which contains descriptions and reviews of thousands of instruments for the measurement of individual characteristics, but very few for the measurement of environments.

This is indeed an unfortunate state of affairs. All theories of learning and behavior make provision for the influence of the environment on the development of human characteristics, but, as noted above, we have not had a corresponding emphasis in our measurement procedures. I would also submit that we have rarely attempted to systematically relate individual test data to environmental data in ways which are designed to increase our understanding of the interactive process between the individual and the environment.

To be sure, we do have some environmental measures. However, the number of instruments designed to measure characteristics of the environment is quite small when compared with the number of instruments designed to measure characteristics of the individual. Also, those environmental measures that are available are usually limited to general measures of social status or economic well-being. However, it would seem that just as a general measure of intelligence or IQ has obscured many important differences among individuals, so a general index of social status or economic well-being has obscured many very important differences among environments. The work of Kahl (10), for example, has been quite revealing in its findings of some of the differences which are to be found *within* a given social class. There is, however, an even more basic difficulty in the use of a general index of social status or economic well-being to characterize an environment. Such indices usually represent a summation of a number of symptoms or surface characteristics of an environment and, as such, give little information about the specific ways in which environmental factors might affect the development of specific behavioral characteristics. This point was emphasized by Bayley in her 1956 address and was one of the central themes in Anastasi's paper, "Heredity, Environment and the Question, 'How' " (1).

In our work, we have attempted to follow some of the guidelines for environmental research suggested by Anastasi, Bayley, Bloom, and others. In doing so, we have utilized a conception of the environment which is different from that used in the development of previous environmental measures in four ways:

1. Instead of viewing an environment as a single entity, we have postulated that a single physical environment may be made up of a number of sub-environments, with each sub-environment operating to influence the development of a specific characteristic. For purposes of measurement, we would conceive of an environment for the development of stature, another environment for the development of general intelligence, another for the development of independence, and so forth. Thus, for us, the problem of measuring an environment was reduced to the identification and measurement of those aspects of the total environment which were likely to be related to the development of selected specific characteristics.

2. Rather than measuring some of the more surface manifestations or symptoms of an environment, we have been interested in measuring environmental variables which we have hypothesized would be likely to *directly* influence the development of specific characteristics. This departure from some of the more conventional approaches has, in the case of measuring the home environment, for example, resulted in our investigating what parents *do* in their interactions with their children rather than what parents are in terms of status level of father's occupation, type of dwelling, source of income, and so forth.

3. We have attempted to summarize and treat environmental data through the use of psychometric procedures. Instead of isolating any particular environmental variable which might be related to a particular characteristic, we have attempted to sample a variety of the processes and conditions in an environment which were hypothesized to be related to a characteristic, and to summarize these in describing an environment. This is essentially the same procedure we follow in testing. That is, we would no sooner describe a student's performance in a certain subject by his response to a single test item than we would

describe an environment on the basis of a single feature of that environment. We have felt, in other words, that a summarization of a number of variables is as important in describing an environment as a number of test items is in describing a student's competence in a subject.

4. We have attempted to systematically relate measurements of the environment to measurements of the individual. This has been an extremely important feature of our work. A number of investigators have developed environmental instruments which attempt to measure selected aspects of an environment, but have not related environmental data to individual test data. In our work, the importance of systematically relating the two kinds of data can be indicated by noting that one of the criteria for the validation of our environmental measures has been individual measures of the characteristic under study.

To briefly summarize, our approach to the measurement of environments has been characterized by conceiving of specific environments for the development of particular characteristics, attempting to measure environmental variables which were hypothesized to be directly related to the development of particular characteristics, summarizing and treating environmental data through the use of psychometric procedures, and relating environmental measurements to individual measurements. The remainder of this paper will be devoted to describing the development of two environmental measures based on the above ideas and presenting some of the results we have obtained.

While environmental influences may range from the most immediate social interaction to the most remote cultural forces, we chose the home as the physical environment for our study. The reason for this was our assumption that the home produces the first, and perhaps most insistent, influences on the development of the characteristics we were concerned with — general intelligence and academic achievement.

The first step in developing our environmental measures was to compile a list of the conditions and processes in the environment which theory and previous research indicated were likely to influence the development of general intelligence and/or academic achievement. The areas from which such variables were drawn included child development, learning theory, motivation, and psychometry. Special efforts were made to select variables which were as close as possible to being environmental counterparts of the components of the individual characteristics. The environmental variables which were identified as likely to be related to academic achievement were: the climate created for achievement motivation; the opportunities provided for verbal development; the nature and amount of assistance provided in overcoming academic difficulties; the activity level of the significant individuals in the environment; the level of intellectuality in the environment; and the kinds of work habits that are expected of the individual. Environmental variables identified as likely to be related to general intelligence were: the stimulation provided for intellectual growth; the opportunities provided for, and emphasis on, verbal development; and the provision for general types of learning in a variety of situations.

No claim is made that these two lists of variables exhaust the range of conditions in the home which can influence the development of academic achievement and general intelligence. Rather, they are viewed as a sample of environmental conditions which were hypothesized to be related to the development of these characteristics.

There are two notable features about these two lists of environmental variables. First, there is a considerable overlap between the two lists, which is largely a reflection of the overlap between the two characteristics, as has been pointed out by Kelley (11) and Coleman and Cureton (4). Second, the variables were stated in quite general terms. This lack of specificity made it necessary to define each environmental variable in an operational form for purposes of measurement. Thus, for each variable, we developed a list of process characteristics consisting of specific behaviors of parents and others in the home which were likely to be related to general intelligence and/or academic achievement.

An illustration or two is perhaps in order. One of the variables hypothesized to influence the development of academic achievement was the climate in the home for achievement motivation. The specific process characteristics which were defined as comprising this variable were: the parental aspirations for the child's education; the parents' own aspirations; parental concern for academic achievement; the social press in the home for academic achievement; the rewards accorded academic accomplishments; parental knowledge of the educational progress of the child; and the preparations made for the attainment of educational goals. The breakdown of each environmental variable into specific measurable characteristics was, in each case, based on relevant theory and previous research. In a similar way, the variable termed "opportunities" which was provided for verbal development — one of the three variables hypothesized as comprising the environment for the development of general intelligence — was analyzed into the following process characteristics: the emphasis on use of language in a variety of situations; the opportunities provided for enlarging vocabulary; parental emphasis on correctness of usage; and the quality of language models available.

Once the process characteristics comprising each environmental variable were specified, it was necessary to develop a set of procedures for gathering evidence about the strength of each characteristic. The procedure finally settled upon involved the use of an extended interview with the mother (and sometimes both parents) in the home without the child being present, and the use of a set of rating scales. Questions were developed to elicit information about the presence and strength of each process characteristic. These questions were then organized into an interview schedule. At the same time, a series of rating scales was constructed. These scales were used to obtain ratings for each of the process characteristics.

Once the interview schedule was developed, it was pilot-tested in a small sample of homes. On the basis of preliminary findings, the interview schedule was revised and retested. This time it was found that enough information could be obtained in an interview lasting about an hour and a half and consisting of 63 questions. The information that was obtained was used in rating the 13 process characteristics comprising the instrument measuring the environment for general intelligence, and the 21 process characteristics which comprised the instrument measuring the environment for academic achievement.

It was noted earlier that the focus of our investigation was what parents did rather than what parents were in terms of status, economic well-being, or some other demographic variable. The questions in the interview schedule were carefully designed to elicit information about what parents actually did insofar as general intelligence and academic achievement were concerned.

One process characteristic which was measured, for example, was the parents' educational aspirations for the child. In the course of the interview, the parents were asked how much education they wished their child to have, how much education they actually expected their child to receive, and the minimum amount of education they felt their child must have. In response to these questions, a number of parents indicated that they hoped their child would receive a college education. Later in the interview, the topic was brought up again, and parents who had indicated that they hoped their child would go on to college were queried as to what plans had been made to finance a college education. The answers to this latter question were most illuminating in differentiating a seemingly homogeneous group — parents who hoped that their child would have a college education. The responses ranged from the total absence of any plan for financing a college education to the most elaborate of plans. At the high end of the scale were several parents who had already established trust funds earmarked for their children's college expenses. It may be noted that parents who rated high on the planning for the attainment of educational goals were also the ones who indicated that education was frequently discussed in the home, and that the child was aware of the educational plans that had been made.

In the main study we selected a sample of 60 homes from a medium-sized Midwestern community comprised of urban, suburban, and rural areas. Preliminary information was collected on all 1,062 fifth grade students in the school system. On the basis of student descriptions of the father's occupation, social class ratings were obtained and a stratified random sample of homes was selected for study. The number of cases drawn from each social class grouping was proportional to the number of male adults in each grouping in the U.S. population according to Department of Labor data.

Contact was made with each of the selected homes and an interview with the mother was requested. In those cases in which the mother was unwilling or unable to participate, an alternate home was selected. The interviews were conducted in the home with the mother when the child was not present. To establish rapport, the mothers were assured at the outset of the interview that all information would be treated confidentially and that the school would receive only group results and not individual reports.

On the basis of the information collected during the course of the interviews, each home was rated on each of the process characteristics comprising the two instruments. For each environmental measure, all homes were rated on one process characteristic, then on another, and so forth. The process characteristics in each instrument were averaged to obtain the environmental variable ratings. These ratings were, in turn, combined to obtain the two total environmental ratings — one for the environment for academic achievement and one for the environment for intelligence. Thus, it was possible to conduct two separate investigations based on the same interview material.

At a later time, achievement test scores on the Metropolitan Achievement Test Battery and intelligence test scores on the Henmon-Nelson Test of Mental Ability were obtained from school files for the 60 fifth graders in the homes selected for study. The environmental data and the test data were then systematically related.

A full discussion of the results of our work is unfortunately beyond the bounds of this paper. However, the major findings can be presented and some implications for further research as well as educational practice can be tentatively drawn. The first major finding concerns the relationship between the total rating for the intellectual environment and measured general intelligence. The correlation between these two variables was +.69. This can be contrasted with the correlation between social status and measured general intelligence which has been found to lie between +.20 and +.40. Accepting the correlation of +.40 as the correct estimate of the relationship between social status and intelligence, it would seem that the newer approach to the measurement of the environment accounts for about three times as much of the variance in general intelligence as a measure of social status.

The second major finding in our work concerns the relationship between the total rating of the environment for the development of academic achievement and achievement test data. The correlation between the total environmental rating and the total achievement battery score was +.80. Again, this may be contrasted with the correlation between social status and academic achievement which has been found to be of the order of +.50. In terms of the proportion of variance accounted for, the newer approach to the measurement of the environment would seem to account for at least two and one half times as much of the variance in total academic achievement as a measure of social status. Taken together, the two correlations between the new environmental measures and intelligence and academic achievement reflect a new level of relationship between measures of the environment and measures of individual characteristics.

This new order of relationship can have a number of important implications for theory, research, and practice. The findings would suggest that the conception of a single physical environment as consisting of a number of sub-environments for the development and maintenance of specific characteristics is a powerful one indeed and, if fully developed, could greatly enhance our understanding of the interactive process between the individual and the environment.

The correlations between the over-all environmental rating and specific achievement sub-test scores ranged from +.55 for arithmetic computation to +.77 for word knowledge. The six sub-tests ranked in decreasing order of their correlation with the over-all environmental rating were: word knowledge; reading; language; arithmetic problem solving and concepts; word discrimination; and arithmetic computation. This ordering of the various sub-test score correlations with the total environmental rating might well have been anticipated. The influence of the environment on academic achievement could be expected to be greatest in the language area because much of the basic socialization of the child is usually accomplished through the medium of language and, in our culture, this responsibility rests pretty clearly in the home. Thus, the fact that the three highest correlations with the total environmental rating were word knowledge, reading, and language might well have been expected. Similarly, the one individual characteristic in which the responsibility for development rests most clearly with the school is arithmetic computation. Again, the fact that arithmetic computation was the lowest correlation with the over-all environmental rating might also have been anticipated. It should be noted that this ordering of the correlations between the achievement sub-test scores and the total environmental rating helps to establish the construct validity of the two instruments as much as do the correlations with over-all academic achievement and with general intelligence.

It was mentioned earlier that the correlation between the over-all rating for the environment for the development of academic achievement and the total achievement battery score was +.80. This represents a high level of relationship and can be considered to be quite sufficient for purposes of prediction. That is, a measure of what parents *do* in the home can be used to predict school achievement with a fairly high degree of accuracy. In actual practice, however, the results of intelligence tests are often used as the basis for making decisions about the placement and educational treatment of students. This fact prompted us to find out what happened when intelligence and environmental data were combined to predict academic achievement. This involved the computation of the multiple correlation between IQ and the over-all environmental rating for academic achievement with the total achievement battery score. The multiple correlation thus obtained was +.87. This represents an extremely high level of relationship. In fact, this is almost the upper limit for such a correlation when one takes into consideration the reliability of the various instruments. Statistically speaking, the amount of variance in academic achievement accounted for by intelligence test scores alone was 58 per cent (r = +.76). When a measure of the environment is added, the amount of variance in academic achievement accounted for rises to 76 per cent. Thus, the addition of a measure of the environment greatly enhances the estimation of academic achievement.

Since the studies described here represent the application of a psychometric approach to the measurement of environments, it is customary to report some technical data about the instruments. The reliabilities of the two instruments were estimated through an analysis of variance procedure suggested by Hoyt (8). The reliability estimates obtained through use of this procedure were .89 for the measure of the environment for the development of intelligence and .95 for the measure of the environment for the development of academic achievement.

Validity data about the two environmental measures inheres largely in the correlations with the measured individual characteristics — general intelligence and academic achievement. Additional evidence about the validity of the measure of the intellectual environment was obtained through a double cross-validation (it's important to remember that the hyphen comes between cross and validation). This procedure was proposed by Mosier (14) and is designed to furnish evidence about the stability of results based on a single testing. It involves the random division of the total sample into two sub-samples, computation of regression weights for each sub-sample separately, and the application of each set of regression weights to the other sub-sample for prediction of the criterion. In our study, the resulting correlations for the two sub-samples were +.663 and +.664. These are extremely close to the correlation of +.690 for the total sample and, as such, tend to fortify the original findings.

There are a number of additional findings but, unfortunately, these will have to wait until a fuller report of our work is prepared. However, with even the sketchy outline presented, some generalizations and implications can be drawn. The first generalization we would offer is that the conception of the environment upon which the present instruments were based does indeed seem to be a fruitful one for measuring and analyzing environments. The high correlations between the environmental measures and the specific characteristics selected for study furnish direct support for this generalization. However, we do not regard our list of environmental variables as by any means complete. Hopefully, further research will result in the identification of other environmental variables, which can be added to, or even replace, some of our variables. This will have to be determined at some future date. The point that we would emphasize is that research on the measurement of environments be directed towards the identification and measurement of ongoing environmental processes in relation to specific human characteristics. It would seem to us that environments for the development and maintenance of such characteristics as dependence, aggression, dogmatism, and others could be delineated and measured and systematically related to measures of the particular characteristic. Such research endeavors could greatly enhance our understanding of the development of many human characteristics.

A second generalization which emerges from our work concerns the relationship between the environment and intelligence and academic achievement. Much of our work gives quantitative support to the ideas of the effect of experience on intelligence and achievement put forth by Hebb (7), Hunt (9) and others. The level of relationship between the measures of the environment, and intelligence and achievement are significantly higher than those found in previous studies which used social status or some other general index as a measure of the nature and quality of the environment. Moreover, these new levels of relationship seem to be quite in line with the expectations of the theoreticians. It is again hoped that further research will be undertaken to test these findings.

If additional research does confirm these findings, then it would seem that there are a number of implications for educational practice. One obvious implication would involve the development of new curricula designed to help overcome identified environmental deficiencies among students. Useful information about the ingredients for programs of compensatory education could be obtained from careful examination of the environmental measures. The reason for this is that the environmental-process variables identified for study are, by and large, educationally malleable ones. That is, the variables represent environmental processes which can be manipulated in an educational situation. For example, one of the process characteristics in the measure of the intellectual environment was the opportunity provided for learning outside the home (excluding the school). If the homes which students come from are found to rate low on this characteristic, it would seem that the school might well undertake to furnish a number of experiences in this area. An extensive program of field trips, for example, could be undertaken to overcome this deficiency. Similarly, if the homes which students come from are found to be deficient in according rewards for academic achievement, the school could undertake to develop a greater system of rewards for such accomplishments and, perhaps, begin to work in concert with the home in insuring that academic achievement is acknowledged and rewarded. The point here is that the measure of environmental processes rather than status characteristics can furnish information which can be of direct service in the planning and conduct of educational programs.

It is rather fortunate that the framework for educational programs designed to overcome environmental deficiencies already exists at the elementary school level. The existence of self-contained classrooms and the recent development of an ungraded primary system could greatly facilitate compensatory educational programs. Even the traditional parent-teacher conference can be of great assistance in educational programs designed to overcome environmental deficiencies. As currently used, the parent-teacher conference serves mainly to inform parents of the child's progress in school. It seems possible that this function could be easily expanded to include the collection of certain standard information about the home environment. Thus, environmental data could be easily gathered and serve as a basis for planning programs of instruction.

There are a number of other practical uses for environmental data which could be suggested. However, the most tantalizing problems are probably the theoretical ones. For example, how stable is the environment or, more importantly, what are the conditions which make for stability in an environment? Also, how can one assess the variety of environments affecting the child as he grows older? What are the points when environmental intervention will have the greatest likelihood for success? These are thorny problems, and we do not have the answers. However, we do believe we have a methodology which can be of service in attacking these questions. Essentially, this methodology involves the application of our measurement procedures to the study of the environment and consists of four steps.

The first step entails the definition of an environment for the development of a specific characteristic. This construct is drawn largely from the relevant theory concerning the development of a particular characteristic. The second step involves the identification of the specific environmental conditions and processes which are likely to directly affect the development of the characteristic. Again, this step involves the extensive use of theory coupled with previous research concerning the development of the characteristic under study. The third step consists of the collection of evidence about the various environmental processes. In our work, we made use of an interview and a series of rating scales. However, there are a number of additional ways to gather such evidence. The work of Pace and Stern (15) suggests that the use of participant observers can furnish important information about an environment. Observational procedures can also be utilized. It would seem that many of the procedures that have been developed to measure individual characteristics can be adapted and applied to the measurement of environments. The fourth step in measuring an environment consists of treating environmental data through the use of psychometric procedures and systematically relating data about the individual to data about the environment. This is an extremely important step since it is only through this process that environmental information can give full meaning to data about individual characteristics.

In addition to the four steps listed above, the same technical considerations about reliability and validity which we are concerned with in the development of tests of individual characteristics must be dealt with. Again, our psychometric procedures can be utilized to gain answers to these technical questions.

It is not necessary to follow the four steps in the order listed above and study the environment in relation to one characteristic, then another, and so forth. A number of environments and individual characteristics can and, perhaps should, be studied simultaneously. It would seem that factor analysis, which has proven so powerful in the identification of the major dimensions on which individuals differ, may prove to be as powerful in defining the dimensions on which environments differ.

In closing, I invite you to give consideration to the prospects for environmental research. I would submit that as we make use of the conception that a single physical environment may consist of a number of sub-environments and seek to develop measurement procedures that capture the operant social and psychological conditions and processes in an environment, we should be able to greatly increase our understanding of the process of how individual characteristics are developed and maintained. Only when we have such understanding can we develop more adequate theories of behavior, and only then can we determine how particular characteristics can be maintained or altered.

REFERENCES

1. Anastasi, A., "Heredity, Environment and the Question 'How?' " *Psychological Review*, 1958, 65:196-207.
2. Bayley, N., "A New Look at the Curve of Intelligence," *Proceedings of the 1956 Invitational Conference on Testing Problems*, Princeton: Educational Testing Service, 1957, 11-25.
3. Bloom, B., *Stability and Change in Human Characteristics*, New York: John Wiley and Sons, Inc., 1964.
4. Coleman, W. and Cureton, E. E., "Intelligence and Achievement: 'The Jangle Fallacy' Again," *Educational and Psychological Measurement*, 1954, 14:347-51.
5. Douglass, J. W. B., *The Home and the School*, London: Macgibbon and Kee, 1964.
6. *Proceedings of the 1949 Invitational Conference on Testing Problems*, Princeton: Educational Testing Service, 1950.
7. Hebb, D. O., *The Organization of Behavior*, New York: John Wiley, 1949.
8. Hoyt, C., "Test Reliability Estimated by Analysis of Variance," *Psychometrika*, 1941, 6:153-60.
9. Hunt, J. McV., *Intelligence and Experience*, New York: Ronald Press, 1961.
10. Kahl, J. A., "Educational and Occupational Aspirations of 'Common Man' Boys," *Harvard Educational Review*, 1953, 23:186-203.
11. Kelley, T. L., *Interpretation of Educational Measurement*, New York: World Book Co., 1927.
12. Kirk, S. A., *Early Education of the Mentally Retarded*, Urbana, Ill.: University of Illinois Press, 1958.
13. Lee, E. S., "Negro Intelligence and Selective Migration: A Philadelphia Test of the Klineberg Hypothesis," *American Sociological Review*, 1951, 16:227-33.
14. Mosier, C. I., "Problems and Design of Cross Validation," *Educational and Psychological Measurement*, 1951, 10:5-11.
15. Pace, C. R., "Psychological Differences Between College Students," paper read before the Annual Meeting of the Western Personnel Institute, Pasadena, California, November 1959.

Social Class and Cultural Group Differences in Diverse Mental Abilities*

GORDON FIFER
Hunter College of the City University of New York

This paper is a necessarily abbreviated report of a study completed at Hunter College under the co-directorship of Professor Gerald Lesser, Dr. Donald Clark, and the author. In an attempt to cover all the most interesting aspects of the study, I may occasionally be too brief. However, the full report will be available within the next few weeks.

The problems of differentiated mental abilities and of their relationships to class and cultural group composition challenge us directly in the urban centers. The provision of suitable educational programs for all of the children in our schools presupposes our knowledge of what these children are like.

Despite the considerable amount of work by psychologists in the field of mental abilities to create so-called culture-free or culture-fair tests, little has been shown to yield consistent and valid results. The problem still remains as to how to evaluate the intellectual potential of children whose backgrounds necessarily handicap them seriously on the usual tests of mental ability.

This study focused on two major aspects of the problem: first, to devise tests which would be as free as possible of any direct class or cultural bias but also would still be acceptable measures of intellectual traits. (It should be noted that no attempt was made to devise any remarkably new or unique tests.) The second goal was to structure a testing situation which would enable each child to be evaluated under optimal conditions.

* This study was made possible by a Cooperative Research Grant from the Office of Education, United States Department of Health, Education and Welfare.

Hypotheses

The specific goal of this study was to examine the patterns among various mental abilities in first-grade children from different social class and cultural backgrounds. We accepted a definition of intelligence that postulates diverse mental abilities; hence, intelligent behavior can be manifested in a wide variety of forms. This provided the premise for the hypotheses tested in the study: that class and cultural influences differ not only in degree but in kind, with the consequence that different kinds of intellectual skills are fostered in various environments.

The specific hypotheses tested were that:

1. significant differences exist among groups of children from different social class and cultural backgrounds in each of the mental ability areas specified below;

2. significant differences exist among groups of children from different social class and cultural backgrounds in the pattern or configuration of scores from these diverse areas of mental ability;

3. significant interactions exist between the variables of social class and cultural background in determining the level of each mental ability and the nature of the patterns among them.

The Sampling

The four ethnic groups selected for the study were Chinese, Jews, Negroes and Puerto Ricans. The design provided for 20 boys and 20 girls at both the middle class and lower class levels in each of the ethnic groups. The total number of subjects was, therefore, 320. The age level selected was from 6 years to 7 years of age. The actual, ultimate sample ranged in age from 6 years 3 months, to 7 years 5 months. The sample was obtained with a few exceptions entirely in urban, congested New York City. In brief, the sampling procedure consisted of a rather detailed study of census-tract information on ethnic group distributions and income-level distributions in the city. Additional data were obtained from a variety of agencies including charitable institutions, government bureaus and market research companies. After careful plotting of the location of our population, the relevant public schools were spotted and the school authorities were contacted with regard to sampling children in the first grade. Once classes were identified, our examiners sampled as randomly as possible the children at these various locations. Children with physical handicaps or emotional disorders, or who became ill for more than a week were discarded. Also, the results of children whose testing was "requested" were omitted. Approximately 500 children were either completely or partially tested in order to try to maintain a fairly random selection within our stratified sample.

The Scales

The scales used in this study are rather extensive modifications of scales previously developed under a U.S. Office of Education grant conducted at Hunter College from 1957 to 1960. The four scales are: Verbal, Reasoning, Numerical and Space. The Verbal scale consists of a 30-item picture vocabulary test and two 15-item word definition tests yielding two-part scores; that is, a Picture Vocabulary score and a Word Vocabulary score plus a total score. Both of these sub-tests are scored on the basis of two points for each correct or completely acceptable response and one point for partially correct responses. The scoring keys were carefully developed by the group of psychologists participating in the study.

The Reasoning scale consists of three tests: Picture Analogies, Picture Arrangements and Jump Peg. The first two tests are relatively well-known types and no major deviation from the usual format is employed. The Jump Peg test, however, is quite novel in that the task required of the subject is unlike that of any of the children's games available on the market, although somewhat similar. The administration of this sub-test is planned to enable each child to start from the same point of knowledge about the task. It is felt that the test is a good measure of reasoning whenever perceptual variance can be neutralized.

The Numerical scale consists of five sub-tests: Enumeration — 6 items; Addition — 10 items; Subtraction — 10 items; Multiplication — 10 items; Division — 10 items. Knowledge of neither numbers nor operational symbols is required on any of these tests. All material is presented pictorially and with the exception of the Enumeration test, in which the subject is permitted to count with his finger touching the picture, all of the numerical operations have to be done in the child's head. It is recognized that this scale is quite unlike the usual number factor tests.

The Space scale consists of 4 sub-tests: Object Completion; Jig-Saws; Estimating Path; and Perspective. The Object Completion sub-test is more a perceptual-visualization measure than a pure space measure, but this type of test has shown to correlate fairly well with three-dimensional measures. The Jig Saws consist of puzzles quite unlike those available in children's games. They require the subject to perceive spatial relationships in order to complete them successfully, but require little motor proficiency and contain none of the usual reasoning cues found in popular jig-saw puzzles. The Estimating Path test was adapted from spatial tests used in the Air Force and probably is the weakest test in our battery owing to its susceptibility to chance. The task of the test is to estimate visually projected paths of airplanes. The Perspective sub-test consists of identifying the field of vision of several persons at various points in a scene. Although it is felt that the hypothetical construct underlying this test is primarily spatial, it is admitted that the method used to evaluate the construct possibly introduces verbal and reasoning variance at the age level studied.

Some general considerations about the construction of the tests should be noted:

1. No reading was required of the subject on any test.
2. All pictorial materials were clearly presented on large cards for easy viewing.
3. All persons were drawn as neutral as possible, with a careful avoidance of the precious, pretty pictures of the usual children's picture books.
4. No items required the naming of an object by the subject except on the Verbal scale. The additional exception to this was the Object Completion test; however, the subject could obtain full credit without naming an object if he could indicate his recognition in some other way.
5. An attempt was made to include only elements which appear commonly in all cultural and class groups in New York City.

Table I contains the Characteristics of the Tests and the Statistics Obtained by the Total Group. It can be noted that the reliabilities of the four total scales are Verbal — .93, Reasoning — .92, Numerical — .96, and Space — .85. Reasonably high reliability coefficients were obtained for all of the fourteen sub-tests with the exception of Estimating Path, Perspective, and the 6-item Enumeration sub-test. The intercorrelations of the four main scales are as follows: Verbal — Reasoning, .58; Verbal — Numerical, .54; Verbal — Space, .44; Reasoning — Numerical, .73; Reasoning — Space, .62; and Numerical — Space, .54. As you can see, these are higher than one would like

Table 1

Characteristics of the Tests and Statistics, Obtained by the Total Group (N = 320)

Scale	Subtest	No. of Items	Total Possible Score	Mean	Stand. Dev.	Rel. Coeff.
Verbal	Pict. Vocab.	30	60	43.4	8.8	.84
	Word Vocab.	30	60	31.0	11.0	.90
	Total	60	120	74.4	18.3	.93
Reasoning	Pict. Anal.	18	18	9.8	3.4	.80
	Pict. Arr.	16	16	7.1	4.5	.91
	Jump Peg	12	12	5.7	3.3	.89
	Total	46	46	22.6	8.6	.92
Numerical	Enumer.	6	6	5.0	1.2	.68
	Add	10	10	5.1	3.0	.88
	Subtr.	10	10	6.0	3.1	.82
	Mult.	10	10	3.5	3.3	.90
	Div.	10	10	3.8	2.9	.88
	Total	46	46	23.4	10.9	.96
Space	Obj. Compl.	16	32	17.0	6.8	.80
	Est. Path	12	12	8.1	1.9	.29
	Jig Saws	16	16	7.7	3.6	.89
	Persp.	10	10	5.2	2.2	.57
	Total	54	70	38.0	10.7	.85

them to be. Fortunately, the high reliabilities enable us to assume sufficient, unique variance to justify utilizing these scales to differentiate these four areas of mental ability.

Testing Procedures

Four psychologists helped with the revision of the scales and administered the tests. One was Jewish, one Negro, one Puerto Rican, and one Chinese. Each had been trained beyond the master's degree level and each had ample experience administering psychological tests, including the standard tests of intelligence. Each of the 320 children in our study was tested by a psychometrician who shared his cultural identity. While the length of the testing period varied and was determined primarily by the examiner's judgment regarding degree of rapport and fatigue, most commonly a child was seen for one session on each of three separate days for 30 to 45 minutes.

The test required neither reading nor writing ability and directions were kept extremely simple. The test could be administered in English or in the child's primary language, or in a combination of the two; children were permitted to respond in any combination of languages. Ample practice material was provided prior to each subtest and examiners were instructed to proceed with scored items only when certain that the child was familiar with the material and understood the task.

There were no formal steps of test administration left to the discretion of the examiner; each step was specified in detail to assure standard procedure. Each examiner saw every other examiner and himself administer the test to a child (during the pretest phase) on video tape. Each examiner also tested a child before a one-way vision screen while the other three took careful notes of the most minute deviation from standard procedure.

With four exceptions, each child was tested in his own school, in a room alone with the examiner. Examiners presented the tasks with a "game" orientation, but were instructed to change the orientation to a serious one if the child seemed more comfortable and productive when dealing with "work."

Class and Ethnic Group Designation

The designation of socio-economic class provided one of the major problems in the study. To make a very long story short, the ultimate technique employed was an adaptation of the Hollingshead and Redlich scale. The three factors of father's (that is, head of the household's) occupation, father's education, and dwelling unit were rated. A composite was then computed resulting in a 5-class scale. Classes I, II and III were designated middle class; Classes IV and V lower class.

A major limitation of our study is the differential level of the class designations of the subjects in the four ethnic groups and the differential separation of the class designations within the various groups. For example, approximately two-thirds of the Jewish middle-class subjects were in Class I on our scale. The Chinese and Negro middle-class subjects were predominantly Class II. The Puerto Rican middle-class subjects were in Classes II and III. The Jewish lower-class subjects were equally divided between Classes IV and V, whereas the other three lower-class groups were almost entirely in Class V. Hence, although we achieved good class separation—that is, at least one class position in our scale was skipped by each ethnic group — the separations did not occur at equivalent points on the scale. However, this situation is the reality in New York City, and we would have distorted our population by sampling to obtain equated groups.

The ethnic designation was somewhat simpler on paper. The parents of each subject had to be known to be of the particular ethnic group. The only problem was the actual determination of the facts for some of the subjects. Cases were discarded, however, if verification of the parentage could not be obtained.

Results

The analyses of the data are too extensive to be presented here in any detail, but certain highlights should prove of interest. The support or lack of support for our three hypotheses can be seen in Table 2 and the graphic representation of them in Figure 1.

Table 2

Means of Normalized Standard Scores on the Four Scales Listed by Ethnic and Social Class Group

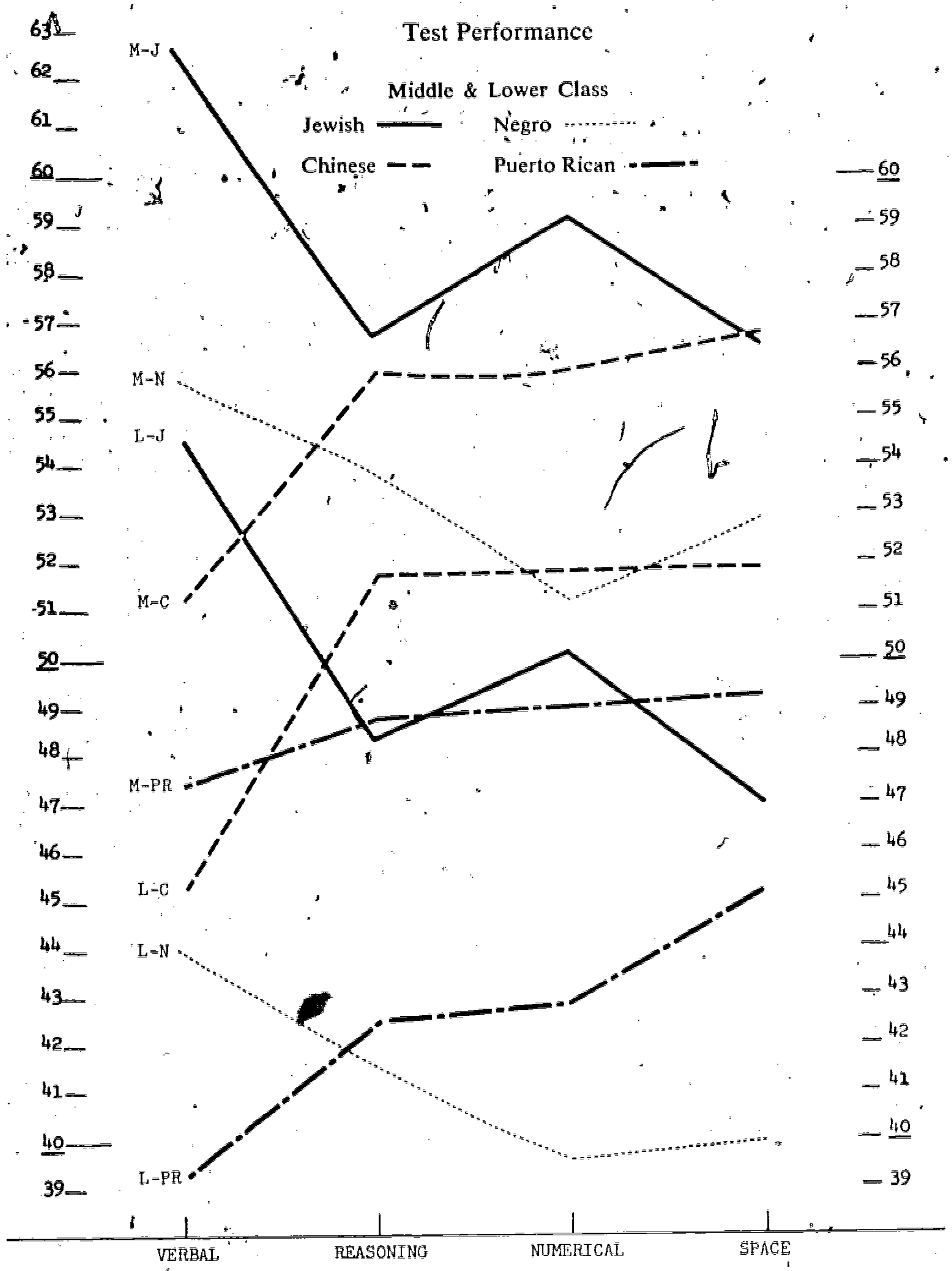
Group	Verbal	Reasoning	Numerical	Space
Jewish Middle Class	62.6	56.7	59.2	56.5
Lower Class	54.7	48.5	50.2	47.0
Total	58.7	52.6	54.7	51.8
Chinese Middle Class	51.2	56.0	56.0	56.8
Lower Class	45.2	51.8	51.8	52.0
Total	48.2	53.9	53.9	54.4
Negro Middle Class	55.8	53.9	51.2	53.0
Lower Class	44.0	41.5	39.6	40.1
Total	49.9	47.7	45.4	46.5
Puerto Rican Middle Class	47.4	48.7	49.2	49.4
Lower Class	39.3	42.5	43.0	45.3
Total	43.3	45.6	46.1	47.4
Total Middle Class	54.2	53.8	53.9	53.9
Lower Class	45.8	46.1	46.1	46.1

Our first hypothesis, that significant differences exist among groups of children from different social class and cultural backgrounds in each of the four mental ability areas, was supported.

Our second hypothesis, that significant differences exist among groups of children from different social class and cultural backgrounds in pattern or configuration of scores from these diverse areas of mental ability, was partially supported.

Our third hypothesis, that significant interactions exist between the variables of social class and cultural background in determining the

FIGURE 1


Test Performance
Middle & Lower Class
Jewish
Negro
Chinese
Puerto Rican
M-J
M-N
L-J
M-C
M-PR
L-C
L-N
L-PR
VERBAL
REASONING
NUMERICAL
SPACE

level of each mental ability and the nature of the patterns among them, was partially supported.

One can see from the graphic representation of the means that social class seems to influence *level* of functioning, and cultural identity seems to influence the *pattern*, or relative high and low ability areas. We have no evidence that these two factors interact in any way that creates significant differences in functioning.

One or two points should be emphasized. Note that the lowest mean group score is earned by the lower-class Puerto Ricans on the verbal tasks. Yet this same group does quite well, compared with the lower-class Negroes, on spatial tasks.

There is a clear and almost uniform difference within each cultural group that is imposed by social class and runs across the four mental ability areas. Yet, notice the difference in degree of separation between the two social classes of Chinese and the two social classes of Negroes. To be sure, this may be partly due to factors built into our study but it still suggests the strong possibility that social-class difference will more strongly affect one's identity and intellectual performance if he is a Negro than if he is Chinese.

It may be wise to point out that these data refer to differences in the performance of *groups* and not the performance of *individuals*. They do not indicate how an individual will perform, but if he belongs to one of these eight groups, they suggest how he is *likely* to perform. His individual deviation, be it high or low, on one score or another, thereby becomes more useful information in understanding his individual abilities.

One approach to utilizing test data such as these for identifying particular patterns for individuals is a classification technique suggested by Tatsuoka. Table 3 illustrates the application of the method to our data. Note that the middle-class Jewish children and the lower-class Negro children were most accurately classified and the lower class Puerto Rican children slightly less so. Note, however, that in addition to the 20 Puerto Rican lower-class children typical of their class and ethnic group pattern, an additional 15 had patterns like other lower-class patterns (that is, Chinese and Negro). It is apparent that the middle-class Puerto Rican children were the most heterogeneous of any of the eight groups.

Table 3

Classification Analysis

This compares the pattern of scores for each individual subject with the pattern profiles of his group and other groups. It yields data concerning the degree to which a subject's profile resembles the profile of his or the other groups. Reference: Tatsuoka, M. M. *Joint Probability of Membership and Success in a Group*, Harvard Graduate School of Education Report 1957.

Group	Group Patterns							
N = 40, each Group	M Ch	L Ch	M J	L J	M N	L N	M PR	L PR
Middle Chinese	13*	10	6	1	5	1	2	2
Lower Chinese	6	14	2	4	3	1	1	9
Middle Jewish	4	0	32	4	0	0	0	0
Lower Jewish	0	1	9	18	7	4	0	1
Middle Negro	5	1	11	10	11	0	0	2
Lower Negro	1	3	0	3	0	28	0	5
Middle Puerto Rican	6	6	3	6	4	0	3	12
Lower Puerto Rican	0	7	1	1	0	8	3	20

*Figures to be read across as follows: The scores of 13 middle-class Chinese subjects fit the middle-class Chinese pattern and level on the four mental ability scales; 10 middle-class Chinese look more like lower-class Chinese; 6 look more like middle-class Jews, 1 more like a lower-class Jew, etc.

Conclusions

It is apparent that our study has raised more questions than it has produced neat answers. Our tests need revision and refinement. Additional analyses currently being completed of the factorial composition of the battery and of the unique non-chance variance of each sub-test will be useful in further interpretation of our data. It is believed that we have strong evidence of differential patterns of mental abilities among four New York City ethnic groups. The data reveal sharp test-performance differences between middle- and lower-class groups regardless of ethnic group and that the differences differ among the ethnic groups. Replications of the study on the same and additional ethnic groups using psychometricians of like and unlike background would be valuable. There is some evidence of differences in patterns among the groups that may be of considerable value in educational planning. The value of the study will be finally realized in the follow-up studies we hope it will stimulate.

REFERENCES

1. Anastasi, Anne, "Cultural Differences." In C. W. Harris (Ed.), *Encyclopedia of Educational Research*. New York: Macmillan, 1960a. Pp. 350-358.

2. Anastasi, Anne, and Cordova, F. A., "Some Effects of Bilingualism upon the Intelligence Test Performance of Puerto Rican Children in New York City." *Journal of Educational Psychology*, 1953, *44*, 1-19.
3. Anastasi, Anne, and D'Angelo, R. V., "A Comparison of Negro and White Preschool Children in Language Development and Goodenough Draw-a-Man I. Q." *Journal of Genetical Psychology*, 1952, *81*, 147-165.
4. Anastasi, Anne, and deJesus, C., "Language Development and Goodenough Draw-a-Man I. Q. of Puerto Rican Preschool Children in New York City." *Journal of Abnormal Social Psychology*, 1953, *48*, 357-366.
5. Block, J., Levine, L. and McNemar, Q. "Testing for the Existence of Psychometric Patterns." *Journal of Abnormal Social Psychology*, 1951, *46*, 356-359.
6. Canady, H. G. "The Effect of 'Rapport' on the IQ: A New Approach to the Problem of Racial Psychology." *Journal of Negro Education*, 1936, *42*, 388-389.
7. Carson, A. S., and Rabin, A. I., "Verbal Comprehension and Communication in Negro and White Children." *Journal of Educational Psychology*, 1960, *51*, 47-51.
8. Davis, F. B., Lesser, G. S., French, Elizabeth G., *et al.*, "Identification and Classroom Behavior of Gifted Elementary-School Children." *Cooperative Research Monographs*, 1960, No. 2, 19032. In "The Gifted Student," OE-35016, Monograph No. 2, U. S. Office of Education, Department of Health, Education, and Welfare.
9. Hollingshead, A. B., and Redlich, F. C., *Social Class and Mental Illness*. New York: Wiley, 1958.
10. McNemar, Q., *Psychological Statistics*. New York: John Wiley 1962.
11. Passow, A. H. (Ed.), *Education in Depressed Areas*. New York: Teachers College Publications, 1963.
12. Tatsuoka, M. M., *Joint Probability of Membership and Success in a Group*. Cambridge: Harvard Graduate School of Education Report 1957.

Participants

1964 Invitational Conference on Testing Problems

ABBOTT, MURIEL M., Teachers College, Columbia University
ABRAHAM, ANSLEY A., Florida Agricultural and Mechanical University
ADKINS, DOROTHY C., University of North Carolina
AGINS, STUART A., Mattlin Junior High School, Plainview, Long Island
AHRENS, DOLORES F., Educational Testing Service
ALLEN, MARGARET E., Portland (Maine) Public Schools
ALMAN, JOHN E., Boston University
ALT, PAULINE M., Central Connecticut State College
ALTER, MILLICENT, The Diebold Group, Incorporated
ANASTASI, ANNE, Fordham University
ANDERSON, DAVID J., East Stroudsburg State College
ANDERSON, GORDON V., The University of Texas
ANDERSON, PAULINE K., New York State Employment Service
ANDERSON, SCARVIA B., Educational Testing Service
ANDREE, ROBERT G., Rich Township High Schools, Park Forest, Illinois
ANGELL, GEORGE W., JR., Massachusetts Institute of Technology
ANGOFF, WILLIAM H., Educational Testing Service
ARCHER, N. SIDNEY, Department of Public Instruction, Harrisburg, Pennsylvania
ARCHIE, WILLIAM C., North Carolina Board of Higher Education
ARMANTROUT, L. L., American Telephone and Telegraph Company, New York City
ARMSTRONG, ROSS, University of Kansas
ARONOW, MIRIAM S., New York City Board of Education
ARONS, SIMONE, Brooklyn, New York
ASHCRAFT, KENNETH B., Colorado State Department of Education
ASTMAN, JOSEPH, Modern Language Association, New York City
ATKINS, WILLIAM H., The Consultation Center, Elizabeth, New Jersey
ATWATER, STEPHEN K., Knox College
AUSTIN, NEALE W., Educational Testing Service
BADAL, ALEN W., Oakland (California) Public Schools
BALL, ALBERT T., The Pingry School, Elizabeth, New Jersey
BALL, SAMUEL, Teachers College, Columbia University
BANDY, SUZANNE J., Aspen Hill School, Rockville, Maryland
BANNON, CHARLES J., Department of Education, Waterbury, Connecticut
BANNON, MARION K., Crosby High School, Waterbury, Connecticut
BARCLAY, GORDON L., Rockland State Hospital, Orangeburg, New York
BARDACK, HERBERT, New York State Department of Social Welfare
BARDON, JACK I., Rutgers, The State University
BARGMANN, ROLF, International Business Machines Corporation

BATTAGLIO, JOSEPH J., Ipswich (Massachusetts) Public Schools
BEATON, ALBERT E., Educational Testing Service
BECKMAN, GEORGE, Rutgers, The State University
BEEZER, ROBERT H., United States Office of Education
BELLOWS, L., The City College of New York
BENNETT, GEORGE K., The Psychological Corporation
BENNETT, MARJORIE G., Bronxville, New York
BENNETT, VIRGINIA, Rutgers, The State University
BENNINGTON, NEVILLE L., New York State Department of Education
BENSON, ARTHUR L., Educational Testing Service
BENTLER, PETER M., Educational Testing Service
BERDIE, RALPH F., University of Minnesota
BERGER, BERNARD, New York City Department of Personnel
BERGESEN, B. E., JR., Personnel Press, Inc., Princeton, New Jersey
BERGSTEIN, HARRY B., Huntington (New York) Public Schools
BERKIN, SHERMAN P., The City College of New York
BERLIN, MARTIN, New Milford (New Jersey) Public Schools
BERNE, ELLIS J., U. S. Department of Health, Education and Welfare
BERRIEN, F. K., Rutgers, The State University
BIGGERS, JULIAN, Texas Education Agency
BILLERA, LOUIS J., Educational Testing Service
BISBEY, GERALD D., State College of Iowa
BIXLER, HAROLD H., Western Carolina College
BLANDAU, HOWARD, The King's College
BLEEKE, DONALD E., College Reading-Study and Counseling Center, Westfield, New Jersey
BLIGH, HAROLD F., Harcourt, Brace and World, Inc.
BLOOM, BENJAMIN S., University of Chicago
BLUM, STUART H., Hofstra University
BOLAND, RUTH F., Cambridge (Massachusetts) School Department
BOLLENBACHER, JOAN, Cincinnati (Ohio) Public Schools
BOLLES, THERESA E., Paterson State College
BONDARUK, JOHN, National Security Agency
BORNSTEIN, HARRY, Gallaudet College
BOUCHARD, ROBERT, New York State Department of Civil Service
BOWES, EDWARD W., University of California
BOYER, E. GIL, Harvard University
BRACA, SUSAN E., Oceanside Senior High School, Forest Hills, New York
BRADLY, R. L., Auckland, New Zealand
BRANDT, HYMAN, Programed Instruction for Industry, New York City
BRANSFORD, THOMAS L., New York State Civil Service Department
BREAM, LOIS GOULD, Cheltenham High School, Wyncote, Pennsylvania
BRETNALL, WILLIAM B., Educational Testing Service

BRIDGES, CLAUDE F., United States Army Enlisted Evaluation Center
BRIGNALL, ED, New York University
BRISTOW, WILLIAM H., New York City Board of Education
BRODSKY, DAVID J., Educational Testing Service
BRODY, ERNESS BRIGHT, Rutgers, The State University
BROOKS, RICHARD B., Longwood College
BROWN, F. MARTIN, Fountain Valley School, Colorado Springs, Colorado
BROWN, FRED S., Great Neck (New York) Public Schools
BROWN, JAMES A., New York State Education Department
BROWN, MARI J. K., Columbia University
BROWNRIGG, HELEN R., Daniel Butler School, Belmont, Massachusetts
BRUSAK, GEORGE F., Bronson (Michigan) High School
BRYAN, GLENN L., Office of Naval Research, Washington, D.C.
BRYAN, J. NED, United States Office of Education
BRYAN, MIRIAM M., Educational Testing Service
BUCKINGHAM, GUY E., Allegheny College
BUDDEKE, SISTER RITA, Trinity College
BUNDERSON, C. VICTOR, Educational Testing Service
BUNNER, ALAN M., Culver Military Academy
BURDOCK, EUGENE I., Columbia University
BURGESS, THOMAS C., Portland State College
BURKE, PAUL J., Bell Telephone Laboratories, Inc.
BURNETTE, RICHARD R., Florida Southern College
BURNHAM, PAUL S., Yale University
BURNS, KEITH, Vashon Island High School, Burton, Washington
BURNS, RICHARD L., Educational Testing Service
BUROS, OSCAR K., Rutgers, The State University
BURR, WILLIAM L., Harcourt, Brace and World, Inc.
BURRILL, LOIS E., Educational Testing Service
BURROWES, ANNE L., Educational Testing Service
BYRNE, RICHARD HILL, University of Maryland
CAHEN, LEONARD S., Stanford University
CAMPBELL, DAVID P., University of Minnesota
CAMPBELL, DONALD W., Board of Education, Newark, New Jersey
CAMPBELL, JOEL T., Educational Testing Service
CAMPBELL, YOLANDA A., Trenton, New Jersey
CANTRELL, CAROL C., Educational Testing Service
CANTRELL, CYRUS D., Princeton University
CANTWELL, ZITA M., Brooklyn College
CARLSON, J. SPENCER, University of Oregon
CARLUCCI, COSIMO, Sears, Roebuck and Company
CARNEGIE, MARY ELIZABETH, Nursing Outlook, New York City
CARNEY, MELVIN, Erie (Pennsylvania) School District

CARR, JAMES V., New Hampshire State Department of Education
CARROLL, JOHN B., Harvard University
CARSON, JOHN, Pennsbury High School, Yardley, Pennsylvania
CAVALLARO, RICHARD F., Rochester Board of Education
CHAMPAGNE, JOSEPH E., South Carolina Committee for Technical Education
CHAPMAN, C. O., New York University
CHAPMAN, DWIGHT W., Vassar College
CHAPMAN, GLORIA, New York University
CHAPPELL, BARTLETT E. S., New York Military Academy
CHAUNCEY, HENRY, Educational Testing Service
CHEVRIER, JEAN-MARC, Institute of Psychological Research, Montreal, Canada
CHOPPIN, BRUCE H., Chicago, Illinois
CIERI, VINCENT P., United States Army Signal Center and School
CLARK, PHILIP I., California Test Bureau, Scarsdale, New York
CLEARY, J. ROBERT, Educational Testing Service
CLEARY, T. ANNE, Educational Testing Service
CLEMANS, WILLIAM V., Science Research Associates
CLENDENEN, DOROTHY M., The Psychological Corporation
CLINKER, B. K., Mt. Healthy High School, Cincinnati, Ohio
CHASE, CLINTON I., Indiana University
COFFMAN, WILLIAM E., Educational Testing Service
COHEN, SANDRA, Educational Testing Service
COHEN, SYLVIA, Psychological Services, Hamden, Connecticut
COLBURN, MARK, Office of Assistant Secretary of Defense
COLLEEN, SISTER MARY, S.S.N.D., Archdiocese of New Orleans
CONKLIN, NANCY W., University of Rochester
CONNOLLY, JOHN A., Educational Testing Service
CONRAD, HERBERT S., United States Office of Education
COOLEY, WILLIAM W., University of Pittsburgh
COOPERMAN, IRENE G., Veterans Administration
COPE, WILLIAM E., JR., United Negro College Fund
COPELAND, HERMAN A., Pennsylvania State Civil Service Commission
CORCORON, MARY, University of Minnesota
CORY, BERTHA HARPER, United States Army Personnel Research Office
COVINGTON, VIRGINIA, Gallaudet College
COX, HENRY M., The University of Nebraska
COX, MRS. HENRY M., Lincoln (Nebraska) Public Schools
CRAMER, AILEEN, Educational Testing Service
CRAWFORD, WILLIAM R., University of Illinois
CREAGER, JOHN A., National Academy of Sciences
CROSS, ORRIN H., West Virginia University

CROSSON, WILHELMINA M., Palmer Memorial Institute
CRUICKSHANK, REBECCA, Stanford University
CURETON, EDWARD E., University of Tennessee
CURETON, LOUISE W., University of Tennessee
CURRIVAN, GENE, New York Times
CURRY, ROBERT P., Cincinnati (Ohio) Public Schools
CYNAMON, MANUEL, Brooklyn College
DAILEY, JOHN T., George Washington University
DALLAS, JOE, JR., New York University
DAMARIN, FRED L., Educational Testing Service
DAVIDSON, HELEN H., The City College of New York
DAVIS, JUNIUS A., Educational Testing Service
DEABLER, L. T., International Business Machines Corporation
DELMAN, MARTIN J., Forest Hills (New York) High School
DE MILAN, SISTER JEAN, Rhode Island College
DERMEN, DIRAN, Educational Testing Service
DERSHIMER, RICHARD A., American Educational Research Association
DESMOND, HELEN E., Summit (New Jersey) High School
DETCHEN, LILLY, Chatham College
DEUTSCH, MARTIN, Institute for Developmental Studies
DE WITT, CHARLES J., Longwood College
DEXTER, DONALD W., Philadelphia Personnel Department
DEXTER, MRS. DONALD, Philadelphia, Pennsylvania
DIAMOND, ESTHER E., Science Research Associates
DIAMOND, LORRAINE K., Teachers College, Columbia University
DICK, WALTER, The Pennsylvania State University
DIEDERICH, PAUL B., Educational Testing Service
DIEHL, MARY JANE, Educational Testing Service
DILLENBECK, DOUGLAS D., College Entrance Examination Board
DIMENGO, CARL, Akron (Ohio) Public Schools
DODD, WILLIAM E., International Business Machines Corporation
DONLON, THOMAS F., Educational Testing Service
DOPPELT, JEROME E., The Psychological Corporation
DOWD, RICHARD A., Department of Education, Montpelier, Vermont
DRAGOSITZ, ANNA, Educational Testing Service
DRIJE, CARLA, Bank Street College of Education
DRUCKER, ARTHUR J., United States Army Personnel Research Office
DUBNICK, LESTER, Huntington (New York) Public Schools
DU BOIS, PHILIP H., Washington University
DUDEK, E. E., United States Navy Personnel Research Activity
DUGAN, ROBERT D., International Business Machines Corporation
DUGGAN, JOHN M., College Entrance Examination Board
DUNN, FRANCES E., Brown University

DUROST, WALTER N., University of New Hampshire
DUTTON, GENE, Rhode Island College
DWYER, MARY A., New York University
DYER, HENRY S., Educational Testing Service
EBEL, ROBERT L., Michigan State University
EDWARDS, DOROTHY S., American Institute for Research
EDWARDS, WINIFRED E., Irvington (New Jersey) High School
EIGEN, LEWIS, U. S. Department of Health, Education and Welfare
ELAINE, SISTER M., S.S.N.D., Our Lady Queen of Heaven School (Louisiana)
ELLIS, ALLAN B., Harvard University
ENGELHART, MAX D., Chicago City Junior College
EPSTEIN, BERTRAM, The City College of New York
ESSER, BARBARA, Educational Testing Service
EVANS, RICHARD, Highland Park (New Jersey) High School
FAIR, J. C. STEACY, Royal Canadian Air Force Central Examination Board, Ontario, Canada
FAUSTEIN, SHIRLEY, New York University
FAN, CHUNG T., International Business Machines Corporation
FARNEN, RUSSELL F., JR., Educational Testing Service
FATERSON, HANNA F., Downstate Medical Center, Brooklyn, New York
FAY, PAUL J., New York State Department Civil Service
FEINSTEIN, ROBERT, Institute for Career Guidance Incorporated, New York City
FENDRICK, PAUL, Western Electric Company, Millburn, New Jersey
FENSTERMACHER, GUY M., Educational Testing Service
FERGUSON, JOHN L., University of Missouri
FERGUSON, LEONARD W., The Journal Press, Provincetown, Massachusetts
FERRIS, RICHARD J., Dean Junior College
FIFER, GORDON, Hunter College
FINDLAY, PATRICIA, Educational Testing Service
FINEGAN, OWEN T., Gannon College
FINK, DAVID R., JR., University of Maine
FINN, JEREMY D., University of Chicago
FISCHER, FREDERIC E., State University of New York at Oswego
FISHMAN, JOSHUA A., Yeshiva University
FISKE, DONALD W., University of Chicago
FITZGERALD, JOHN F. M., Hillcrest School, Brookline, Massachusetts
FITZGIBBON, T. J., Harcourt, Brace and World, Inc.
FLANAGAN, JOHN C., American Institute for Research
FLAUGHER, RONALD L., Educational Testing Service
FLEISCH, SYLVIA, Boston University
FLEMMING, EDWIN G., New York, New York
FLESCH, MARY H., Educational Testing Service

FLOWERS, JOHN F., Ontario College of Education
FLOYD, WILLIAM, West Lafayette (Indiana) School System
FOLEY, JOSEPH J., Worcester State College
FORBES, DEAN W., Baltimore (Maryland) Public Schools
FORD, SHIRLEY, New York University
FORLANO, ANNA MARIE, New York City Public Schools
FORLANO, GEORGE, New York City Board of Education
FORMICA, LOUIS A., Ridgefield (Connecticut) Public Schools
FORSTALL, LIONEL J., New York University
FORSTER, C. R., JOIN, City of New York
FRANKEL, MARVIN, Educational Testing Service
FREDERIKSEN, JOHN, Educational Testing Service
FREE, JOHN E., University of Pennsylvania
FREMER, JOHN J., Staten Island, New York
FRENCH, BENJAMIN J., New York State Department of Civil Service
FRENCH, J. L., The Pennsylvania State University
FRENCH, JOHN W., New College
FREUND, LUCY D., Educational Testing Service
FRICKE, BENNO G., University of Michigan
FRIEDMAN, JERRY J., Kapiat Junior High School, Spring Valley, New York
FRIEDMAN, SIDNEY, Bureau of Naval Personnel, Washington, D. C.
FROE, OTIS D., Morgan State College, Maryland
FRY, EDWARD, Rutgers, The State University
GAINES, LYNN, Educational Testing Service
GALLAGHER, HENRIETTA L., Educational Testing Service
GAMES, PAUL, Ohio University
GANNON, FRED, Educational Testing Service
GARBER, HERBERT, University of Connecticut
GAROD, FLORENCE, City College of New York
GASIOROWSKI, MARY, City University of New York
GEE, HELEN HOFER, National Institute of Child Health and Human Development
GELBERG, ANITA, Bureau of Child Guidance, New York City
GIBLETTE, JOHN F., University of Maryland
GIBSON, W. O., Queens College of the City University of New York
GILBERT, JEAN P., New York University
GILDAR, SANDRA, New Jersey State Department of Education
GIRAFFE, FRANK, Chester County (Pennsylvania) Public Schools
GLASER, ROBERT, University of Pittsburgh
GLASIER, CHARLES A., New York State Education Department
GLASS, DAVID C., Russell Sage Foundation
GLESER, GOLDINE C., University of Cincinnati
GLESER, LEON J., Columbia University

GLICKMAN, ALBERT S., United States Department of Agriculture
GOBETZ, WALLACE, New York University
GOGGIN, VIRGINIA, Roanoke (Virginia) Public Schools
GOLDBERG, MIRIAM L., Teachers College, Columbia University
GOLDBERGER, FRANCES R., New York City Department of Personnel
GOLDMAN, LEO, City University of New York
GOLDMAN, NORMAN, New Jersey Education Association
GOLDSTEIN, LEO S., New York Medical College
GORDON, LEONARD V., United States Army Personnel Research Office
GORHAM, WILLIAM A., United States Civil Service Commission
GOSLIN, DAVID, Russell Sage Foundation
GOTKIN, LASSAR G., New York Medical College
GRAFF, FRANKLYN A., Westport (Connecticut) Public Schools
GRANT, DONALD L., American Telephone and Telegraph Company
GRAY, FRANK M., Harcourt, Brace and World, Inc.
GRAY, LYLE BLAINE, Diagnostic and Remedial Center, Baltimore, Maryland
GREELEY, DAVID MC L., Educational Council for Foreign Medical Graduates
GREEN, DOROTHY, United States Civil Service Commission
GREENBERG, BERNARD L., Gallaudet College
GREENBERG, SALLY, Gallaudet College
GREENE, FRANK, Rutgers, The State University
GREENE, JAY E., Board of Examiners, Brooklyn, New York
GREENHILL, L. P., Pennsylvania State University
GREER, HARRY H., JR., Scarsdale, New York
GROESBECK, BYRON L., University of Michigan
GROSE, ROBERT F., Amherst College
GROSSMAN, LEN, Hunter College
GROVE, LEONARD C., Franklin and Marshall College
GRUMMON, DONALD L., Michigan State University
GRUPPER, RUTH T., New City, New York
GUERRIERO, MICHAEL A., The City College of New York
GUGGENHEIM, CORINNE L., Haverstraw-Stony Point (New York) Schools
GUGGENHEIM, FRED, New York City Board of Education
GULLIKSEN, HAROLD, Educational Testing Service
GUNDERSON, DORIS V., United States Office of Education
GUTTMAN, ISAIAH, Science Research Associates
GUTTMAN, LOUIS, The Hebrew University and the Israel Institute of Applied Social Research, Jerusalem
HAAG, CARL H., Educational Testing Service
HAAGEN, C. HESS, Wesleyan University
HAGEN, ELIZABETH, Teachers College, Columbia University
HAGGERTY, HELEN, United States Army Personnel Research Office

HAGIN, ROSA, New York University
HAGMAN, ELMER R., Greenwich (Connecticut) Public Schools
HALL, ALFRED E., Carnegie Institute of Technology
HALL, NORMAN, Lampasas (Texas) School System
HALL, ROBERT O., Manter Hall School, Cambridge, Massachusetts
HALPERN, GERALD, Educational Testing Service
HAMBURGER, MARTIN, New York University
HAMILL, CHARLES O., Department of Education, Hato Rey, Puerto Rico
HANRIEDER, BARBARA D., Educational Testing Service
HARVIN, BRISTOW, JR., West End School, Roanoke, Virginia
HAROOTUNIAN, BERJ, University of Delaware
HARRIS, BEATRICE, Brooklyn College
HARRIS, CHESTER W., The University of Wisconsin
HARRISON, FOREST, University of Chicago
HARTMAN, THOMAS F., International Business Machines Corporation
HARTSHORNE, NATHANIEL, Educational Testing Service
HARTZOG, ERNEST E., New York University
HARVEY, JAMES G., Harvard University
HARVEY, PHILIP R., Educational Testing Service
HASTINGS, J. THOMAS, University of Illinois
HAUSDORFF, HENRY, University of Pittsburgh
HAUSMAN, HOWARD J., National Science Foundation
HAVEN, ELIZABETH, Educational Testing Service
HAWES, GENE, Columbia University Press
HAYWARD, PRISCILLA, Educational Testing Service
HEATH, C. NEWTON, Lincoln-Sudbury (Massachusetts) Regional High School
HEATH, ROBERT W., Educational Testing Service
HEENAN, DAVID K., Michigan State University
HEFLIN, VIRGINIA B., Bureau of Educational Research, New York City
HEISER, RUTH BISHOP, Glendale, Ohio
HELMICK, JOHN S., Educational Testing Service
HEMPHILL, JOHN K., Educational Testing Service
HENRY, EDWIN R., Standard Oil Company
HENRY, SALLYANN, The Psychological Corporation
HERMAN, DAVID O., The Psychological Corporation
HESLIN, PHYLLIS, National League for Nursing
HIGGINS, J. C., Fordham University
HILTON, THOMAS L., Educational Testing Service
HINDS, JEAN, Howard University
HO, WAI-CHING, Educational Research Council of Greater Cleveland
HOCK, LOUISE E., New York University
HOCKING, ELTON, Purdue University

HOFFMAN, E. LEE, Tulane University
HOLLENBECK, GEORGE P., International Business Machines Corporation
HOLLIS, ESTHER R., The Psychological Corporation
HOLLISTER, JOHN S., Educational Testing Service
HORN, DOROTHY M., Teachers College, Columbia University
HOROWITZ, GEORGE F., Columbia University
HOROWITZ, MILTON W., Queens College of the City University of New York
HORSMAN, DONALD, McGraw Hill Book Company
HORTON, CLARK W., Dartmouth College
HOUIS, JANE, Educational Testing Service
HOWARD, ALLEN H., University of Illinois
HOWELL, JOHN J., New Jersey Educational Research
HUMPHRY, BETTY, Educational Testing Service
HUNT, G. HALSEY, Educational Council for Foreign Medical Graduates
HUTCHINS, EDWIN B., Association of American Medical Colleges
HUTCHINSON, MARY G., Wilby High School, Waterbury, Connecticut
HUYSER, ROBERT J., Educational Testing Service
HUYSER, SARAH L., Educational Testing Service
HYPPS, IRENE C., District of Columbia Public Schools
IFFERT, ROBERT E., United States Office of Education
IMPELLITTERI, JOSEPH T., City University of New York
INDOW, TAROW, Harvard University
IRBY, ALICE J., Educational Testing Service
IRVINE, PAUL JR., Board of Cooperative Educational Services, Yorktown Heights, New York
IWAMOTO, DAVID, National Education Association
JACKSON, DOUGLAS N., University of Western Ontario
JACKSON, ROBERT W. B., University of Toronto
JACOBS, JANET, Northfield (Massachusetts) School
JANEBA, HUGO B., Rutherford (New Jersey) Senior High School
JANOWITZ, CECELIA B., Public School 106, Bronx, New York
JASPEN, NATHAN, New York University
JINKS, ELSIE H., Grosse Pointe (Michigan) Public Schools
JOHNSON, A. PEMBERTON, Newark College of Engineering
JOHNSON, LEWIS W., Personnel Department, Philadelphia, Pennsylvania
JOHNSON, RICHARD T., Rutgers, The State University
JOHNSON, VICTOR B., Florida State Department of Education
JONES, KENNETH J., Harvard University
JONES, MEVERIL, Roosevelt Junior High School, Westfield, New Jersey
JONES, OCTAVIA M., New York University
JONES, ROBERT A., University of Southern California
JONES, THELMA J., New York University

JORDAN, ARTHUR M., University of North Carolina
JORESKOG, KARL G., Educational Testing Service
JOSEPHINA, SISTER, Boston College
JOYCE, LYNN, Fordham University
JURGENSEN, CLIFFORD E., Minneapolis Gas Company
JUSTMAN, JOSEPH, Board of Education, Brooklyn, New York
KABACK, GOLDIE R., The City College of New York
KAISER, HENRY F., University of Wisconsin
KANE, SARA H., Public School 57, Bronx, New York
KAPLAN, KYRA M., National League for Nursing
KARAS, SHAWKY F., Educational Testing Service
KARSTETTER, ALLAN, Pennsylvania State Department of Public Instruction
KASEL, FRANCIS X., Educational Testing Service
KASTRINOS, WILLIAM, Educational Testing Service
KATZ, MARTIN R., Educational Testing Service
KAUFMANN, JAMES D., Warwick Valley High School, Warwick, New York
KAYE, MILDRED, Brooklyn College
KEITH, JAMES A., Rutgers, The State University
KELLER, FRANKLIN J., Governor's School of North Carolina
KELLER, ROBERT J., University of Minnesota
KELLEY, H. PAUL, College Entrance Examination Board
KELLEY, PAUL R., National Board of Medical Examiners
KENDALL, LORNE M., Educational Testing Service
KENDALL, W. E., The Psychological Corporation
KENDRICK, S. A., College Entrance Examination Board
KENNEDY, JOHN JOSEPH, New York University
KENNEY, HELEN J., Harvard University
KENT, CLARENCE L., Virginia State Department of Education
KERSTING, ETHEL, Educational Testing Service
KIDD, SUZANNE E., Educational Testing Service
KIEFFER, JOHN E., Harker Preparatory School, Potomac, Maryland
KIGHT, HOWARD R., State University of New York at Buffalo
KILPATRICK, JEREMY, Stanford University
KIRKWOOD, VIRGINIA, Roanoke City (Virginia) Public Schools
KLARE, GEORGE R., Ohio University
KLEIN, PETER, New York Department of Social Welfare
KLEIN, STEPHEN P., Educational Testing Service
KLEINHANDLER, ANNE, The City College of New York
KLEINKE, DAVID J., New York State Department of Education
KLINE, WILLIAM E., Washington Pre-College Testing Program
KLOPF, GORDON, Bank Street College of Education
KNAG, HOWARD A., Queens College
KNAPP, THOMAS R., University of Rochester

KOFFLER, JULIAN S., New York University
KOGAN, LEONARD S., Brooklyn College
KOGAN, NATHAN, Educational Testing Service
KOSTIN, IRENE, Educational Testing Service
KOSTIN, MORTON, Educational Testing Service
KRASNER, LEONARD, Educational Testing Service
KRISTOF, WALTER, Educational Testing Service
KROFTA, JAMES M., Spring Valley (New York) High School
KRUG, ROBERT E., Peace Corps
KUBIS, JOSEPH F., Fordham University
KUJAWSKI, CARL J., The Atlantic Refining Company
KUNTZ, JAMES E., Texas Technological College
KURFMAN, DANA, Educational Testing Service
KURLAND, NORMAN D., New York State Department of Education
KVARACEUS, WILLIAM C., Tufts University
LAGEMANN, JOHN LORD, New York, New York
LAKE, BARBARA W., Fordham University
LAMB, HOWARD E., University of Delaware
LANE, HUGH W., National Merit Scholarship Corporation
LANG, GERHARD, Fairleigh Dickinson University
LANGMUIR, CHARLES R., The Psychological Corporation
LANKTON, ROBERT S., Detroit (Michigan) Public Schools
LANNHOLM, GERALD V., Educational Testing Service
LARSON, CEDRIC A., Rutgers, The State University
LARSON, RUTH, New York, New York
LATHROP, ROBERT, The Pennsylvania State University
LAWRENCE, SISTER HELEN, Trinity College
LAZAR, RICHARD L., Orange (New Jersey) Public Schools
LEARY, WILLIAM J., JR., The Pennsylvania State University
LEES, DIANA M., Educational Testing Service
LEHMKUHL, CARLTON B., Boston College
LEIBLER, SETH N., University of Rochester
LEIGHTON, P. L., East Stroudsburg State College, Pennsylvania
LEVERETT, HOLLIS M., Hollis M. Leverett and Associates, Inc.
LEVINE, HAROLD G., New York State Department of Education
LEVINE, MILTON, National Science Foundation
LEVITZ, BARBARA F., The City College of New York
LEWY, ARIEH, University of Chicago
LIBERMAN, SAMUEL S., Columbia University
LILLY, ROY S., Educational Testing Service
LINDEMAN, RICHARD H., Teachers College, Columbia University
LINDLEY, JESSE B., Educational Testing Service
LINDQUIST, E. F., State University of Iowa

LINDVALL, C. M., University of Pittsburgh
LINN, ROBERT L., University of Illinois
LIPPER, LOIS, Biometrics Research
LIPSITZ, HOWARD C., New York University
LITTLE, JAMES W., University of North Carolina
LOCKETTE, LOUISA L., Junior High School No. 3, Trenton, New Jersey
LO MONACO, LEON J., New York City Board of Education
LONG, LILLIAN D., Professional Examination Service, New York City
LONG, LOUIS, The City College of New York
LORD, FREDERIC M., Educational Testing Service
LOREE, M. RAY, University of Alabama
LORET, PETER G., Educational Testing Service
LORETAN, JOSEPH O., New York City Board of Education
LORR, MAURICE, Veterans Administration Hospital
LOTTES, JOHN J., State University of New York at Geneseo
LOWERY, ZEB A., College Board Test Center, Rutherfordton, North Carolina
LUCAS, DIANA D., Educational Testing Service
LUCAS, NORMA LEE, Mary Institute
LUSK, LOUIS T., Independent Vocational Counselor, Norwalk, Connecticut
LUTZ, ORPHA M., Montclair State College, New Jersey
LYMAN, HOWARD B., University of Cincinnati
LYNCH, ELEANOR, National League for Nursing
LYONS, WILLIAM A., New York State Department of Education
MACALUSO, C. J., United States Naval Examination Center
MAC GINITIE, WALTER H., Teachers College, Columbia University
MACOMBER, HERRICK M., The Phillips Exeter Academy, New Hampshire
MADAUS, GEORGE F., State College at Worcester, Massachusetts
MADDEN, RICHARD, San Diego State College, California
MADDOX, CLIFFORD R., Cedarville College
MAHAN, THOMAS W., JR., Connecticut State Department of Education
MAHNEN, HERMAN A., West Point Military Academy
MAIER, MILTON H., Educational Testing Service
MALCOLM, DONALD J., Educational Testing Service
MALKIEL, JUDITH A., Educational Testing Service
MANN, SISTER M. JACINTA, Seton Hall College
MANUEL, HERSCHEL T., University of Texas
MARCO, GARY L., Educational Testing Service
MARCO, MILLICENT, Long Branch (New Jersey) Public Schools
MARITA, SISTER, Rosary Hill College
MARSH, JAMES V., Educational Testing Service
MARSHALL, NORA A., Worcester (Massachusetts) Public Schools
MARTIN, JOSEPH B., Baylor School, Chattanooga, Tennessee
MASLOW, ALBERT P., United States Civil Service Commission

MASON, ANDREW W., Maryland State Department of Education
MAYHEW, LEWIS B., Stanford University
MAYO, SAMUEL T., Loyola University
MAYRAND, FRED R., JOIN, City of New York
MAZUR, JOSEPH L., Cleveland (Ohio) Public Schools
MC ADOO, HAROLD L., United States Civil Service Commission
MC CALL, W. C., University of South Carolina
MC CALLUM, WILLIAM SHAW, Cohasset (Massachusetts) High School
MC CANN, FORBES E., McCann Associates, Philadelphia
MC CARTHY, DOROTHEA, Fordham University
MC CONNELL, JOHN C., Windward School, White Plains, New York
MC CORD, RICHARD B., Philadelphia Personnel Department
MC CULLOCH, AGNES L., International Business Machines Corporation
MC DANIEL, SARAH W., Hofstra University
MC DONALD, R. P., Educational Testing Service
MC GEE, LINDBERGH, New York University
MC GUIRE, JOHN T., New York University
MC INTIRE, PAUL H., Boston University
MC KENNA, MAE R., Wilby High School, Waterbury, Connecticut
MC KEON, JAMES, Educational Testing Service
MC LAIN, JOSEPH C., Mamaroneck (New York) High School
MC LAUGHLIN, KENNETH F., United States Office of Education
MC LEAN, LESLIE DAVID, Teachers College, Columbia University
MC MORRIS, ROBERT F., State University of New York at Albany
MC MULLIN, THOMAS E., University of Pennsylvania
MC NAMARA, SISTER M. PATRICE, Office of Education, Milwaukee, Wisconsin
MC NATT, SUSANNE, Educational Testing Service
MC NULTY, THEODORE P., Educational Testing Service
MC PEEK, W. MILES, Educational Testing Service
MC QUITTY, JOHN V., University of Florida
MC QUITTY, LOUIS L., Michigan State University
MEDLEY, DONALD M., The City University of New York
MELBYE, BENNETT C., Frazee (Minnesota) High School
MELKER, FELICE, Veterans Administration
MELTON, RICHARD S., Educational Testing Service
MELVILLE, S. DONALD, Educational Testing Service
MERRITT, ROBERT T., College Entrance Examination Board
MERRY, MARGARET H., Boston University
MERRY, ROBERT W., Harvard University
MERWIN, JACK C., University of Minnesota
MESSICK, SAMUEL J., Educational Testing Service
MIANO, KATE P., Bureau of Curriculum Research, New York City
MIDDENDORF, LORNA, Rutgers, The State University

MIERZWA, JOHN A., The University of Pennsylvania
MILHOLLAND, JOHN E., University of Michigan
MILLARD, KENNETH A., U. S. Department of Health, Education and Welfare
MILLER, DOROTHY F., Clayton (Missouri) School District
MILLER, E. JOYCE, Seton Hall University
MILLER, HOWARD G., North Carolina State University
MILLER, MARIAN B., Delaware State Department of Public Instruction
MILLER, PAUL VAN R., JR., Educational Testing Service
MILLER, ROBERT M., Havre DeGrace (Maryland) High School
MILLMAN, JASON, Cornell University
MIRKIN, LOUISE, Educational Testing Service
MITCHELL, BLYTHE C., Harcourt, Brace and World, Inc.
MITZEL, HAROLD E., The Pennsylvania State University
MOLDOVAN, LEONARD, Junior High School No. 38, Bronx, New York
MOLLENKOPF, WILLIAM G., Proctor and Gamble Company
MOORE, OMAR K., Rutgers, The State University
MOORE, RUTH, Rutgers, The State University
MOORING, DANA K., McGraw-Hill Book Company
MORIARTY, DORIS M., Educational Testing Service
MORRISON, ALEXANDER W., Polytechnic Institute of Brooklyn
MORRISON, FRANCES CROOK, State University of New York at Albany
MOSS, SIMEON F., New Jersey Department of Labor and Industry
MUIRHEAD, DAVID B., University of Colorado
MUKHERJEE, G., Windsor Mountain School, Lenox, Massachusetts
MULLER, RAYMOND I., Montgomery County (Maryland) Public Schools
MULLIGAN, CATHERINE C., St. Francis College
MUNGER, A. M., The Psychological Corporation
MURRAY, KAY E., State University of New York at Albany
MYERS, ALBERT E., Yale University
MYERS, CHARLES T., Educational Testing Service
MYERS, ROBERT, Administration Building, Akron, Ohio
MYERS, SHELDON S., Educational Testing Service
NEALE, A. D., U.S. Department of Health, Education and Welfare
NEIVERT, SYLVIA, Harcourt, Brace and World, Inc.
NELSON, H. ROBERT, Highland Park (New Jersey) High School
NELSON, LEWIS J., New York University
NEULINGER, JOHN, Russell Sage Foundation
NOLL, VICTOR H., Michigan State University
NORRELL, GWEN, Michigan State University
NORTH, ROBERT D., Educational Records Bureau
NOVICK, MELVIN R., Educational Testing Service
NULTY, FRANCIS X., Educational Testing Service

OATHOUT, D. V., American Telephone and Telegraph Company
OHNMACHT, FRED W., University of Maine
OILL, BENJAMIN C., New York City Department of Personnel
OLANOFF, MARTIN, Board of Education, Brooklyn, New York
OLSON, LE ROY A., Michigan State University
O'NEAL, JOHN F., Long Island University
O'NEILL, JOHN J., Rutgers, The State University
OPPENHEIM, DON B., Educational Testing Service
OPPENHEIMER, VICTOR, Harvard University
ORAHOOD, ELIZABETH, Kansas City (Missouri) Public Schools
ORLEANS, ISAAK D., United States Army Personnel Research Office
ORLEANS, JOSEPH B., Ramaz School, New York City
ORR, CORA I., Muskingum College
ORR, DAVID B., American Institute for Research
ORSHANSKY, BERNICE, Bronxville (New York) Senior High School
OSTRAM, ELIZABETH, New York State Department of Civil Service
PACE, C. ROBERT, University of Southern California
PAGE, ELLIS B., University of Connecticut
PAINE, THOMAS E., U. S. Department of Health, Education and Welfare
PALMER, ORVILLE, Educational Testing Service
PALUBINSKAS, ALICE L., Tufts University
PARSLEY, KENNETH M., University of Alabama
PASCIUTI, E. C., New Rochelle (New York) High School
PAYNE, DAVID A., Syracuse University
PAYNE, HENRY, New Rochelle (New York) High School
PEDHAZUR, E., New York University
PEMBERTON, W. A., University of Delaware
PERRY, W. D., University of North Carolina
PETERS, FRANK R., State University of New York at Stony Brook
PETERSON, RICHARD E., Educational Testing Service
PETRELLA, JOSEPH, Junior High School 118, Bronx, New York
PETRULLO, LUIGI, Office of Naval Research
PFEIFFER, ANN, Baltimore City (Maryland) Public Schools
PHILIPS, C. LEE, University of Miami
PHILLIPS, MORTON, Queens, New York
PHIPPS, LILA O., Westfield (New Jersey) Senior High School
PICARD, OTTO R., Frank K. Hehnly School, Clark, New Jersey
PIERS, ELLEN V., The Pennsylvania State University
PIERSON, ELLERY M., Educational Testing Service
PIKE, LEWIS, Educational Testing Service
PITCHER, BARBARA, Educational Testing Service
PLUMLEE, LYNNETTE B., Sandia Corporation, Albuquerque, New Mexico
POLLACK, NORMAN C., New York State Civil Service Department

PRESCOTT, GEORGE A., University of Maine
PRIOR, JOHN J., Pratt Institute
PRUZEK, ROBERT, University of Wisconsin
PURCELL, WILLIAM D., Summit (New Jersey) Public Schools
QUINN, JOHN S., Harcourt, Brace and World, Inc.
RA, JUNG BAY, Harcourt, Brace and World, Inc.
RABINOWITZ, WILLIAM, The Pennsylvania State University
RAMSAY, JAMES O., Educational Testing Service
RAPH, JANE B., New York, New York
RAPHAEL, BROTHER ALOYSIUS, F.S.E., Bishop Loughlin Memorial High School, Brooklyn
RAPPARLIE, JOHN H., The Owens-Illinois Glass Company
RASKIN, EVELYN, Brooklyn College
RAYMOND, T. CICCHINO, Harvard University
READING, ETHEL, Educational Testing Service
REED, SAMUEL C., Educational Testing Service
REELING, GLENN E., Montclair (New Jersey) Public Schools
REID, JOHN W., Indiana State College
REINER, WILLIAM B., New York City Board of Education
REMMERS, HERMANN H., Purdue University
REUTER, WILLIAM H., Educational Testing Service
REYNOLDS, HARLAN J., International Business Machines Corporation
RHOADES, DAVID, University of Nebraska
RICKS, JAMES H., JR., The Psychological Corporation
RITENOUR, LOUISE, Educational Testing Service
ROBERTS, G. TRACEY, State Civil Service Commission
ROBERTS, S. O., Fisk University
ROBERTSON, D. MOTT, Roanoke City (Virginia) Public Schools
ROBERTSON, GARY J., Harcourt, Brace and World, Inc.
ROBINS, ALEXANDER R., U. S. Department of Health, Education and Welfare
ROBINSON, DOROTHY F., Paterson State College
ROCA, PABLO, Inter-American University, Puerto Rico
ROCK, DONALD A., Educational Testing Service
ROHRBAUGH, FRANCES G., Educational Testing Service
ROMANELLA, ALAN E., Port Jefferson (New York) Public Schools
ROMBERG, THOMAS, Stanford University
ROSEN, JULIUS, New York University
ROSENBLATT, SIDNEY M., The City College of New York
ROSENTHAL, ELSA, Educational Testing Service
ROSNER, BENJAMIN, Educational Testing Service
ROSS, PAUL F., Standard Oil Company of New Jersey
ROSS, ROBERT T., California State Department of Mental Hygiene

ROTHENBERGER, RUSSELL R., Professional and Occupational Affairs (Pennsylvania)
ROYSON, HARRY, The Rehabilitation Center, New York City
RUDMAN, HERBERT C., Michigan State University
SACHS, LORRAINE P., National League for Nursing
SAFRAM, MORRIS, Franklin High School, New York City
SAKODA, JAMES M., Brown University
SAMLER, JOSEPH, Veterans Administration
SAND, LOUISE, University of South Florida
SASAJIMA, MASU, Educational Testing Service
SASLOW, MAX S., New York City Department of Personnel
SAUNDERS, DAVID R., Princeton, New Jersey
SCATES, ALICE Y., United States Office of Education
SCHEINHAUS, PHYLLIS, City College of New York
SCHELDON, ETHEL, Brooklyn, New York
SCHERER, FRANK H., Rutgers, The State University
SCHLEIFER, JOSEPH L., Scarsdale, New York
SCHNEIDER, HARRIET, National League for Nursing
SCHNITZEN, JOSEPH P., University of Houston
SCHOEPPS, MARJORIE E., Educational Testing Service
SCHRADER, WILLIAM B., Educational Testing Service
SCHUBERT, HERMAN J. P., Veterans Administration Hospital
SCHULMAN, JAY, Institute for Career Guidance, New York City
SCHULTZ, CHARLES B., University of Pennsylvania
SCHULTZ, DOUGLAS, Western Reserve University
SCHULZ, DELPHIN L., The Lutheran Church-Missouri Synod
SCHUMACHER, CHARLES F., National Board of Medical Examiners
SCHWARTZ, PAUL, New York University
SCHWARZ, MARIETTE, Educational Testing Service
SCOFIELD, LEONARD G., Dean Junior College
SCOTT, C. WINFIELD, Rutgers, The State University
SCOTT, GARY J., The Pennsylvania State University
SEALY, ROSE F., New York University
SEASHORE, HAROLD, The Psychological Corporation
SEBASTIAN, JOHN P., New York University
SEIBEL, DEAN, Educational Testing Service
SEIBEL, ROBERT, The Pennsylvania State University
SENFT, ALBERT, Shaker Heights (Ohio) High School
SERAFINO, ROBERT P., Educational Testing Service
SERAYDARIAN, THOMAS, Princeton (New Jersey) High School
SERLING, ALBERT M., Educational Testing Service
SETZER, CHARLES J., New York City Department of Personnel
SEUTER, DONALD R., Detroit (Michigan) Public Schools

SEYFRIT, PAUL L., Kentucky Village Department of Child Welfare
SFORZA, RICHARD F., United States Military Academy
SHAHINIAN, SIROON, Consultant on Services for Aged
SHAPIRO, BERNARD J., Harvard University
SHARP, CATHERINE G., Educational Testing Service
SHAYCOFT, MARION F., American Institute for Research
SHEEHE, MARCIA, University of Rochester
SHEINMEL, BRENDA, City College of New York
SHENDALMAN, ANITA C., Educational Testing Service
SHERMAN, EDGAR M., Irvington (New Jersey) High School
SHIELDS, MARY, National League for Nursing
SHIMBERG, BENJAMIN, Educational Testing Service
SHIMBERG, HELEN, Princeton, New Jersey
SHIRKEY, EDWIN C., U. S. Army Enlisted Evaluation Center
SIEGEL, ARTHUR, Applied Psychological Services
SILVA, MILTON N., College Entrance Examination Board, Puerto Rico
SILVERBERG, EDWARD B., New York City Department of Personnel
SINGLETON, CARLTON M., Wesleyan University
SKAGER, RODNEY W., Educational Testing Service
SMITH, ALEXANDER F., Southern Connecticut State College
SMITH, ALLAN B., Rhode Island College
SMITH, ANN Z., Educational Testing Service
SMITH, MARSHALL P., Trenton State College
SMITH, MARION S., South Tenth Street School, Newark, New Jersey
SOAR, ROBERT S., Temple University
SOLOMON, HERBERT, Stanford University
SOLOMON, ROBERT, Educational Testing Service
SOMMER, JOHN, Houghton Mifflin Company
SONTAG, MARVIN, New York City Board of Education
SOSULSKI, M. C., St. John Fisher College
SOTHERLAND, DONALD W., New York Department of Civil Service
SOUTHER, MARY T., Tower Hill School, Wilmington, Delaware
SPANEY, EMMA, Queens College of the City University of New York
SPAULDING, GERALDINE, Educational Records Bureau, Winston-Salem, North Carolina
SPEER, GEORGE S., Illinois Institute of Technology
SPENCER, RICHARD E., University of Illinois
SPREMULLI, ESTELLE, Educational Testing Service
SQUIRES, JO ANNE S., Hampton Roads Speech and Hearing Center, Newport News, Virginia
STAHL, SUZANNE, Educational Testing Service
STAKE, ROBERT E., University of Illinois
STANLEY, JULIAN C., University of Wisconsin

STATLER, CHARLES R., University of Iowa
STEINMAN, ARTHUR M., Trenton State College
STELLER, DANIEL J., Pace College
STEPHEN, SISTER MARY, Villa Julie College
STEPHENS, ROBERT W., Elma, Washington Schools
STEWART, CLIFFORD, University of South Florida
STEWART, E. ELIZABETH, Educational Testing Service
STEWART, NAOMI S., Cranford, New Jersey
STICKELL, DAVID W., Bucknell University
STILES, GRACE ELLEN, University of Rhode Island
STIMPEL, SIDNEY, Martin C. Johnson Reporting Company
STOKER, HOWARD W., Florida State University
STOLTZ, ROBERT E., Science Research Associates
STONE, PAUL T., Huntingdon College
STONE, SELMA, The City College of New York
STOUGHTON, ROBERT W., Connecticut State Department of Education
STRICKER, LAWRENCE J., Educational Testing Service
STUDDIFORD, WALTER B., Princeton University
STUNKARD, CLAYTON L., University of Maryland
SUPER, DONALD E., Columbia University
SUSSMAN, LEAH, Newark (New Jersey) School System
SUTTON, JOSEPH T., Stetson University
SWAN, BEVERLEY B., Florida State Department of Education
SWANSON, EDWARD O., University of Minnesota
SWINEFORD, FRANCES, Educational Testing Service
TABER, VICTOR A., New York State Department of Education
TANNER, DANIEL, The City University of New York
TATSUOKA, MAURICE, University of Illinois
TAYLOR, BRUCE, Educational Testing Service
TAYLOR, JUSTINE, Educational Testing Service
TAYLOR, SAMUEL J., New York State Civil Service Department
TCHORNI, BERNARD, Educational Testing Service
TERRAL, JOSEPH E., Educational Testing Service
TERWILLIGER, JAMES S., Peabody College
THANOS, CYNTHIA, The City College of New York
THEVAOS, DENO, The Pennsylvania State University
THOMAS, WILLIAM F., University of Wisconsin
THOMPSON, RAYMOND E., Educational Testing Service
THOMSON, MURIEL, Los Angeles, California
THORNDIKE, ROBERT L., Teachers College, Columbia University
THORNE, MARGARET A., Educational Testing Service
THURSTONE, THELMA GWINN, University of North Carolina
TRAXLER, ARTHUR E., Educational Records Bureau

TRIGGS, FRANCES, Committee on Diagnostic Reading Tests, Inc.
TRISMEN, DONALD, Educational Testing Service
TUCKER, DONALD K., Northeastern University
TUCKER, LEDYARD R, University of Illinois
TUFTE, CORINNE G., Teachers College Computer Center, New York City
TULLY, EMERSON, Florida State University
TURNBULL, WILLIAM W., Educational Testing Service
TYLER, GLADYS K., Psychological Services, Hamden, Connecticut
UMANS, SHELLEY, New York City Board of Education
UNGER, R., New York, New York
URQUHART, HELEN L., Brown University
URRY, VERN W., U. S. Army Enlisted Evaluation Center
VALLEY, JOHN R., Educational Testing Service
VAN AUSDALL, LEIGH, Science Research Associates
VANICA, SUMNER, Akron, Ohio
VECCHIONE, NICHOLAS, Educational Records Bureau
VISCEGLIA, JOHN A., Camden (New Jersey) Board of Education
VITELES, MORRIS S., University of Pennsylvania
VOLANTE, WILLIAM, The Pingry School, Elizabeth, New Jersey
VON FANGE, ERICH A., Concordia Lutheran Junior College
VRIEND, JOHN, New York University
WAGNER, EDWARD P., Bloomesburg State College, Pennsylvania
WAGNER, MAZIE EARLE, State University of New York at Buffalo
WAHLGREN, HARDY L., State University of New York at Geneseo
WAIDLEY, JOHN W., New York University
WALKER, HELEN M., Teachers College, Columbia University
WALKER, ROBERT N., Personnel Press, Inc., Princeton, New Jersey
WALL, VICTOR D., LaJunta (Colorado) High School
WALLACE, WIMBURN L., The Psychological Corporation
WALSH, JEANNE, Washington Square College
WALTHEW, JOHN K., Educational Testing Service
WALTON, WESLEY W., Educational Testing Service
WALTZMAN, HAL, New York State Department of Civil Service
WALWER, FRANK K., Columbia University
WANTMAN, MOREY, Educational Testing Service
WARD, ANNIE W., Volusia County (Florida) Public Schools
WARE, DOROTHY A., Detroit Psychological Clinic
WARRINGTON, WILLARD G., Michigan State University
WATERS, MAX, The Pennsylvania State University
WATKINS, RICHARD W., Educational Testing Service
WATT, ROBERT J., Glenbrook North High School, Northbrook, Illinois
WEBER, BURTON J., Ridgefield (New Jersey) Memorial High School
WEBER, EWALD, New York City Department of Personnel

WECHSLER, DAVID, Bellevue Psychiatric Hospital, New York
WEINBERG, CARL, University of Delaware
WEINER, MAX, Brooklyn College
WEISBRODT, JEROME, Purdue University
WEITZ, HENRY, Duke University
WELNA, CECILIA, University of Hartford
WESMAN, ALEXANDER G., The Psychological Corporation
WEST, LEONARD J., The City University of New York
WHALEY, FRANCIS L., The Pennsylvania State University
WHETSTONE, ROBERT D., University of Colorado
WHITE, HERBERT L., New York City Department of Personnel
WHITLA, DEAN K., Harvard University
WIENER, ROSA D., Learning Institute of North Carolina
WILDE, GERRIT J., Institute of Clinical and Industrial Psychology, Holland
WILDER, DAVID H., Bucknell University
WILDER, GITA, Educational Testing Service
WILDS, PRESTON LEA, Medical College of Georgia
WILKE, MARGUERITE M., Hunter College
WILKE, WALTER H., New York University
WILLARD, RICHARD W., Massachusetts Institute of Technology
WILLEY, CLARENCE F., Norwich University
WILLIAMS, CLARENCE M., University of Rochester
WILLIAMS, E. BELVIN, Teachers College, Columbia University
WILLIAMS, GERALD, The Pennsylvania State University
WILLIAMS, IRENE, Educational Testing Service
WILLIAMS, JOHN M., Educational Testing Service
WILLIAMS, MALCOLM J., United States Coast Guard Academy
WILLIAMS, ROGER K., Morgan State College
WILLIS, WARREN K., University of Illinois
WILSON, MARGARET, Educational Testing Service
WINGO, ALFRED L., Virginia State Department of Education
WINIEWICZ, CASIMERE S., United States Naval Examining Center
WINTERBOTTOM, JOHN A., Educational Testing Service
WIRSIG, JANE, Educational Testing Service
WODTKE, KENNETH H., The Pennsylvania State University
WOHLHUETER, JAMES F., Educational Testing Service
WOLF, RICHARD, University of Chicago
WOLINS, LEROY, Iowa State University
WOLMAN, BENJAMIN B., Adelphi University
WOMER, FRANK B., University of Michigan
WOOD, BEN D., Columbia University
WRIGHT, DANIEL W., New York University
WRIGHT, WILBUR H., State University of New York at Geneseo

WRIGHTSTONE, J. WAYNE, New York City Board of Education
YEVAK, MARIE E., Paterson State College
YOUNGQUIST, OWEN F., JR., Huntington (New York) School District
YUNDT, JOHN H., Moorestown (New Jersey) Public Schools
ZALKIND, SHELDON, The City College of New York
ZARRO, PASQUALE J., Philadelphia Personnel Department
ZELIN, MARTIN, Queens College of the City University of New York
ZIMILES, HERBERT, Bank Street College
ZOOK, DONOVAN Q., Board of Examiners for the Foreign Service
ZUBIN, JOSEPH, Department of Mental Hygiene, New York