DOCUMENT RESUME

ED 080 540 TM 003 045

TITLE Invitational Conference on Testing Problems (New York, October 29, 1966).
INSTITUTION Educational Testing Service, Princeton, N.J.
PUB DATE 29 Oct 66
NOTE 128p.

EDRS PRICE MF-$0.65 HC-$6.58
DESCRIPTORS Achievement; *Adjustment (to Environment); Computer Oriented Programs; *Conference Reports; *Educational Change; Educational Innovation; Educational Objectives; *Evaluation; Instructional Innovation; Prediction; Program Evaluation; Psychometrics; Statistical Analysis; *Testing Problems

ABSTRACT

The 1966 Invitational Conference on Testing Problems dealt with the innovations of the new age of flexibility and the problems of evaluating and preparing for them. Papers presented in Session I, Innovation and Evaluation, were: (1) "Innovation and Evaluation: In Whose Hands?" by Nils Y. Wessell; (2) "The Discovery and Development of Educational Goals" by Henry S. Dyer; (3) "The Meaning of Impact" by Martin Trow; (4) "Unconventionality, Triangulation, and Inference" by Eugene J. Webb; and (5) "The Prediction of Academic and Nonacademic Accomplishment" by John L. Holland. The luncheon address was "Education's Age of Flexibility" by Francis Keppel. Papers presented at Session II, Natural Language and Computers in Education, were: (1) "An Interactive Inquirer" by Philip J. Stone; (2) "The Natural-Language Approach to Psychometrics" by Carl E. Helm; and (3) "Grading Essays by Computer: Progress Report" by Ellis Batten Page. A list of conference participants concludes the report. (KM)

FILMED FROM BEST AVAILABLE COPY

ED 080540

TE 003 045

PROCEEDINGS

ED 080540

Invitational Conference on Testing Problems

October 29, 1966
Hotel Roosevelt
New York City

JULIAN C. STANLEY
Chairman

EDUCATIONAL TESTING SERVICE
Princeton, New Jersey
Berkeley, California
Evanston, Illinois

Testing Problems in Perspective

At the 1966 Invitational Conference, Dr. Anne Anastasi, who edited TESTING PROBLEMS IN PERSPECTIVE, *was presented with a copy of the anthology by William W. Turnbull, Executive Vice President of Educational Testing Service. They are shown above with Henry Chauncey, President of* ETS, *and Robert Quick, Director of Publications for the American Council on Education.*

The Invitational Conference on Testing Problems, in its history of a little more than 25 years, has spanned a period in which educational measurement has changed from a field of specialized interest to a central focus for educational development and planning at all levels, from the local school to the federal government. Throughout this period the Invitational Conference has provided a forum in which leaders in this field have expressed their thoughts about testing problems. Thus, the *Proceedings* of this conference have consistently mirrored the best and most provocative thinking in each stage of the development of testing as an art and a science.

It seemed, therefore, that the field of educational and psychological measurement would be well served if a group of papers from past conferences could be selected for their continuing timeliness and value, organized topically, and published in a single volume. To our great good fortune, Dr. Anne Anastasi, Professor of Psychology at Fordham University, agreed to

serve as editor of the book. The papers were selected by Dr. Anastasi with the advice of former conference chairmen and staff members of Educational Testing Service.

Those who attended the 1966 Invitational Conference were the first to see the results of Dr. Anastasi's work. At the opening session of the conference, Dr. Anastasi was presented with a copy of *Testing Problems in Perspective*, which was published two days later, on October 31, by the American Council on Education.

Testing Problems in Perspective contains 58 papers by 47 authors and deals with three significant areas of concern in the field of testing: Test Development and Use, Psychometric Theory and Method, and Special Problems in the Assessment of Individual Differences. The introduction by Dr. Anastasi provides a history of the conference, and her commentary on each major topic points up the significant developments in that area of measurement.

We owe thanks to Dr. Anastasi and all those who worked with her on this book. We hope it will be a successful and informative reader in the field of measurement for students and for those in the profession.

William W. Turnbull
EXECUTIVE VICE PRESIDENT

Foreword

In his luncheon address to the 1966 Invitational Conference on Testing Problems, Dr. Francis Keppel sounded the keynote of the conference when he stated that ". . . we have to prepare young people, and older people as well, for a persistently changing world. This cannot be done unless we help to make them more amenable to change, more flexible individuals. To do its part of the job, education must itself wholeheartedly enter the new age of flexibility . . ."

The innovations of this new age and the problems of evaluating and preparing for them formed the basis of this year's conference. Speakers at the morning session raised a number of important questions about the problems of shaping our educational system to meet the demands of a changing society: How can educators discover proper goals for education when no one knows what the world in 20 years will require of students? How can we determine which aspects of the college and university experience strengthen those qualities of self-confidence and the ability to learn that should be an important part of the impact of education? What can be done to eliminate the lag between the development of an exciting approach to education and its use in the classroom? These and other questions about evaluation were followed by a discussion of some exciting innovations now under way in education. Speakers described what they have been doing and what they hope to do with computers in grading essays, analyzing language, and helping to formulate scientific theories.

These considerations of our present problems and glimpses into the future combined to provide a program that was well-balanced and exciting. I should like to extend our thanks to Dr. Julian Stanley who, as chairman, made this program possible, and to those distinguished speakers whose papers appear in these *Proceedings.*

Henry Chauncey
PRESIDENT

Preface

As organizer and chairman of the 1966 Invitational Conference on Testing Problems, sponsored by Educational Testing Service, I was aided greatly by the accumulated suggestions of prior ICTP chairmen, particularly Robert L. Ebel and Chester W. Harris. Also, a personal visit to ETS in January of 1966 resulted in many other excellent recommendations. Throughout the planning for this conference, Anna Dragositz and her efficient co-workers at ETS made my task interesting and far easier than it otherwise would have been. President Henry Chauncey obtained Francis Keppel as the luncheon speaker and chaired that session.

My approach was simple. With the above assistance, I sought out eight highly able persons doing interesting research relevant to "testing problems," broadly defined, and asked them to talk about whatever aspects of their investigations they wished. Seven of these are psychologists of various ages and persuasions. The other (Martin Trow) is a sociologist especially concerned with higher education in the United States and England.

All nine speeches seemed well received by a large audience. Details that otherwise might have been too technical for the occasion were presented with verve and humor. Many ingenious ideas which should further measurement principles and practices are contained in the nine papers, several of which relate to each other in heuristic or provocative ways. I feel confident that you will be repaid amply for reading the entire volume without delay, even though you may have heard the speeches.

Julian C. Stanley
CHAIRMAN

Contents

Session I

Theme:
Innovation and Evaluation

Innovation and Evaluation: In Whose Hands?

NILS Y. WESSELL
Institute for Educational Development

In the few months that the Institute for Educational Development (IED) has been in existence, one overriding conclusion has become clear. It is that the concerns implied by the title of my remarks are troubling thoughtful persons in all sectors of our national life, public and private. Officials of government agencies, foundation executives, teachers, school administrators, curriculum specialists, educational researchers, and officers of corporations producing educational materials all are asking the same difficult questions. All express an urgent desire to hear answers to these questions.

How is educational innovation best encouraged and sustained? What steps can shorten or eliminate the lag between development and testing of a promising new approach or product and its use in the classroom? How can the resources, in talent and money, represented by public and private agencies come together to upgrade our schools? What measures can effect, in society's interests, better rapport and cooperative endeavor between those individuals and agencies dealing with education on a nonprofit basis and those properly concerned with making a profit in the production and sale of educational materials?

How can new ideas and new products best be appraised and the results of such appraisal made quickly available to the consumer in usable form? What is the role of government in the setting of standards? Have we defined adequately as a prior step the criteria and objectives of education, with reference to broad goals, or even with reference to specific courses and curricula? If we can reach some agreement on criteria and objectives, do we have available the techniques for determining the extent to which such criteria and objectives

are met by school systems as a whole or by p ses or particular materials? What are the proprieties w ould obtain between government and industry in the field of education?

This list of questions is only a sampling of the queries raised. A lingering doubt seems always to be present in the minds of the questioners, a suggestion that either our ignorance is great or the road is long, or both.

An answer as to where we start or how we proceed will not evoke the same unanimity that attaches to the underlying concern that "begin we must." It was with this conviction that the Institute for Educational Development was created and set to work. But while its first tenet is one of optimism, it lays no claim to an ability to come up with the answers to all of the questions raised. Its commitment rather is to proceed, to select priorities, and to be prepared for failure, ambiguity, and the development of expertise on blind alleys. While IED does have convictions about its special role and potential, it exercises eminent domain over no area. On the contrary, with a feeling of urgency, it invites all with similar concerns to enter the field.

The Institute for Educational Development received its charter in 1965 as a nonprofit educational corporation in the state of New York. The early ideas and direct assistance toward formation of IED came from Educational Testing Service. Its original trustees were six in number and included John Corson, then a professor of public and international affairs, Woodrow Wilson School, Princeton University; Henry Chauncey, President of Educational Testing Service; John Fischer, President of Teachers College, Columbia University; Albert H. Bowker, Chancellor of the City University of New York; Wallace Macgregor, Executive Vice President of American Metal Climax; and Harold Howe II, then Director of the Learning Institute of North Carolina. When Mr. Howe became U.S. Commissioner of Education, he resigned as a trustee of IED, and his place was taken by Charles Brown, Superintendent of Schools in Newton, Massachusetts. Dr. Brown serves also as chairman of a project advisory committee whose membership is drawn from persons in educational research or in school administration. Vice presidents of IED are John L. Kennedy, for eight years Chairman of the Department of Psychology at Princeton University, and Donald E. Barnes, formerly with the University of Chicago Press, financial corporations, and the Center for Programmed Instruction. A small professional staff has been drawn from school administration, instructional technology, and educational research.

The Institute's program concentrates on learning and instruction, with emphasis on curriculum materials, equipment, systems, and evaluation.

To the extent that IED has a special role, it is probably to bring together for the common good the resources of the business community, the educational world, and government agencies in ways that will ensure the full utilization of their resources which in their totality are almost without limit in their promise for the improvement of education. Dealings based upon suspicion or inadequate communication, misunderstanding, and the development of programs and approaches in isolation are clearly not in the interest of society or of education.

A second broad conclusion is also clear to us. No single sector of our society has a monopoly on men of talent and of good will. Some of the strongest advocates of high standards and objective evaluation are found in the commercial firms interested in education for the profits that can be made. Persons in the academic world and persons in the business world do have motives that differ, but the extent to which they overlap is impressive. Moreover, it is quite possible to respect and to value motives which are different from your own. Of course shoddy workmanship, deceiving salesmanship, and an undue concern for profits do characterize some commercial concerns in the educational field, but unfounded suspicion of all firms is not the best route to effective rapport.

Now let us turn from these broad generalizations to the critical and disturbing questions raised in my opening paragraphs. In the vernacular of the times, these are "gut" questions and too often caution or timidity or unreasonableness posing as pristine academic virtue has inhibited even trial-and-error approaches to solutions.

Commissioner of Education Harold W. Howe II (2), in his address in August to the American Management Association Conference on Industry and Education, described three ways of approaching the problem of evaluation and maintenance of proper standards. These can serve as three answers to the question of who should bear the final responsibility for evaluation:

1. An educational Consumer's Union, modeled on the operation that has for some years assisted buyers of commercial products, is his first approach. He suggests it be nonprofit and supervised by a standard-setting group representing education, business, foundations, state school departments, and federal officials.

2. His second solution is a committee on educational development which would be patterned after the Committee on Economic Development. Representatives from government, education, and industry would comprise the committee, but it would be beholden to no group.

3. His third suggestion is a regulatory agency similar to the United States Food and Drug Administration.

Howe's model of the regulatory agency is a last resort, I trust, which will never require serious consideration. May circumstances never invite or demand its consideration. While the FDA may well be the one appropriate approach in its domain, it will be a sad day indeed for all of us if we allow a situation to develop which will demand government policing in the production and marketing of edu- .tional materials. The blame will belong to all of us.

It is a mistake also to assume that clear black-and-white distinctions can be made or are even desirable with respect to the evaluation of programs and materials. One dimensional "seals of approval" are unreliable, if not useless. Even rank-order ratings involve physical and temporal unrealities in many instances. (For example, how could 10,000 textbooks—or any other kinds of products for that matter—be so ranked or approved?) Those who advocate seals of approval and rank ordering ignore the diversity of objectives among schools, grades, teachers, and class populations. They also assume the existence of evaluation techniques of demonstrated feasibility as well as proven reliability and validity.

Evaluation of all kinds is certainly taking place—some of it careful and reasonably objective, but much of it best described as "willy-nilly." This seems to suggest a first order of business—the development of a taxonomy of evaluation. At the risk of bringing down from on high the wrath of my "scientific" colleagues in psychology and in educational research, may I suggest that in the practical school situation, there are circumstances in which the adequate and feasible approach consists simply of making a long distance telephone call to an acknowledged authority. If we call this method one end of our evaluation spectrum, then at the other end would be carefully controlled, long-range studies of matched groups. The taxonomy of evaluation to which we refer would be developed by testing the many points or approaches between these two extremes, and including

these extremes, determining the degree of confidence to be attached to each, and improving the reliability, validity, and feasibility of all of them. Developing a taxonomy of evaluation must include a concerted effort to close the gap between existing methodology and its use, a general concern which applies to so much in education. Appropriate training in adequate numbers of competent personnel is no small part of the problem.

Much needs to be done also in interpreting the needs and standards of the academic world to the business world and the needs and standards of the business world to the academic world. Is it reasonable to expect a commercial firm to invest \$50,000 in the evaluation of a product which costs only \$20,000 to produce? In the competitive commercial world a company's advantage over a competitor may be only in lead time and not in quality of the product. Yet statements by the commercial producer regarding the evaluation procedures followed with respect to a particular product must be clear and not misleading, must state fully and without exaggeration what was done to determine the adequacy of the product. We may be naive or optimistic, but we do believe the economics of marketing can be shown to require such candor by the producer.

Promising and important though the development of a taxonomy of evaluation may be, we cannot await its ultimate refinement. Decisions and choices must be made now—*are* being made now by teachers and administrators and school boards—using the methods available. For this reason, IED is embarking at the same time on a totally different approach to evaluation which has within it the possibility of more immediate usefulness. The project is known as the Educational Products Information Exchange, or EPIE for short. It is based on the premise that there exists today in fragmented form much useful information about educational products based on the experience of those using them. When brought together, collated, and interpreted, such information, coupled with available information from the producers of the materials, can raise immediately and by several steps the soundness of decision making in our schools. From a network of schools representative of many kinds of communities and diverse educational objectives, information about the actual experience of teachers with educational materials will be gathered, interpreted, and made available to specific schools and to producers of materials. We are seeking advice from both educators and commercial producers in the development of EPIE. While in its early stages it will require sup-

port from government agencies and foundations, we confidently expect that in about three years EPIE will be self-sustaining.

An important part of this development will be the improvement of the evaluation and reporting techniques used by the schools providing information to EPIE. The project itself will require some research and testing, but other projects undertaken by IED will also contribute to making EPIE a more reliable and useful source of evaluation information. The program referred to earlier as a taxonomy of evaluation will obviously have such an impact on EPIE.

IED will also be concerned with the derivation of proper objectives and criteria for school programs and curricula. Evaluation to be significant must be related to clearly described and relevant criteria and objectives. These can vary from school to school, from grade to grade, and within a given grade on the basis of individual differences among the students. School and community environments also have a bearing, and more often than not, the larger context of the course as well as the larger context of the community must be taken into account. If I seem to dismiss the importance of such considerations by so casually referring to them, this is not my intention. The full space allotted to me for these remarks could well be addressed to them.

While designers and producers of educational materials will certainly find useful the kinds of information gathered and processed by the Educational Products Information Exchange, it is clear that EPIE is but one device or one approach. This was implied by my references to a taxonomy of evaluation and to other programs in evaluation being advanced by IED. Yet even these overlook quite another approach—a direct involvement in product design. Evaluation in the broadest and best sense should be a continuing process and should not be limited solely to the appraisal of finished products in use. However, such services to producers of materials, whether they are corporations operated for profit or nonprofit agencies, must be kept clearly separate from and independent of the kind of assessment represented by EPIE or by a number of other approaches. It may be that to be most effective and to maintain credibility IED should give counsel with respect to the *concept* and *process* of evaluation and not with respect to specific product design. Here particularly we have much to learn. A period of trial and error is clearly ahead of us.

But the title of my remarks refers to *innovation* as well as to evaluation. Thus far, I have addressed myself almost entirely to evaluation although I must point out quickly that innovation and evaluation are

very often inseparable, inextricable, and mutually interdependent. I grant that the relationship can be negative as well as positive. The application of irrelevant, complex, or unreasonable standards of evaluation can hinder or discourage innovation. Improved and more relevant criteria and objectives must accompany the evaluation of many innovations, for old goals and expectations sometimes bear no relation to the new and the original. For example, evaluation yardsticks designed to measure specific knowledge and skills in a particular subject matter field may shed little light on the usefulness of a new program whose emphasis is on learning methods and cognitive processes. On the other hand, it is equally short-sighted to condemn all forms of evaluation that go beyond the personal and the subjective. It is strange indeed that some curriculum innovators do not perceive that there are innovators in evaluation also.

It seems to us that innovation in methods of assessment may be as important a kind of educational innovation as there is. Without it, promising innovations in curriculum or approaches to learning or materials may never have their promise revealed and may lapse into disuse only because the results cannot be identified and appraised with confidence. To this large and complex task, IED will also direct some of its energies.

As Launor Carter (1) pointed out at last winter's meetings of the American Educational Research Association, the sequence from research to development to utilization of research results is very seldom a smooth one. He referred to the extensive study completed for the Department of Defense by the Arthur D. Little Company in which it was found that the transition did not proceed necessarily in logical fashion, and that phases assumed to be sequential often occurred simultaneously. Moreover, communication between those who recognized a need and those who were capable of generating ideas in answer to the need was often quite informal and not well organized. In fact, informal personal communication often pre-empted the exchange of formal reports or documents. Max Tishler (4), President of Merck, Sharp and Dohme, a pharmaceutical research enterprise, makes a similar point that often a university researcher comes to Merck, Sharp and Dohme seeking an answer to a question which he could have obtained on his own campus, sometimes on the floor below his own laboratory.

The Arthur D. Little study also pointed out that success in pushing an original idea from the research stage to actual utilization often

depended upon having the same people and the same management involved at all stages. As president of a newly formed nonprofit organization, I was particularly impressed by the further finding that the funds which launched an important event were most often discretionary, rather than specific-project, funds. Finally, it was found that an adaptive rather than an authoritarian organization made for the best environment for innovation.

While the A. D. Little study involved the Department of Defense and professionals who were engineers and physical scientists, it is possible that there are implications for education. The question then does not concern in whose hands innovation rests, but the circumstances, the organization, and the climate in which innovation is likely to prosper.

A further relevant consideration with which I find myself in great sympathy was emphasized by Emmanuel G. Mesthene (3) of Harvard University in his concluding remarks at last summer's American Management Association Conference:

> The most fundamental obstacle to achievement of the necessary updating of the enterprise of education is our failure as yet to recognize the full implications of the new tools, educational and otherwise, that our technology gives us. There is a tendency to think of a tool as a better way to do a known job. Yet the meaning of tools and technology throughout the ages has been that they have changed the job by making new things possible . . . If we see the future exclusively in terms of old values, we will sell the future short. For the values of society are determined importantly by the tools of the society. That is why one has to search his tool box carefully. There are unsuspected possibilities in it that a bit of craftsmanship may well fashion into greater values still.

There is one more point which needs to be made. For all of the emphasis we have heard in the past on the subject of individual differences, much of what transpires in our schools is still designed primarily with the large middle group of students in mind. I mean "middle" in every sense of the word, not just economic. Innovation that proceeds on the premise that the middle group is our only, or our main, concern will serve only to widen the gap between the economically disadvantaged and the culturally deprived on the one hand, representing perhaps one-third of our total school population, and the rest of our educational society on the other. The teacher must be persuaded that this one-third of our school population represents a promising *intellectual* market just as the commercial producer needs

to be persuaded that it represents a promising *economic* market.

Innovation and evaluation—in whose hands? Obviously, they must be in every competent and qualified person's hands. My plea or my hope is simply that the hands of industry, of education, and of government will work together, for only by the joining of such hands can the best interests of society and of our schools be served.

REFERENCES

1. Carter, L. From research to development to use. Address delivered to the American Educational Research Association. February 1966.
2. Howe, H. II. Address delivered to the American Management Association's Second International Conference and Exhibit. August 1966.
3. Mesthene, E. G. Values and education. Address delivered to the American Management Association's Second International Conference and Exhibit. August 1966.
4. Campbell, T. L. Reflections on research and the future of medicine. *Science*, July 22, 1966, Vol. 153, No. 3734, 442-449.

The Discovery and Development of Educational Goals

HENRY S. DYER
Educational Testing Service

Since World War II most professional philosophers, with some notable exceptions, have backed away from rows over the goals of education and have stuck more or less consistently to analyzing the absurdities in all such forms of discourse (14). Before the philosophical silence set in, however, practically every major philosopher, from Confucius and Plato and Aristotle down to Whitehead and Russell and Dewey, had had a good deal to say about the aims of education and its functions in society. Since then there has been an increasing volume of writing on the subject by eminent non-philosophers inside and outside the academic community. No less than two Presidential Commissions have taken a crack at the problem (10, 20), and their efforts have been supplemented and extended by such documents as the Harvard report on objectives of general education (13), the Russell Sage reports on elementary and secondary school objectives (9, 15), and the two taxonomies by Benjamin Bloom and his collaborators (3, 17).

One would think that the accumulation of so much high-level verbiage on the subject of goals over at least two and one-half millenia would have exhausted the subject if not the discussants. One would suppose that by now the question of educational goals would have been fairly well settled, and the problem of how to define them would have found some useful answers. But the question is still very much open. The problem of goals is today, more than ever, a top-priority, and largely unsolved, problem. It is symptomatic that a recent book on the preparation of instructional objectives (11) starts off with an echo from Charles Dudley Warner's famous remark about the weather: "Everybody talks about defining educational objectives, but almost nobody does anything about it."

The trouble is that in spite of all the hard thinking and earnest talk about educational goals and how to define them, the goals produced have been essentially nonfunctional—and I mean even when they have come clothed in the so-called behavioral terms we so much admire. They have had little or no effect on the deals and deliberations that go on in faculties and school boards and boards of trustees and legislative chambers where the little and big decisions about education are being made. As you watch the educational enterprise going through its interminable routines, it is hard to avoid the impression that the whole affair is mostly a complicated ritual in which the vast majority of participants—pupils, teachers, administrators, policy makers—have never given a thought to the question *why*, in any fundamental sense, they are going through the motions they think of as education. In spite of the tardy recognition in a few quarters that there are some ugly situations in the schools of the urban ghettos and rural slums, the general attitude still seems to be that if we are spending 50 billion dollars a year on the education of 50 million children, and if over 40 percent of them are now getting to go to college, as compared with less than 20 percent a few years back, then "we must be doing something right," even though we haven't the remotest idea of what it is. This blind faith in quantity as proof of quality is precisely the faith that, in the long run, could be our undoing.

Perhaps in a simpler age a disjunction between educational purpose and educational practice was tolerable. A hundred years ago, such a small part of the population went to school that the opportunities open to educators for inadvertently damaging the lives and minds of the generality of mankind were neither potent nor pervasive. The situation today, as the headlines hardly permit us to forget, is somewhat different. We have more knowledge than we know what to do with, more people than we know how to live with, more physical energy than we know how to cope with, and, in all things, a faster rate of change than we know how to keep up with. So we dump the problem on the schools and hope that somehow they can program the oncoming generation for the unforeseeable complexities of the twenty-first century, now less than 34 years away.

Henry Adams (1, p. 496), as far back as 1905, had already figured out what we would be up against. As he saw it then, "Every American who lived into the year 2000 would know how to control unlimited power. He would think in complexities unimaginable to an earlier mind." This being 1967 rather than 1905, the near prospect of

unlimited power in the hands of every American (and European and Asian and African) has finally scared us into a rash of educational innovations that we hope will help the oncoming generation "think in complexities unimaginable" to *us*. But the rising curve of proposed innovations itself is adding to the burden of our complexities by swamping the schools with more untested devices, strategies, administrative arrangements, and curricular materials than those who run the educational system are prepared to absorb or evaluate. This is why it is more important than ever to reconsider the problem of goals. Somehow we have to arrive at goals that are so clear and compelling that the movers and shapers of education can and will use them in deciding on the tradeoffs that are going to have to be made if the system is to be kept from stalling under the mounting load of new ideas and conflicting demands.

II

Why is it that the goals formulated in the past—even the recent past—have been largely nonfunctional? I think there are three principal reasons: too much reliance on the magic of words, too little public participation in formulating the goals, and too great a readiness to suppose that the goals are already given and require only to be achieved.

In the 1947 report of the President's Commission on Higher Education (20, p. 9), there is the following paragraph:

> The first goal in education for democracy is the full, rounded, and continuing development of the person. The discovery, training, and utilization of individual talents is of fundamental importance in a free society. To liberate and perfect the intrinsic powers of every citizen is the central purpose of democracy, and its furtherance of individual self-realization is its greatest glory.

This is an example of word-magic. It is an expression of an ideal to which presumably the great majority of Americans would enthusiastically give verbal assent, *without having the foggiest notion of what*

the words are saying. And this failure is not to be chalked up as a flaw in the thinking of the American people. For it is no mean task for anybody, however sophisticated in words and their ways, to translate into specifiable operations such metaphoric expressions as "full, rounded, and continuing development of the person" or "liberate and perfect the intrinsic powers of every citizen." Phrases like these sing to our enthusiasms, but they don't tell us what to do about them. The difficulty is that the metaphors in which they are couched are extremely hard to translate in terms of what little we really know of human growth and functioning. How do you know, for instance, when you have liberated and perfected the intrinsic powers of a citizen? Or how do you calibrate the roundedness of his development?

To ask such questions is to suggest why the word-magic has not worked and why such goal statements leave school people with barely a clue for determining what the lines of progress ought to be or whether the system is making any headway in the desired directions. And this failure has led to more than a little disillusionment about the practical utility of any kind of goal statements and to a considerable degree of offhand cynicism about pious platitudes that have no relevance for practical operations beyond that of providing useful window dressing to keep the public happy.

A second reason that the usual statements of goals fail to function is that there has not been enough genuine participation by the public in the goal-making process. The typical approach to working out educational objectives for pupils or schools or school systems is for a group of educators or academicians or psychometricians or some mixture of these to hole up and bring their combined expertise to bear on working out what *they* think should happen to people as a consequence of going to school. In the presentation of their findings they have occasionally involved representatives of the citizenry at large, but this wider involvement has been usually little more than a series of gestures aimed at getting acceptance rather than participation. The result, again, is usually assent without understanding, and the goals produced turn out to be a dead letter.

The approach of the experts is back-end-to. It should not be one of trying to convince the public of what it *ought* to want from its schools but of helping the public to discover what it *really* wants; and among the public I include those who will be in charge in the next 15 years or so—namely, the pupils themselves, as well as their teachers, their parents, their prospective employers, and behind all these, the school

boards and legislators who make the ultimate decisions.* This is partly what I mean by the discovery and development of educational goals. By its nature this process of discovery will be necessarily tedious and often frustrating and, most important, never-ending. So far as I know, it has never been given a serious trial on any broad or continuous basis to the point where the actual needs and desires of individuals and of society become the determiners of such subsidiary matters as whether school budgets are to be voted up or down, whether school districts will be consolidated, or kindergartens shall become mandatory, or whether a foreign language shall be taught to all children or some children or no children at all in the third grade.

It is easy to dismiss this idea, the idea of the public search for goals, as utopian. How can one possibly bring about genuine public involvement in the goal-making process or expect that anything really useful will come of it when everybody knows that 90 percent of what happens in and to the schools is determined by the power blocs and pressure groups and influence agents whose prime interest is keeping taxes down, or getting bus contracts, or simply gathering in the symbols that add up to prestige and power for their own sake? Nevertheless, in an essay on "Who Controls the Schools?" Neal Gross (12), who has looked these hard realities square in the eye, can still make the hopeful observation that:

> The control is ultimately, of course, in the hands of the people. If they really want it, they can have it any time, since it is they, after all, who elect the school boards.

The problem is to get them to take control and to know what they want their schools to deliver. The chances of a solution will be much improved when the experts stop talking exclusively to themselves and broaden their conversations to include the public.

The third reason that educational goals have been nonfunctional is that too frequently they have been assumed as, in some sense, already given, and the only problem has been to figure out how to attain them. This assumption is as old as Plato and as recent as Clark Kerr. According to Plato (19), the reason the guardians of the state must study geometry is that it forces "the soul to turn its vision round to

*The goal-making efforts of the State Board of Education and interested citizen groups in the Commonwealth of Pennsylvania suggest the practical possibilities. See *A Plan for Evaluating the Quality of Educational Programs in Pennsylvania* (Harrisburg, Pa.: State Board of Education, 1965) Vol. I, pp. 1-4; pp. 10-12; and Vol. II, pp. 158-161.

the region where dwells the most blessed part of reality . . . for geometry is the knowledge of the eternally existent." Clark Kerr's brief comment on the purposes of a university (16, p. 38) is in the same vein:

> The ends are already given—the preservation of the eternal truths, the creation of new knowledge, the improvement of service wherever truth and knowledge of high order may serve the needs of man.

Interestingly enough it was Aristotle (2) who wondered whether things were all that simple. He recognized that there could be diversity of opinion in these matters:

> Confusing questions arise out of the education that actually prevails, and it is not at all clear whether the pupils should practise pursuits that are practically useful or morally edifying, or higher accomplishments—for all these views have won the support of some judges, and nothing is agreed as regards the exercise conducive to virtue, for, to start with, all men do not honour the same virtue, so that they naturally hold different opinions in regard to training in virtue.

The fact that "all men do not honour the same virtue" is precisely what makes the structuring and conduct of education in a free society so complicated and frequently so frustrating. If schools are to keep at all, they must somehow accommodate themselves to the pluralism in the values of those whom they serve and from whom they derive their support. Any system that tries to operate on the assumption that there is one fixed set of goals to which all people must aspire is bound to be so far out of touch with the actualities of the human condition that such effects as the schools may have are likely to be altogether unrelated to the needs of the pupils in them or to the society they are expected to serve.

Each individual and each generation has to create its own truth by which to know the world of its own time and place, and, by the same token, it has to create its own goals for ordering its efforts to cope with its world. Thus, the discovery and development of educational goals has to be part of the educational process itself, starting with the child and continuing with the adult as he works his way through to the personal, social, and economic decisions that determine the shape of the free world he is to live in. This, as I understand him, is what John Dewey (7, p. 71) had in mind when he said that "freedom resides in the operations of intelligent observation and judgment by which a purpose is developed." He was thinking in this particular

instance of the child in the classroom trying to find goals that make sense for him, but the principle applies with equal force to such adult groups as school boards, where there is no authoritarian teacher hovering in the background ready to pounce in favor of the eternal verities; only superintendents, curriculum experts, and others who are equally sure they have all the answers.

I realize that there can be profound disagreement with this relativistic conception of educational goals, but I think it is time we stopped kidding ourselves that the misty absolutes we have inherited from the ancients can serve to unravel the ambiguities in education that are inescapable in our half of the twentieth century.

There is an inevitable dilemma in the business of goal making that has to be faced candidly if we are going to make any headway in the process. On the one hand, as we have been saying for decades, we require goals that specify definite performance levels for pupils as they move through and out of the schools, so that we can gauge how the educational system is doing in its attempts to help them deal with the occupational, social, cultural, and moral demands of the world they are to enter. On the other hand, it is impossible to predict with much certainty anymore what the world is going to be like in 15 or 20 years when the children now in elementary school will be taking over the social controls. Margaret Mead put the problem succinctly a few years ago (18). She said:

> If we can't teach every student . . . something we don't know in some form, we haven't a hope of educating the next generation, because what they are going to need is what we don't know.

The easy answer to this problem is that instead of teaching youngsters the substance of what they will need to know, we must teach them the "process of discovery" and express our goals in terms of the mastery of that, and its close relatives flexibility, tolerance for ambiguity, adjustment to the environment, and the like. The danger is that we can still get caught in the word-magic. We can be too quickly satisfied that we know what we mean by the terms before we have worked out any more than a few "for instances" of the operations they might actually entail.

III

What is the way out? And what is the role of educational measurement in the search for educational goals?

I think the way out is to hold the search for long-term goals in abeyance for awhile and concentrate on getting a clearer idea of what is happening in the schools right now and making up our minds about how much we like what we see.

Every morning, Monday through Friday, 50 million children leave 18 million homes and are funneled into 120 thousand schoolhouses where they have an uncountable number of experiences affecting their thoughts, feelings, aspirations, physical well-being, personal relations, and general conception of how the world is put together. The extraordinary fact is, however, that in spite of the mountains of data that have been piled up from teachers' reports, tests, questionnaires, and demographic records of all kinds, we still have only very hazy and superficial notions of what the effects of the school experience actually are.

There are some things we are beginning to suspect that leave us more or less comfortable—mostly less. For instance, all but a very few children learn to read, at least up to the point where most of them can and do enjoy comic strips.* It has been estimated that by the time students reach college, half of them will admit to some form of academic dishonesty (4, p. 64), but the grade norms for this form of academic achievement are not yet known. According to the Project TALENT data, the career plans most students make in high school are unrealistic and unstable (8, p. 179), but nobody knows for sure whether this situation is good or bad or how far the schools can or should be held accountable for it. In elementary school, according to the recent Educational Opportunities Survey by the Office of Education (6, p. 199), 10 percent of white children and 18 percent of Negro children have acquired an attitude that prompts them to agree with the proposition: "People like me don't have much of a chance to be successful in life;" and in high school 15 percent of whites and 19 percent of Negroes say they have reached the conclu-

*A poll by the American Institute of Public Opinion, released February 20, 1963, estimated that 50 million adults (45 percent of the adult population) read comic strips. As a "cultural" diversion this activity ranked second in popularity to watching westerns on television.

sion: "Everytime I try to get ahead, something or somebody stops me." To what extent can attitudes like these be attributed to school experiences and how much to the education supplied by the city streets? Again, we don't know, but information of this sort seems indispensable to the process of arriving at educational goals and deciding on priorities among them.

The point I am trying to make is very simply this: People are more likely to get clear in their minds what the outcomes of education *ought* to be if they can first get clear in their minds what the outcomes actually *are*. To know that a considerable number of pupils are learning to cheat on examinations or learning that the cards are stacked against them should help to suggest, if only in a negative way, what educational outcomes are to be preferred.

It has been customary to take the view that before one can develop measures of educational outcomes, one must determine what the objectives of education are. What I am suggesting is that it is not possible to determine the objectives until one has measured the outcomes. This sounds more like a paradox than it really is. Evaluating the *side* effects of an educational program may be even more important than evaluating its intended effects. An up-to-date math teacher may be trying to teach set theory to fourth graders and may be doing a good job at it, but one wants to know whether he is also teaching some of the youngsters to despise mathematics.

In a recent essay on "Education as a Social Invention," Jerome Bruner (5) makes the point that "however able psychologists may be, it is not their function to decide upon educational goals," but it *is* their function to be "diviner(s) and delineator(s) of the possible." And he goes on to say that if a psychologist "confuses his function and narrows his vision of the possible to what he counts as desirable, then we shall all be the poorer. He can and must provide the full range of alternatives to challenge society to choice."

The same argument holds with equal if not greater force for the educational tester who is intent on doing his full duty to society. He must provide instruments and procedures for displaying and accurately ordering as many of the behavioral outcomes of the educational process as he, with the help of everybody involved, can imagine, regardless of whether these outcomes are to be judged good or bad, helpful or harmful, desirable or undesirable. The educational tester must not allow his thinking to become trapped in the traditional

categories of the curriculum such as English, mathematics, and science; he must be concerned with the whole spectrum of human behavior as possible outputs of the educational process and he must try to find ways of categorizing it and measuring it that will make sense to the general public that decides on what schools are for.

In the *Taxonomy of Educational Objectives, Handbook II: Affective Domain*, David Krathwohl and his collaborators have made an enormous contribution to this effort if for no other reason than that they insist one must attend to human functioning beyond the cognitive. Their focus, however, is on "classifying and ordering responses specified as *desired* outcomes of education" (17, p. 4). What is now required, it seems to me, is a taxonomy of all possible educational outcomes without reference to whether they are desirable or undesirable, good or bad, hurtful or helpful.* Only as this requirement is met are we likely to approximate testing programs that will begin to tell us all we need to know for evaluating educational programs.

Any achievement testing program that is limited to measuring performance in the basic skills and mastery of academic subject matter—and this, I suspect, is the pattern of most such programs—is almost certain to do more harm than good by not raising the question whether excellence in performance in such things as reading and mathematics and science and literature is not being bought at the expense of something left unmeasured, such as academic honesty and individual sense of self-worth. Granted the tremendous importance of mastery of the basic intellectual tools for these times, it seems axiomatic that they hardly compare in importance with common honesty and mutual trust as the indispensable ingredients of a viable free society.

It is easy to argue that the present state of the art leaves much to be desired in the measurement of the affective and social outcomes of the educational system. It is easy to argue that such instruments as we have for these purposes are productive of soft data, full of superficialities and pitfalls that can lead people astray in assessing what the educational system is really doing to students. This is all too true, and anyone with a conscience rooted in sound measurement knows it only too well. But such arguments only point to the need for firming up the soft data by going after the correlates of behavior that get

*Krathwohl and his collaborators hint at this possibility in a footnote on page 30.

beneath the semantic confusions inherent in self-report devices.* They also point to the need for keeping a spotlight on the limitations of the data we have, when, for want of anything better, such data have to be consulted. Not to consult them at all is to keep our eyes shut to many of the products of schooling that most need attention.

Finally, educational measurement has its uses not only in the discovery, but also in the development, of the goals. In an ideal world, this developmental process is a continual series of approximations—an unending iterative process for constantly checking the validity of concepts against the behavior of the measures derived from them, and checking the validity of the measures against the concepts from which they have been derived. This back-and-forth process begins in the vague concerns of the public for what it wants but has not defined—personal fulfillment, effective citizenship, the good life, the open society, and so on. All of which terms are still word-magic. They are no good in themselves as goals. But as symbols of human hope, they cannot be neglected in the *search* for goals. They have an extremely high heuristic value in getting the search started. The first practical approximation in the search, however, is some combination of tests and other measures that can begin to delineate, for all to see, the dimensions along which we think we want to progress. This is to say that, in the last analysis, an educational goal is adequately defined only in terms of the agreed-upon procedures and instruments by which its attainment is to be measured. It is to say that the development of educational goals is practically identical with the process by which we develop educational tests. It is to imply what in some quarters might be regarded as the ultimate in educational heresy: *teaching should be pointed very specifically at the tests the students will take as measures of output;* otherwise, neither the students nor their teachers are ever likely to discover where they are going or whether they are getting anywhere at all.

A great problem—probably the greatest problem—in the development of meaningful goals is that of making sure that the tangible tests that come out of the process bear a determinable relationship to

*See, for instance, the approach taken by Sears and Sherman in their case studies of self-esteem: Pauline S. Sears and Vivian S. Sherman, *In Pursuit of Self-Esteem* (Belmont, California: Wadsworth Publishing Company, 1965); also the approach of Sandra Cohen in her study of the attitudes of primary school children to school and learning: "An Exploratory Study of Student Attitudes in the Primary Grades," in *A Plan for Evaluating the Quality of Educational Programs in Pennsylvania*, Volume II, pp. 61-130.

all the vague individual and collective concerns that go into it. The only way this relationship can be assured is through some sort of continuous dialogue among testers, students, educators, and the public bodies that control the educational enterprise. As anyone who has tried it knows, this is not an easy dialogue to get going or keep going in fruitful directions, but without it there is small likelihood that anyone will be able to figure out where American education is, or where it ought to be headed, or how it must tool up to get there.

Educational measurement, in the full sense of the term, is one field in which insulation of the experts is intolerable, for measurement in education is the only process by which a society can externalize and give effect to its hopes for the next generation.

REFERENCES

1. Adams, H. *The education of Henry Adams: an autobiography.* Boston: Houghton Mifflin Company, 1930.
2. Aristotle. *Politics,* book vii, paragraph i.
3. Bloom, B. S. (Ed.) *Taxonomy of educational objectives, handbook i: cognitive domain.* New York: Longmans, Green and Co., 1956.
4. Bowens, W. J. Student dishonesty and its control in college. New York: Bureau of Applied Social Research, Columbia University, 1964. Mimeographed.
5. Bruner, J. S. *Toward a theory of instruction.* Cambridge: Harvard University Press, 1966. Pp. 23-24.
6. Coleman, J. S. *et al. Equality of educational opportunity.* Washington, D. C.: U. S. Office of Education, 1966.
7. Dewey, J. *Experience and education.* New York: Collier Books, 1963.
8. Flanagan, J. C. and Cooley, W. W. *Project Talent one-year follow-up studies.* Pittsburgh: University of Pittsburgh, 1966.
9. French, W. & Associates. *Behavioral goals of general education in high school.* New York: Russell Sage Foundation, 1957.
10. Gardner, J. W. National goals in education. In The President's Commission on National Goals, *Goals for Americans.* New York: Prentice Hall, Inc., 1960. Pp. 81-100.

11. Gilpin, J. B. In the foreword to R. F. Mager, *Preparing objectives for programmed instruction.* San Francisco: Fearon Publishers, Inc., 1962.

12. Gross, N. Who controls the schools? In S. E. Harris (Ed.) *Education and public policy.* Berkeley: McCutchan Publishing Corporation, 1965. Pp. 19-29.

13. Harvard University Committee on the Objectives of General Education in a Free Society. *General education in a free society.* Cambridge: Harvard University, 1945.

14. Kaufmann, W. Educational development from the point of view of a normative philosophy. Harvard Educational Review, summer 1966. 247-264.

15. Kearney, N. C. *Elementary school objectives.* New York: Russell Sage Foundation, 1953.

16. Kerr, C. *The uses of the university.* Cambridge: Harvard University Press, 1963.

17. Krathwohl, D. R., *et al. Taxonomy of educational objectives, handbook ii: affective domain.* New York: David McKay Company, Inc., 1964.

18. Mead, M. Changing teachers in a changing world. In *The education of teachers: new perspectives*, Official Report of the Second Bowling Green Conference. Washington, D. C.: National Education Association, 1958. Pp. 121-134.

19. Plato. *The republic*, book vii, paragraph ix.

20. The President's Commission on Higher Education. *Higher education for American democracy*, volume i, *Establishing the goals.* Washington, D. C.: U. S. Government Printing Office, 1947.

The Meaning of Impact

MARTIN TROW
University of California, Berkeley

I want to amend the title of this paper to read "*Some* Meanings of Impact." For there are, of course, many meanings, as many as there are ways in which higher education affects individuals, other institutions, and the larger society. But if we confine ourselves for the moment to the supposed effects of higher education on students who experience it, we may usefully distinguish three broad kinds of outcomes:

First, the skills and knowledge acquired which closely reflect the manifest intention of the curriculum and syllabus.

Second, changes in a wide range of attitudes, values, orientations, and aspects of personality which occur over the course of the years in college and to which the college experience itself contributes.

Third, certain attitudes, behaviors, and styles of thought and action among adults who have been to college, which are of importance for the quality of life in the society, and which we may reasonably believe to have been affected by some aspect of their experience of higher education. Here we are speaking of the long-range influence of college over the individual's whole lifetime.

Although tests that attempt to measure changes in skills and knowledge are as old as formal education, and studies of changes in other characteristics of students during their college years currently make up a thriving research industry, studies of the long-range effects of college experience are still rare. Yet this long-range effect is the kind of impact that is ultimately of greatest interest to the educator and researcher. It provides the criteria against which, in principle, we would want to evaluate higher education.

There is yet another meaning of impact which I will have in mind during these remarks: not the impact of the college experience on students and graduates, but rather the impact of mass higher education on American society. And I will be discussing the first kind of impact—of institutions on individuals—chiefly for what it can tell us about the role of higher education in contributing to certain kinds of moral and intellectual resources of adult citizens in the peculiar society that is emerging on this continent. The qualities I mean to discuss are (1) a sense of personal effectiveness in social action; (2) certain civic virtues that we might call "civic responsibility;" and (3) the capacity to learn and adapt to new circumstances throughout the adult career.

The effects I am referring to have been relatively little studied by social scientists for a number of reasons. For one thing, it is difficult to devise good, reliable, and economical measures of them. For another, the researcher must wait a long time for them to appear. Higher education may strengthen, or even create, these qualities, but only as *potentialities*; they show themselves, at least in the forms that interest us most, only much later, and under circumstances which themselves are variable and difficult to predict. And finally, it is extremely difficult, in studying them, to disentangle the role of the individual's experience in higher education from all the other influences, prior to, after, and even during the college years, which are in varying senses of the word "independent" of the specific experience in college or university.

The justification for speaking of these "outcomes" (and thus indirectly of these kinds of "impact") of higher education, despite the difficulties of studying them systematically and with precision, lies in the dual fact that they are, on one hand, among those outcomes that educators themselves are most concerned to achieve, while on the other hand they are, quite apart from the intentions of educators, qualities that heavily affect individual lives and, in their aggregate, the character of the society in which those lives are lived. Let us look at these qualities a bit more closely.

One of the gains of higher education is an increased belief in one's own capacities to handle broad responsibilities, contribute to the solution of important problems, have an impact on the larger society. Relatively uneducated people tend to have a much narrower conception of their range of effective action. We cannot say, as we could of the medieval peasant or can of most men in traditional societies

today, that their horizons are bound by family, village, or work group. The mass media and other institutions ensure that almost all Americans are constantly exposed to the idea and the image of distant places and remote issues. But to consume passively the news of the great world beyond one's own personal experience is very different from feeling able to affect and participate in those events. To most people in America today, as to most people in most times and places in history, society and its institutions resemble the world of nature as it appears to primitive man—something to which one adapts or defends oneself against, but cannot significantly alter. But while this is true for most people, in varying degrees, it is less true for men and women who have had some exposure to higher education. Higher education is one, and an increasingly important, aspect of what Max Weber spoke of as the continuing process of rationalization in all spheres of life—the tendency to find logical and coherent patterns in the flux of events. Today we tend to seek those linkages of cause and effect that are congruent with empirical evidence. Higher education, with, as we know, quite varying degrees of effectiveness, is to a considerable degree devoted to cultivating the capacity to make such linkages and moreover—especially in the social sciences—to communicating what we think we know about social institutions and the nature and levers of social change. It also, I suggest, plants or nurtures this still rare and fragile notion that an individual can significantly affect events. The contribution of higher education to men's capacity to understand the relation of cause and effect in social life is only one, though an important, part of its contribution to the individual's sense of himself as the kind of person who can intervene to shape the course of events beyond the boundaries of his immediate milieu.

Students of political behavior, such as Angus Campbell and his associates, have studied one aspect of this self-assurance in the form of a quality they call "feelings of political efficacy," which they find to be strongly related to formal education. But this sense of the ability to affect political events is one facet of a more general feeling of potential effectiveness which shows itself in how men feel about their ability to affect the behavior of other institutions in which they are involved. And this broader sense of effectiveness, like the sense of political efficacy which is the aspect that has been most closely studied, is, I believe, strongly associated with having had some experience of higher education.

If we accept that this sense of personal competence and effective-

ness in social action is somehow influenced by experience in college and university (and that, of course, is an empirical question), we might well ask what kinds or aspects of that experience have this effect. I would suggest, for investigation, that the experience of effectiveness, and especially of distinction, during the college years may enhance this more generalized sense of competence. To be admitted to an honors program, to gain the personal attention and encouragement of an admired teacher, to earn a degree with distinction—these may all have something of the character of self-fulfilling prophecies. The young men and women who gain such distinction are, almost by definition, already more than usually effective people. But the rewards and distinction they gain may themselves enhance their sense of their capacity to deal with large affairs competently and successfully.

These rewards may take the form of academic distinction; they may, much more fundamentally, center on the student's experience of being distinguished by his teachers from everyone else, of having, for some of them anyway, a distinct face, name, voice, and certain unique qualities—above all, the quality of uniqueness. Some institutions are much more sparing of these rewards than others, quite independently of the objective qualities of their students. For example, those rewards are conspicuously rare under the conditions of mass impersonal processing of students that sadly has come to be thought of as "The Berkeley Syndrome." The real implications of that way of organizing undergraduate education may well lie in its short- and long-run effects on the student's conception of himself and his own capacities rather than in its ability to transmit skills and knowledge.

Attendance at a college or university of recognized distinction may provide something of this enhancement of self-regard that elsewhere accrues to the small minority who achieve distinction. The interplay between personal and institutional distinction in the shaping of self-concepts is a fascinating problem. For example, the selection to M.I.T., as we all know, is extremely severe; only students of very high achievement and aptitude gain admittance. Most of those prize winners and high school valedictorians get a rude shock in their first few weeks at M.I.T. when they discover that in the land of the highly gifted they are, for the most part, only mediocre. Students who never got less than an *A* in high school suddenly find themselves flunking exams and earning *C*s and *D*s. That, as I suggested, is a severe shock to their self-conceptions and a source of stress and painful personal reassessment. But over four years, many of these "mediocre" students

appear to regain a large measure of self-confidence, in part, I think, through a process that involves a kind of borrowing of the prestige and distinction of the institution they attend. So finally, even to graduate with a quite ordinary record, as necessarily most of them must, is felt itself to be a mark of distinction. Moreover, it is not just the public reputation of the institute that makes the difference in how they feel about themselves. They have been taught by leaders in their fields; they have been addressed by prominent men in public, professional, and academic life; they have been told, overtly and implicitly, that they are, as a body, very unusual and talented young men and women. They come thus to feel that they are an élite, and part of a larger élite, and that large things are expected of them.

The process by which young men and women gain these feelings of potential effectiveness in college is difficult to study empirically. Moreover, we are interested in how these feelings affect what they do with their lives; and we can see the problems of separating the influence of their own high abilities, the sheer technical qualities of their education, and the advantages which a degree from a distinguished college or university gives to men in many fields of endeavor, from the sense of potential effectiveness and personal capacities gained in the course of attendance at an élite institution.

Moreover, we Americans are a bit shy about studying the processes of élite formation; though we recognize their existence, somehow our egalitarian values tend to direct our attention to less invidious subjects. But I suggest that in the formation of intellectual and cultural élites we may see processes which are similar to those at work in less clearly visible form throughout our system of higher education. For in one sense what we are doing in this country is to expand greatly the number and variety of élites, and through our system of mass higher education, to extend greatly the distribution of qualities which heretofore have characterized, and elsewhere still do characterize, relatively small élites. It is not difficult to understand how young men from the English upper classes who pass through Eton and Oxford come to feel that they have special talents for leadership—everything in their life experience, much of it by design, has served to strengthen those convictions. But it is a more subtle and difficult matter to ask this question about the millions of young Americans from the broad middle and lower-middle and working classes who attend state colleges and universities or less well-known private colleges. Few of them emerge with the unquestioned, if gracefully borne, assump-

tions of superiority of the product of Eton and Oxford or its continental counterparts—and perhaps that is just as well. But I suggest that our popular institutions do transmit to many who pass through them some sense of competence, if not superiority; some sense, to put it in negative terms, that one is at least not disqualified by limitations of talent and training from taking large responsibilities or from making important contributions in the larger society. The feeling that "I am as good as the next man," the rejection of claims to personal superiority, reflect, of course, old and very strong egalitarian values in American life. What mass higher education is doing, more or less well, is to give substance to these claims, to make of them more than the empty barroom boast. And perhaps it is just because of our tendency to deny claims to superiority, our lack of deference, in public life as in personal relations, that the wide diffusion of élite characteristics is so important to the quality of American life. When there is a tradition of deference to traditional élites, it is the training of the élites that is the crucial question. But where deference is denied, as in a populist democracy, it is especially important that the qualities of élites be widely diffused in the population. And the crucial question then becomes the quality of our mass higher education.

I would like to speak more briefly of two other presumptive effects of the college experience. One of these, about which we have a good deal of evidence, is the possession of various civic virtues, of attitudes and orientations and behaviors appropriate to the functioning of a democratic political order in a complex and heterogeneous society. Among these is a readiness to take part in political life, and in voluntary associations devoted to improving the natural and human environment through education, conservation, pollution control, rapid transit and city planning, and the like. The relation of education to participation in both politics and voluntary associations is strong and well-documented. We know even more about the relation of formal education to another civic virtue: the readiness to support or at least to tolerate the exercise of their civil rights and liberties by unpopular and despised minorities. I cannot report this literature in detail; besides, much of it is, I am sure, familiar to you. In studies by Herbert Hyman, for example, of attitudes toward racial integration in the general population, and in the now classic studies of Samuel Stauffer on attitudes toward civil liberties, among many others, we find educated people distinctly more likely to hold tolerant views; moreover,

these differences hold up when we control for social class and occupation, geography, age, ethnicity, and other plausible factors. Further, in a number of studies we find differences in these regards between freshmen and seniors enrolled in the same college. And in some as yet unpublished studies done by Burton Clark, Paul Heist, and myself, we have followed students at eight very different kinds of colleges and universities right through their college years, and have been able to observe at least some aspects of this liberalizing process occurring. In all these institutions, despite the wide variation in their size and character, we find students becoming, in time, more likely to hold tolerant and libertarian views—though, interestingly, the largest gains are found in those colleges whose students were most liberal on entry. But even in the most conservative institutions, the students tended to move in the same direction, though not quite so uniformly nor so far. Of course, there are students who become less tolerant during their college years, and I suspect we could find institutions where they outnumber the students moving in the other direction. But the enormous diversity within and among our nearly 2,000 colleges and universities should not obscure their broad common characteristics, among which is this pervasive liberalizing influence. Variations in the form and strength of these influences are very great; nevertheless, I suspect the liberalizing influences of most colleges reflect their use of the leading institutions as models; the influence of college faculties, which are in these respects much more alike across institutions than are their students; and of course, the intrinsic character of colleges and universities as institutions devoted to reason and the pursuit of understanding, values which on one hand tend to undermine racial prejudice and on the other tend to support the rule of law and its due processes.

There is one further characteristic of educated men which shows itself over the whole course of their adult lives and which I believe also reflects the specific influence of their college years. This is the capacity to learn new skills, to take on new tasks and responsibilities throughout life. In the world of work this shows itself as a flexibility, an adaptability to new circumstances, jobs, and opportunities. At the lower end of the occupational structure we know of the difficulties governmental and private agencies have of retraining poorly educated men who have lost not just their jobs but their occupations as a result of some change in production techniques or consumption patterns. We pay less attention to, because we take for granted, the contrasting high capacity of educated men to change their patterns of

work when they change jobs, when jobs change under them, or when they themselves change the nature and content of their own jobs. My colleague Harold Wilensky has observed that "over his worklife the average man holds a dozen or more jobs, most of them related neither in function or status," while "half our young people will one day hold jobs not now in being." This rapid and continuous transformation of the occupational structure, under the spur of technological, organizational, and cultural changes, is in large measure made possible by the flexibility and adaptability in the labor force that I am speaking of. But it also makes that quality, in its most general form as the capacity to learn throughout life, perhaps the most valuable of the skills acquired in formal schooling. It is this capacity to learn and to adapt to new circumstances that distinguishes the beneficiaries from the victims of rapid social and economic change.

Confidence in one's capacity to affect the social environment and the ability to respond flexibly and sensitively to its changing requirements and opportunities are both individual characteristics: They are the old liberal virtues of self-reliance and self-help adapted to the requirements of a society of large organizations and rapid social change. Mass higher education, by producing very large numbers of people who are prepared to operate large organizations in a changing environment, thus itself shapes the character of the society for which it prepares its graduates. The classic picture of large bureaucratic organization was of a series of offices, hierarchically graded, governed by formal rules and routinized procedures. Not even the post office looks like that today. What we see more commonly are industrial, educational, and governmental agencies undergoing a more or less continuous process of internal reorganization, defining or adapting themselves to new functions, devising new modes of operation, perennially and often somewhat anxiously seeking for new ideas about how to deal with new problems, or old problems in new guises. What Fritz Machlup has called "the knowledge industry" absorbs a large and rapidly growing proportion of the labor force, and just that segment of the labor force that includes the largest proportion of college-educated people. In the most advanced sectors of the economy there is no shortage of new problems, nor, in many cases, of material resources for meeting them. But there is, despite the enormous output of American higher education, a chronic shortage of imagination and initiative. The society we are shaping rewards the qualities I have been describing, as it also punishes their absence with low pay, hard

work, job insecurity, early obsolescence, redundancy, and unemployment. This alone would account for the continually rising proportions of the population gaining some experience of higher education.

Let me close, in this conference on research, with a question for research. I have not meant to celebrate the myth of our national genius, nor to imply that the problems of war, racial injustice, and poverty can be solved by the magic of higher education. Yet it would be wrong to be mesmerized by our problems to the neglect of our national resources, not least among which are the human qualities of feelings of competence, civility, and the ability to learn, which I suggest are strengthened by experience in higher education. But obviously colleges vary greatly in their power to shape these qualities and in the kinds of students whom they are able to reach in these ways. Moreover, it is not at all clear what aspects of the college experience have these presumed effects, or through what processes and social or psychological mechanisms they operate. Here are familiar research questions; our answers to them may tell us something not only about the long-range impact of college on students, but also something about the even longer-range impact of mass higher education on American society.

Unconventionality, Triangulation, and Inference

EUGENE J. WEBB
Northwestern University

All three of the nouns in this paper's title—unconventionality, triangulation, and inference—are imbedded in a more general concept: multiple operationalism as a way of knowing. With educational psychologists making significant contributions, the mistaken belief in the single operational definition of learning, of performance, or of values has been eroded.

Most students today would agree that it is appropriate to draw simultaneously on multiple measures of the same attribute or construct—multiple measures hypothesized to overlap in theoretically relevant components, but which do not overlap on measurement errors specific to individual methods (16, 17, 7, 19, 38).

In 1953, E. G. Boring (3) wrote:

> As long as a new construct has only the single operational definition that it received at birth, it is just a construct. When it gets two alternative operational definitions, it is beginning to be validated. When the defining operations, because of proven correlations, are many, then it becomes reified.

The most persuasive evidence and the strongest inference comes from a triangulation of measurement processes. Feigl (14) spoke of fixing a concept by triangulation in logical space, and the partition of sources of variance can do just that.

But just as we ask if a correlated x and y are more highly correlated with z, it is also reasonable to ask if the components being converged or triangulated are truly complementary. Are we fully accounting for known sources of error variance?

This is a serious question with most of the multimethod studies now

available. "Multimethod" has usually been defined as multiple scales or behaviors collected under the condition in which the subject knew he was being tested. Humphreys (19), for example, when talking of multiple measures of reasoning, spoke of "series analogies and classification items." The multiple methods thus have tended to be multiple variants within a *single* measurement class such as the interview.

Every data-gathering class—interviews, questionnaires, observation, performance records, physical evidence—is potentially biased and has specific to it certain validity threats. Ideally, we should like to converge data from several data classes, as well as converge with multiple variants from within a single class.

The methodological literature warned us early of certain recurrent validity threats, and the evidence has markedly accelerated in the last few years. It has been 30 years, for example, since Lorge (20) published his paper on response set, and 20 years since Cronbach (11) published his influential paper on the same topic in *Educational and Psychological Measurement*. Further, there is the more recent work of Orne and his associates on the demand characteristics of a known research setting (24, 25, 27, 26) and Rosenthal's stimulating work (29, 30, 31) on the social psychology of the experiment. All these investigations suggest that reliance on data obtained only in "reactive" settings (9) is equivocal.

As a guide to locating the strengths and weaknesses of individual data classes—to better work the convergent multiple-methods approach—my colleagues at Northwestern and I have tried to develop a list of sources of research invalidity to be considered with any data class (38). An outline of these sources of invalidity is contained in Chart 1.

To bring under control some of the reactive measurement effect, we might employ data classes which do not require the cooperation of the student or respondent. By supplementing standard interview or pencil-and-paper measures, more dimensionality is introduced into triangulation.

In a recent paper which described the use of observation methods in the study of racial attitudes, Campbell, Kruskal, and Wallace (8) studied seating aggregations by race. Two colleges were picked in the Chicago area—one noted for the liberal composition of its student body and the other more associated with a traditional point of view. Going into lecture halls, they observed seating patterns and the clustering of Negro and white students during class. With a new

Chart I

Sources of Research Invalidity

I. Reactive Measurement Effect
  1. Awareness of being tested
  2. Role playing
  3. Measurement as change
  4. Response sets

II. Error from Investigator
  5. Interviewer effects
  6. Change—fatigue/practice

III. Varieties of Sampling Error
  7. Population restriction
  8. Population stability over time
  9. Population stability over areas

IV. Access to Content
  10. Restrictions on content
  11. Stability of content over time
  12. Stability of content over areas

V. Operating Ease and Validity Checks
  13. Dross rate
  14. Access to descriptive cues
  15. Ability to replicate

statistical test developed by Kruskal, they were able to demonstrate a greater racial mixture in the more "liberal" college. They also found, however, that the seating mix in the liberal college was significantly less than that expected by chance.

The linkage of secondary records is another way to develop control over reactivity. An example of this approach is DeCharms and

Moeller's (12) study of achievement imagery. They first gathered the number of patents issued by the United States Patent Office from 1800 to 1950. These data (controlled for population) were then matched to achievement imagery found in children's readers for the same period. There was a strong relationship between the level of achievement imagery in their sample of books and the number of patents per million population. Both data series are non-reactive, and although other rival, plausible hypotheses might explain the relationship, it remains as one piece in the inferential puzzle, uncontaminated by awareness of being tested.

For matching of other archival records, we can note Lewis Terman's (37) study estimating Galton's IQ (not far from 200) and Galton's own early studies of hereditary genius (15).

Another class of data comes from physical evidence, one example of which is Fredrick Mosteller's creative study of the degree to which different sections of the *International Encyclopedia of the Social Sciences* were read (22). He estimated usage by noting the wear and tear on separate sections: dirty edges of pages, frequency of dirt smudges, finger markings and underlinings on pages. He sampled different libraries and even used the *Encyclopedia Britannica* as a control.

Thus far, the emphasis has been on data sources and overlapping classes of data. We might also profitably explore the possibility of using multiple samples. Again, this is different from the usual definition of multiple samples. In addition to sampling a number of different classrooms, or groups of students or cities, one may ask if there are different types or categories of samples available for the variable under study. Is there a group of natural outcroppings among occupations, already formed social and interest groups, or people who have common experiences? Can we economically exploit for research purposes the broad spectrum of already formed groups which may be organized along some principle of direct substantive applicability to the investigation?

Professor James Bryan of Northwestern and I have been interested in the use of these "outcropping" groups as a middle-level sampling strategy—one that straddles the elegant but cumbersome national probability sample and the more circumscribed "N=80 volunteer males from the introductory psychology class" populations.

Because one sometimes doesn't know the universe for a study and because of cost restraints, subjects are most often selected because of

proximity. Our subjects are typically drawn from the subject pool of the introductory class, from friends, friends of friends, or those unlucky enough to be members of the same institution as the investigator, be it the school, the hospital, or the prison.

Consider some convenience samples which may supplement conventional groups. Becker, Lerner and Carroll (1) used caddies loafing about a golf course waiting for jobs as a subject pool. E. E. Smith (33) suggested firemen in a fire house. They have almost unlimited time available for questioning and offer the very happy situation of a naturally formed, real group, whose members know each other very well. This is a good setting in which to replicate findings derived from experimentally formed groups in laboratories or from natural groups.

Sometimes these convenient aggregates offer a special opportunity to get a high concentration of usable subjects. To study somatotyping among top athletes in different track and field events, Tanner (35, 36) went to the 1960 Olympic Village at Rome. In a study of proposed brand names for new products, in which one of the criteria was relative invulnerability to regional accents, MacNiven (21) sent interviewers to a nearby airport where they asked travellers to read off lists of names while the interviewers noted variable pronunciations.

In trait measurement, one may define altruism by one or by a series of self-report scales. But it may also be profitable to examine extant groups with some face-valid loading on altruism—say, volunteer blood donors, contributors to charitable causes, or even such groups as those who aided Jews in Nazi Germany.

Bryan and Test (5) have recently reported on a provocative study of the influence of modeling behavior on altruism. Their objective in a field experiment was to see whether or not people stopped to help someone who had a flat tire. The experiment involved two women stranded with flat tires one quarter of a mile apart on a highway and a model, a man who had stopped to help one of them. In one part of the experiment, the traffic passed the woman and the model and then, farther up the highway, passed the other woman. In the other part of the experiment, the traffic passed only one woman and no model.

Other clusters of groups may help to define or locate a particular ability. Occupational categories may be particularly useful here. For studies of superior depth perception there are natural occupational outcroppings such as magnetic core threaders, jugglers, or grand prix automobile drivers.

Each of these groups possesses other attributes, and one might

consider the same group of automobile race drivers as a high risk-taking sample and link them with other high risk-taking groups such as sport and military parachute jumpers (13).

Or, for studies of deviance, there are the self-help deviant groups of Alcoholics Anonymous, Gamblers Anonymous, and prisoners who volunteer for therapy. All presumably share a common characteristic, but the setting of the phenomenon is varied.

As an expansion of this idea, consider Ernest Haggard's exemplary chapter on isolation and personality (18). Haggard reviewed studies of isolation: How is personality affected by the restraint of habitual body movement in restricted, monotonous, or otherwise unfamiliar environments? Instead of limiting himself to the laboratory experimentation on sensory deprivation, he went abroad to the large literature of "naturally" occurring isolation. There are research findings on interstate truck drivers, pilots flying missions alone at night or at high altitudes, orthopedic patients in iron lungs, and anecdotal reports of prisoners in solitary confinement, shipwrecked sailors and explorers. Haggard reports the commonalities among these widely *differing* groups, which overlapped on the isolation dimension, and which shared common sensory and personality phenomena. He compares, for example, the anecdotal reports of Admiral Byrd (6) and the scientific investigation of Rohrer (28) on International Geophysical Year personnel, both of whom found the individual cutting back on information input under isolated conditions—even when a mass of material was available to consume.

As an aside on the nature of isolated man, Haggard quoted Bombard's (2, p.x) comments on the sinking of the *Titanic:*

> When the first relief ships arrived, three hours after the liner had disappeared, a number of people had either died or gone mad in the lifeboats. Significantly, no child under the age of ten was included among those who had paid for their terror with madness and for their *madness* with *death*. The children were still at the age of reason.

In another isolation investigation, Sells considered many of the same data in his applied study, "A model for the social system for the multiman extended duration space ship" (32). Thinking of such long journeys as a Mars shot, Sells assembled data from many isolated groups, both natural and artificial. His analysis was careful and based on theory. He related the findings from different studies to a general model of an isolated social system—evaluating the degree

to which results from the individual studies were likely to transfer to a space vehicle setting. Thus, data from submarine and exploration parties were most applicable, while the findings from shipwreck and disaster studies were least likely to transfer. Naroll (23) has suggested similar procedures to differentially weight data derived from documentary sources of varying credibility, and Stanley (34) has offered a broader approach for treating data in the general multitrait-multimethod matrix format.

In this paper, I have stressed two main points. One is the utility of different data-gathering techniques applied concurrently to the same problem. The other is the laying of these techniques against multiple samples which are natural outcroppings of a phenomenon.

From E. G. Boring (4):

> . . . The truth is something you get on toward and never to, and the way is filled with ingenuities and excitements. Don't take the straight and narrow path of the stodgy positivists; be gay and optimistic, like Galton, and you will find yourself more toward than you had ever expected.

REFERENCES

1. Becker, S., Lerner, M., & Carroll, Jean. Conformity as a function of birth order, payoff, and type of group pressure. *Journal of Abnormal and Social Psychology*, 1964, 69, 318-323.
2. Bombard, A. *The voyage of the heretique.* New York: Simon & Schuster, 1953.
3. Boring, E. G. The role of theory in experimental psychology. *American Journal of Psychology*, 1953, 66, 169-184.
4. Boring, E. G. Personal communication, 1966.
5. Bryan, J. H. & Test, Mary Ann. Models and helping. Unpublished paper, Northwestern University, 1966.
6. Byrd, R. E. *Alone.* New York: Putnam's, 1938.
7. Campbell, D. T. & Fiske, D. W. Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychological Bulletin*, 1959, 56, 81-105.
8. Campbell, D. T., Kruskal, W. H., & Wallace, W. P. Seating aggregation as an index of attitude. *Sociometry*, 1966, *29*, 1-15.

9. Campbell, D. T. and Stanley, J. C. Experimental and quasi-experimental designs for research on teaching. In N. L. Gage (Ed.) *Handbook of research on teaching*. Chicago: Rand McNally, 1963. Pp. 171-246. (Also published as a supplemental monograph by Rand McNally in 1966.)

10. Cook, S. W., & Selltiz, C. A multiple-indicator approach to attitude measurement. *Psychological Bulletin*, 1964, 62, 36-55.

11. Cronbach, L. Response sets and test validity. *Educational and Psychological Measurement*, 1946, 6, 475-494.

12. DeCharms, R., & Moeller, G. Values expressed in American children's readers: 1800-1950. *Journal of Abnormal and Social Psychology*, 1962, 64, 136-142.

13. Epstein, S. The measurement of drive and conflict in humans: theory and experiment. In *1962 Nebraska Symposium on Motivation*. Pp. 127-206.

14. Feigl, H. The mental and the physical. In H. Feigl, M. Scriven, & G. Maxwell (Eds.), *Minnesota studies in the philosophy of science*. Vol. II. *Concepts, theories and the mind-body problem*. Minneapolis: University of Minnesota Press, 1958.

15. Galton, F. *Hereditary genius*. New York: D. Appleton, 1870.

16. Garner, W. R. Context effects and the validity of loudness scales. *Journal of Experimental Psychology*, 1954, 48, 218-224.

17. Garner, W. R., Hake, H. W., & Eriksen, C. W. Operationism and the concept of perception. *Psychological Review*, 1956, 63, 149-159.

18. Haggard, E. A. Isolation and personality. In P. Worchel & D. Byrne (Eds.), *Personality change*. New York: Wiley, 1964.

19. Humphreys, L. G. Note on the multitrait-multimethod matrix, *Psychological Bulletin*, 1960, 57, 86-88.

20. Lorge, I. Gen-like: halo or reality. *Psychological Bulletin*, 1937, 34, 545-546.

21. MacNiven, M. Personal communication, 1965.

22. Mosteller, F. Use as evidenced by an examination of wear and tear on selected sets of ESS. In K. Davis *et al.*, A study of the need for a new encyclopedic treatment of the social sciences. Unpublished manuscript, 1955. Pp. 167-174.

23. Naroll, R. *Data quality control.* Glencoe, Ill.: Free Press, 1962.

24. Orne, M. T. The nature of hypnosis: artifact and essence. *Journal of Abnormal and Social Psychology*, 1959, 58, 277-299.

25. 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.

26. Orne, M. T. & Evans, F. J. Social control in the psychological experiment: antisocial behavior and hypnosis. *Journal of Personality and Social Psychology*, 1965, 1, 189-200.

27. Orne, M. T., & Scheibe, K. E. The contribution of nondeprivation factors in the production of sensory deprivation effects: the psychology of the "panic button." *Journal of Abnormal and Social Psychology*, 1964, 68, 3-12.

28. Rohrer, J. H. Human adjustment to antarctic isolation. Armed Services Technical Information Agency.
Publication AD 246610, Arlington Hall Station,
Arlington, Virginia, 1960.

29. Rosenthal, R. On the social psychology of the psychological experiment: the experimenter's hypothesis as unintended determinant of experimental results. *American Scientist*, 1963, 51, 268-283.

30. Rosenthal, R. Experimenter outcome-orientation and the results of the psychological experiment. *Psychological Bulletin*, 1964, 61, 405-412.

31. Rosenthal, R. Changing children's IQ by changing teacher's expectations. Paper read at American Psychological Association meetings, New York, September 1966.

32. Sells, S. B. A model for the social system for the multiman extended duration space ship (NASA report No. NGR 44-009-008), undated.

33. Smith, E. E. Obtaining subjects for research. *American Psychologist*, 1962, 17, 577-578.

34. Stanley, J. C. Analysis of unreplicated three-way classifications, with applications to rater bias and trait independence. *Psychometrika*, 1961, 26, 205-219.

35. Tanner, J. M. *The physique of the Olympic athlete.* London: Allen and Unwin, 1964.

36. Tanner, J. M. Physique and athletic performance. In S. A. Barnett & Anne McLaren (Eds.), *Penguin Science Survey B: 1965.* Middlesex: Penguin, 1965. Pp. 112-133.

37. Terman, L. M. The intelligence quotients of Francis Galton in childhood. *American Journal of Psychology*, 1917, 28, 209-215.

38. Webb, E. J., Campbell, D. T., Schwartz, R. D. & Sechrest, L. *Unobtrusive measures.* Chicago: Rand McNally, 1966.

The Prediction of Academic and Nonacademic Accomplishment

JOHN L. HOLLAND
American College Testing Program

About 10 years ago, my colleagues and I became interested in the whole area of originality, creativity, or creative performance. Like many others, we wondered how we could distinguish an original from a nonoriginal person, whether or not we could predict creative behavior, and how we could define it.

As a first step we decided to define creative performance as "a performance which is accorded public recognition through awards, prizes, or publication, and which may therefore be assumed to have exceptional cultural value." With this definition as a guide, we then derived a list of achievements at the high school level by reviewing the secondary school achievements of National Merit Finalists.

The items were divided by content into two scales: Creative Science and Creative Arts. Some typical items follow:

- Won a prize or award in a scientific talent search.
- Invented a patentable device.
- Had a scientific paper published in a science journal.
- Won one or more speech contests.
- Had poems, stories, or articles published in a *public* newspaper or magazine or in a state or national high school anthology.
- Won a prize or award in an art competition (sculpture, ceramics, painting, etc.).
- Received the highest rating in a state music contest.
- Composed music which has been given at least one public performance.
- Won literary award or prize for creative writing.

Our first attempt at a definition was discouraging. The estimated reliabilities ranged only from .36 to .55 for groups of Finalists. Student accomplishments meeting our ambiguous definition were hard to find, and we had to settle for a few accomplishments that obviously did not meet our definition. Nevertheless, we pressed on.

These meager criteria, along with a variety of measures thought to be associated with creativity and academic performance, were administered to a large group of Finalists. Generally, the relationships found between the criteria of scientific and artistic performance, and personal, demographic, and parental variables were extremely low and often negligible. But these low relationships did suggest that "creative" performance at the high school level occurs more frequently among students who are independent, intellectual, expressive, asocial, and consciously original. Our results also indicated that at an extremely high level of academic aptitude high school grades and academic aptitude measures were essentially unrelated to our brief checklists of accomplishment.

In a sentence, we found many expected relationships, but they were so small as to be of no practical value. We did, however, acquire an important lesson in researchmanship: Do not use words like "original" or "creative" if you want to get on with editors and colleagues. In all subsequent reports, we substituted terms like "nonacademic accomplishment" for "creative behavior," but we maintained the same criteria of creative behavior with only slight revisions. As a result, we have had no mo. editorial controversy. Our current definition is somewhat more explicit: "Students with high scores on one or more of these simple scales have attained a high level of accomplishment which requires complex skills, long-term persistence, or originality, and which generally received public recognition."

In the next eight or nine years, we proceeded to find useful resolutions for the many problems raised by our first investigation as well as the work of others. I will now briefly describe how we coped with the various subsidiary problems that make up the big problem. Although I will discuss these subproblems as if they were dealt with one at a time, we usually worked on and worried about them at the same time.

Criteria of Accomplishment

The search for reliable and comprehensive criteria has been moderately successful. From the scientific and artistic criteria of low reliability and limited content, we moved first to six criteria—science, leadership, art, music, writing, and dramatic arts—to assess notable extracurricular accomplishment at both the high school and college levels. In their current form, their reliabilities range from .65 to .84 for high school students and from .44 to .80 for college students. More recently, we developed new scales to assess such additional accomplishments as the following: social participation, social service, business, humanistic-cultural, religious service, social science, and interpersonal competency. As a result, we have, in addition to college grades, 13 criteria for assessing a student's accomplishment in college. And, although we began the search for criteria with creativity in mind, we have developed a set of criteria or standards for assessing a student's progress toward many of the goals of a general or liberal education.

When is a Scale a Scale?

In the process of developing more comprehensive criteria, we did a variety of analyses which established that our c. ria were relatively independent of one another and of academic aptitude, and that the individual items did form homogeneous scales. We performed item analyses to see if scale items had been assigned to the appropriate scales. We intercorrelated the difficult and easy items (that is, achievements rarely or frequently attained) for a given scale to learn if they were performing similar functions. Judges in various academic fields were asked to review our scales for their face validity, to cull out poor items, to suggest better items, and to help us build new scales. Although the current criterion scales are only brief checklists, they possess useful reliability and obvious content validity—people who get high scores clearly are more competent, skilled, or original than people who do not.

John L. Hollar^d

THE SEARCH FOR PREDICTORS

Along with developing comprehensive and reliable criteria we also had to search for predictors, or ways to identify high school students who would produce a record of notable accomplishment in college. Over a period of nine years we explored the predictive value of about 300 variables, including measures of parental attitudes, interests, activities, grades, originality, personality, and aptitude. This search led to several mild depressions and the following conclusions:

1. A student's record of nonacademic accomplishment in high school was the best predictor of collegiate accomplishment in the same area; that is, leaders in high school tend to become leaders in college, writers become writers, and so on.

2. Brief lists of activities can be used to form good predictors.

3. Brief lists of competencies also do about as well.

For groups of National Merit Finalists, we obtained predictive validities averaging .38 using records of activities and accomplishments in high school. And in a recent study employing two diverse groups of colleges, the predictive validities of these records of accomplishment average .40. In short, we have developed some simple ways to assess a student's potential for notable accomplishment that have useful reliability and validity.

Potential and Competency Scales

It became apparent as we went along that the use of records of notable accomplishment might favor the student who matures early, or the student from an affluent or large high school where there are more prizes to win, contests to enter, and the like. Consequently, we established short activity scales--similar to interest scales—to assess potential for notable accomplishment in college. A recent study indicates these scales work about as well as the high school records of accomplishment. Finally, we developed simple scales or lists of things a student claimed he could do. These scales, which also proved to have moderate validity, provide a beginning for assessing competency when the opportunity for notable accomplishment is limited.

Academic and Nonacademic Accomplishment

In all of these studies, we have tried to learn whether or not nonacademic performance was independent of academic performance or potential. We have examined our data to see if the lack of relationship found in some earlier studies was due to a narrow range of talent; it was not. Our results make clear (they don't suggest) that these are different kinds of talent and performance. People who have academic talent may or may not have these other kinds of talent.

One important qualification should be made. Because our criteria are only a sample of the existing important accomplishments, we cannot say that all notable accomplishments have negligible relationships with academic potential. On the other hand, our work and the work of Barron, Gough, MacKinnon, Taylor, and others has reduced the possibilities for finding many substantial relationships.

Do Students Lie?

Even if we have established the validity of our scales of accomplishment, there remains another nagging problem: Can you count on students to tell the truth, when the chips are down? To deal with this problem we developed a six-item validity scale to detect students who either exaggerate their accomplishment or get confused in their use of the answer sheet. Using this scale, we discarded less than one percent of the students in several samples and recalculated the relationships between aptitude and grades, and nonacademic accomplishment. In every case we obtained correlational differences only in the second or third decimal places. With the use of this validity scale, we can easily detect the *grand* liar, but the subtle exaggerator we will never detect. On the other hand, it seems unwise to delay helping the vast majority of students because some small percentage will beat the game.

What we have done can be summarized quite simply. Actually, we have only engineered what every layman and mother knows: To find out if a person is going to become an outstanding performer, simply add up his little performances as he moves through life.

John L. Holland

PRACTICAL APPLICATIONS

Perhaps the educational implications of our findings are more important than any immediate application. Why do college and high school grades have little or no relationship to a student's notable accomplishments? Such occurrences raise serious doubts about the effects colleges have on students.

It seems reasonable to assume that when colleges are committed to the goals of liberal or general education, good grades and other notable accomplishments would go together, but they do not. Equally important, the pervasive allegiance to grades as measures of overall worth has had and will continue to have long-term detrimental effects upon students. Somehow we need to inculcate the notion that there are many dimensions of talent, that the absence of academic talent does not spell perdition any more than the lack of musical aptitude does. We can do this by applying broader conceptions of talent to our students and to ourselves.

The American College Testing Program incorporates these ideas in a fifth test, the Student Profile Section. This brief information blank helps the student present himself as a person with a variety of talents, ambitions, and needs. By formalizing this information we reduce the overemphasis on academic potential. At the same time, a college receives in advance a more complete account of the needs and talents of its entering student body. (Of course, many admission blanks perform a similar function, but have you ever tried to tell about your outstanding accomplishments in a space two and one-half inches long?) The mechanization of this nonintellectual information also makes possible a useful profile of a college's entering class. Successive class profiles provide a simple way both to study the effects of changing admission policies and to comprehend the educational needs of entering students.

It is our hope that these nonintellective materials will not become another hurdle in a highly selective admissions procedure. But, as Cronbach points out, when good decisions require information about many aspects of a person, psychometrically it is better to use many psychological devices with moderate reliability and validity against several criteria than a few—or only one—instrument with high reliability and high validity against a single criterion. Certainly, decisions about college attendance are of the latter kind, and the formal use of nonintellective devices broadens the base for student access to

higher education. Moreover, not using such devices is much more detrimental to the student than using them. If, for example, we had required these new methods to meet the highest standards of our old revered aptitude and achievement tests, we would have remained fixated in the aptitude and achievement test era. Unless we provide some running room for research and for revision of services, testing agencies will serve largely to immobilize rather than to facilitate educational practice.

REFERENCES

Barron, F. *Creativity and psychological health.* Princeton, N.J.: Van Nostrand, 1963.

Buel, W. D. Biographical data and the identification of creative research personnel. *Journal of Applied Psychology*, 1965, 45, 318-321.

Chambers, J. A. Relating personality and biographical factors to scientific creativity. *Psychological Monographs*, 1964, 78, 7 (No. 584).

Flescher, I. Anxiety and achievement of intellectually gifted and creatively gifted children. *Journal of Psychology*, 1963, 56, 251-268.

Gough, H. G., Hall, R. E., & Harris, W. B. Admissions procedures as forecasters of performance in medical education. *Journal of Medical Education*, 1963, 38, 983-998.

Holland, J. L. Creative and academic performance among talented adolescents. *Journal of Educational Psychology*, 1961, 52, 136-147.

Holland, J. L., & Astin, A. W. The prediction of the academic, artistic, scientific, and social achievement of undergraduates of superior scholastic aptitude. *Journal of Educational Psychology*, 1962, 53, 132-143.

Holland, J. L., & Nichols, R. C. Prediction of academic and extracurricular achievement in college. *Journal of Educational Psychology*, 1964, 55, 55-65.

Holland, J. L., & Richards, J. M., Jr. Academic and nonacademic accomplishment: correlated or uncorrelated? *Journal of Educational Psychology*, 1965, 56, 165-174.

Holland, J. L., & Richards, J. M., Jr. *Academic and non-academic achievement in a representative sample from a population of 612,000.* Iowa City: American College Testing Program, 1966.

Hoyt, D. P. *The relationship between college grades and adult achievement: a review of the literature.* Iowa City: American College Testing Program, 1965.

Locke, E. A. The development of criteria of student achievement. *Educational and Psychological Measurement*, 1963, 23, 299-307.

MacKinnon, D. W. What do we mean by talent and how do we test for it? In *The search for talent*. New York: College Entrance Examination Board, 1960. Pp. 20-29.

Nichols, R. C., & Holland, J. L. Prediction of the first year college performance of high aptitude students. *Psychological Monographs*, 1963, 77 (7, Whole No. 570).

Price, P. B., Taylor, C. W., Richards, J. M., Jr., & Jacobsen, T. L. Measurement of physician performance. *Journal of Medical Education*, 1964, 39, 203-211.

Richards, J. M., Jr., Holland, J. L., & Lutz, Sandra W. *The assessment of student accomplishment in college*. Iowa City: American College Testing Program, 1966.

Richards, J. M., Jr., Holland, J. L., & Lutz, Sandra W. *The prediction of student accomplishment in college*. Iowa City: American College Testing Program, 1966.

Taylor, C. W., Smith, W. R., & Ghiselin, B. The creative and other contributions of one sample of research scientists. In C. W. Taylor and F. Barron (Eds.), *Scientific creativity: its recognition and development*. New York: Wiley, 1963. Pp. 228-250.

Wallach, M. A., & Kogan, N. *Modes of thinking in young children: a study of the creativity-intelligence distinction*. New York: Holt, Rhinehart, and Winston, 1965.

Luncheon Address

Education's Age of Flexibility

Francis Keppel
General Learning Corporation

Any speaker who follows luncheon and precedes an afternoon meeting is a kind of intellectual digestion tablet. He should be bland and seek to avoid creating a rumble. Gentle speculation is what is needed, an unimaginative tone with nothing very startling to cause tension. This is what you will get from me, and it is perhaps all that could be expected from a recently departed servant of government. And by the way, the tablet is designed for less than 25 minutes.

To set the tone and establish some kind of analogy, let's turn our minds back to the beginning of the century, a time most of us have been taught to regard as an era of some stability and of a good deal of complacency in our society.

One complacent group in those days, I am told, were the physical scientists many of whom felt that they had the universe pretty well taped. What was needed was to make careful observations of their Newtonian universe, measure things more accurately, figure out a lot of things along lines that were already understood, and before long they would know how everything operated. They were wrong, of course, and back in the uneasy corners of their minds many knew it. There were some mysteries they couldn't explain, such as radioactivity or the way that light behaves—sometimes like a ripple of waves and sometimes like a stream of solid particles—and there were some contradictions, too, but these things were swept under the rug pending a more thorough housecleaning at some later time.

Then something extraordinary happened. In 1900, Max Planck found evidence that energy came, not in all convenient sizes but in separate chunks, something like atoms—very tiny, but nonetheless measurable and discrete. Five years later Albert Einstein delivered his extraordinary stroke of mathematical genius and upset the scientists' well-balanced reasoning about the universe.

A revolution had taken place. Not many people were aware of it, but there was nonetheless a kind of underground buzz of growing excitement among physicists. In time, they found that the alchemists might have been right after all—that atoms could be smashed and that matter could be transmuted from one element to another. Some people were even willing to risk the derision of their colleagues by claiming that perhaps energy could be extracted from atoms. Then came the Second World War, when energy was somewhat spectacularly extracted from atoms. Science's long-awaited, thorough housecleaning had finally taken place.

In the flick of an eye, the phenomenon that had until then been the private knowledge of just a handful of human beings suddenly involved all mankind. The revolution spread out from a small corner of the world of science to the larger world, from scientific speculation to moral, military, and industrial questions. And now the atom and its powers form a part of our everyday lives. Most of us still do not truly understand the differences between the universe of Newton and the universe of Einstein, but we accept it. And increasing numbers, at least of our younger men and women, understand it.

I suppose we ought to be reconciled to the idea that in certain areas of human knowledge there will necessarily be some very tiny societies of human beings who can comprehensibly communicate only with each other. Maybe that is the way with all knowledge when it is very young. But we no longer seem able to afford the luxury of permitting knowledge to stay young very long.

That brings me back to the analogy. At the start of the twentieth century, the world of science was in what used to be called an "interesting condition"—that is, pregnant. The period of gestation was somewhat protracted but, as in the case of the elephant, the issue was weighty. I think education today is in a similarly interesting condition. We seem to be at the threshold of some major new discoveries about learning and the processes of education. We would do well to be prepared for them.

There was a time, not too long ago, when education was thought of, more often than not, as its own little universe, as a thing apart from the rest of society. That is no longer nearly so true. Education has become more and more involved with the rest of society, with government, with industry, with all manner of agencies and institutions. The problems that beset all of us—urbanization, the population explosion, automation, communications, and so on—are also education's problems, both in the sense that they affect education and in the sense that education is helping to solve them.

There is still another new aspect of education that is even more indicative of major changes to come. In the past, education consisted apparently of fixed amounts of knowledge to be absorbed in fixed periods of time, of known concepts and known blocks of factual matter. In such a framework, the various elements of education—instruction, materials, architecture, testing—had fairly explicit and well-determined roles. Now that is less the case than ever before. Education daily becomes more fluid and dynamic, in terms not only of its own processes, but also of its objectives and its end products. What is most significant, however, is that this is not just a temporary state of affairs, not just a symptom of its present interesting condition. It is rather a characteristic of its new role in society, and continuing change may well be the rule rather than the exception, just as it is for an increasing number of institutions in our society. All the forces within education will have to adapt to changes that will continue to come from a number of different directions. There are at least four areas in which the need for such adaptation is fairly obvious:

1. First, of course, there is new knowledge of all kinds, proliferating in almost every direction. From new insights into religion obtained from the Dead Sea Scrolls to new theories of chemical bonding, all this will become part of mankind's consensus of knowledge. It will not only be taught to the young, but will move into the content of the necessary continuing education that most of us will be constrained to undergo.

2. Next are new approaches to the content of education: new curricula, such as modern mathematics, the wave theory approach to physics, and a host of interdisciplinary approaches in the humanities and sciences.

3. Third, we will need to adapt to the new and improved tools for teaching and for learning. New kinds of hardware, as well as such

new techniques as linear and branched programmed instruction. will surely give us greater accessibility to the mind of the learner.

4. Finally, we have reason to hope that we may be approaching a new appreciation of the mind and how it appears to work. The growing knowledge and familiarity with cognition, memory, transfer, and conceptual understanding will surely give us insights into all mental processes, including the learning process.

I called these the "obvious" areas of adaption, and I know that all of you are more familiar with these developments than I. What is less obvious—at least to me—are some of the ways we need to adapt to these changes—in short, the kind of flexibility that is required.

Should we, for example, build elements of flexibility into our teaching and learning environment, at least to the extent that the requirements of architecture and basic creature comfort permit? This is far more difficult than it may appear to be at first blush. To a certain extent, all environment is learning environment. Since the home and its surroundings make up the dominant environment of the young, we can observe that this becomes an extremely flexible learning environment for some, and a fairly rigid learning environment for others. What is unfortunate is that the least flexible environment engulfs those who are already disadvantaged in other ways.

Another area of flexibility, it seems to me, is in testing, and I know that you are well started on this road. By becoming increasingly sensitive to the consequences of education, testing can bring greater flexibility to the whole learning process. Such electronic memory and logic devices as the computer show great new promise with their capacity for making minute measurements of the pupil's progress, and for integrating the instruction and testing processes.

Yet flexibility comes no more easily to education than it comes to other institutions in society or to you and me when we must shake off old habits and routines. Education, as a matter of fact, has had a long heritage of rigidity throughout most of the world. It may be worth going back to the record in other lands if we wish to get some measure of the problem we face here.

In many European countries, including those which served as the wellspring of our own educational institutions, central government agencies tend to prescribe both the content and conduct of teaching. The teacher must adhere to the syllabus, and external examinations are devised to test how closely the syllabus has been followed.

The former colonies of the European powers are frequently "more royal than the king" and exaggerate this characteristic to the point of vice, a fact that many leading educators in the new nations deplore. Their schools, they feel, are designed to train clerks, not to educate men and women. Otonti Nduka, lecturer in philosophy at the University of Nigeria, has written this: "Part of our trouble . . . is that our educational system is one that tends to produce students with a textbook mentality. The emphasis tends to be more on the memorizing of facts, with a view to passing examinations, and less on the method of finding out facts and learning to apply them." In one secondary school classroom in Kenya, the teacher was once upbraided by his students with shouts of "N.E.! N.E.!" The letters stood for "non-examination." The teacher had had the effrontery to introduce material that would not be on the standard examination. The principal of Makerere University, Y. K. Lule, has no fondness for the rigid syllabus, but feels "it is necessary because of the quality of the teachers available." And a recent report of the Kenya Education Commission said this: "One of the results of the employment of large numbers of unqualified teachers is that they so greatly influence the general tone and methods of the school in a conservative direction, as to make it hard for the newly qualified teacher, trained in up-to-date methods and anxious to try them out, to put his training into practice." Can we honestly say that we differ in kind from this statement—or just in degree?

The problems of education throughout most of the rest of the world do not seem to be much different. Asia's teachers generally are not well trained, by U. S. standards, and educational systems throughout most of Asia tend to discourage their use of initiative and ingenuity.

The educational philosophy of Latin America is patterned after the Spanish, which has been described as one of keeping the social and economic classes in fixed positions, and thus working against vertical mobility within the society. As an example of the general approach to learning, a Ford Foundation consultant in Chile has pointed out that "the professor doesn't want his students to have books, because books threaten the authoritative stance which the teacher has in relation to the students."

I have taken this hasty trip around the world to illustrate the point that lack of flexibility is so often synonymous with poor teaching practices. Much of the contribution that the United States has made

to education generally has been to replace rigid practices with more free-wheeling inquiry. Compared with the fact-cramming techniques of most educational systems, American education seems far more to be based on a problem-solving approach.

Except by contrast to a good deal of formal education in other lands, I'm not too certain, however, that American education really deserves such plaudits. Far too often, I'm afraid, despite our intentions, we teach to the test, rather than test the teaching. When that happens, the designer of the test takes over the role of shaping the curriculum rather than following it and reporting on how well it has been learned.

There are many heartening signs of a willingness to innovate in American education, to try a wide assortment of curricular experiments, and to accept or reject them on their merits. Witness the hundreds of schools and school systems across the country that have adopted the new mathematics, the Physical Science Study Committee physics courses, the new biology and chemistry courses, as well as a host of new approaches to language arts and social sciences.

It has been this kind of flexibility that has already brought about a considerable amount of bootstrap lifting all across the spectrum of American education. The upgrading that has already taken place led President James A. Perkins of Cornell University to observe: "On the qualitative side, secondary education has improved dramatically, particularly since our rude awakening by Sputnik in 1957. As a result, the responsibilities for general education have slowly been assumed by the high school and the preparatory school. In the university, general instruction has given way to far more sophisticated work in the first two years."

Yet it is clear that educational institutions need to demonstrate still more willingness to innovate and to experiment in more new directions. One new tool, for example, is systems analysis, which has already been used successfully in both industry and government. There is every reason to believe that, with the application of sufficient brainpower, it could work equally well for education.

The resource now available to education that is by far the most flexible is the teacher. To take advantage of that fact, systems analysis may help to make better use of the strengths of the teacher. Few today could argue that the present administrative arrangements provide full use of teacher flexibility. The case can be made that present arrangements, by and large, do not encourage teachers to become more adapt-

able to changed situations. Rather than seek to have the teacher reach out for new techniques, new methods, and new subject matter, they may tend to switch the teacher onto fairly narrow-gauge tracks that help simplify the problems of administration itself.

A good job of systems analysis and planning would not only seek, therefore, to achieve maximum effectiveness from all kinds of teaching materials and equipment, but would build a high degree of teacher flexibility right into the system.

As far as education is concerned, of course, the major stumbling block to reaching such a goal is reaching agreement on goals and objectives. We need to know what we want to be flexible for, and there is no more difficult task.

The society for which we are preparing young people is no longer so much fixed as fluid, no longer so much stable as changing, often dynamically and drastically. The rules, no longer rigid, sometimes seem to bend over double. The knowledge, no longer neatly packaged, now keeps breaking out at the seams.

So now we have to prepare young people, and older people as well, for a persistently changing world. This cannot be done unless we help to make them more amenable to change, more flexible individuals. To do its part of the job, education must itself wholeheartedly enter the new age of flexibility.

All of this has, it seems to me, some major bearing on the field of testing. I am aware that testers and testing technicians have long been in the van of those asking for criteria, for standards. State your educational goals, they say, and we will devise ways to measure whether you have achieved these goals.

This is clearly an eminently reasonable and logical approach. But it may not be good enough. All of us may have a certain intuitive awareness of our appropriate goals. But the great challenge facing education today is to state those objectives in a way that will satisfactorily approximate a consensus on a variety of topics, and change as the needs and the consensus change. The assignment is difficult enough to demand the best efforts of all of us—scholars, administrators, teachers, testers, and educational suppliers. But nobody should be let off that hook.

It seems to me that one of the important initiatives taken in this direction is that of the Carnegie Corporation's Exploratory Committee on Assessing the Progress of Education. This committee has enlisted the help of a wide selection of specialists from within the field of edu-

cation itself in seeking a way to report on the results of American education. In the process, the Committee may be building a degree of consensus by the very process of having lay panels determine those objectives that they deem worth pursuing.

What they are seeking, in effect, is more knowledge, more information about education that might appropriately be added to the publicly held store of common knowledge. This has been likened to the sharply felt need for more information about the national economy during the depression of the 30's, and the subsequent development of the Gross National Product as the new measuring stick of our economic achievement.

We need units that are far different from the degree, the diploma, the certificate, or the child-year that we have often used to quantify education in the past. This might very well be where the testing specialist comes in—working, of course, with other educational specialists. Clearly, this becomes anything but a simple matter, as you know so well. Criteria as to whether learning has actually, or only seemingly, taken place, as to whether it is merely superficial or fundamental, whether or not it has taken root so that it can grow by itself, how long it has been retained, and how well it can be applied—all of these need far more development. Certainly it would seem to be one of the great creative challenges before us all.

In education we have been called technologically backward. Many of our tools and techniques have not changed for decades, even centuries. This either means that the best ways to teach and learn were discovered hundreds of years ago, or it betokens resistance to change and a lack of flexibility. I honestly believe there is something of the truth in both inferences. But while we can continue to live with the first, we can no longer tolerate the second. The problem of how much and how well people must learn is so great and so pervasive that we must try many things in order to discover how learning can become more effective.

I think we can look with great hope to the future, to changes that are already under way, to other changes that lend great promise to the future, and to a mounting spirit of willingness to accept change in education. To give the context in which such change should take place, I would like to close by quoting the words of a great teacher, words that look to both the past and the future. These are the words of Rabindranath Tagore, and they are inscribed on a plaque hung in

the hallway of the Central Institute for Teacher Training in New Delhi. The sign reads:

> A teacher can never truly teach unless he is still learning himself. A lamp can never light another lamp unless it continues to burn its own flame. The teacher who has come to the end of his subject, who has no living traffic with his knowledge, but merely repeats his lessons to his students, can only load their minds. He cannot quicken them. Truth not only must inform, but also must inspire. If inspiration dies out, and the information only accumulates, then truth loses its infinity. The greater part of our learning in the schools has been waste, because for most of our teachers their subjects are like dead specimens of once living things, with which they have a learned acquaintance, but no communication of life and love.

Session II

Theme:
Natural Language and Computers in Education

An Interactive Inquirer

PHILIP J. STONE
Harvard University

This paper considers some new explorations based on our past experience with a form of automated language processing called content analysis. We will first briefly describe our basic analysis techniques and then relate them to some new ventures in dialogues with the computer as applied to education.

Content analysis procedures are concerned with the identification of repeated symbols or themes in text. These procedures have been shown to be relevant to research in psychology, sociology, political science, anthropology, and education. The variety of textual material studied includes autobiographies, thematic apperception tests (TAT's), folktales, college admission essays, acceptance speeches by presidential candidates, newspaper editorials on the Common Market, diplomatic notes, personal letters collected over a number of years, therapy protocols, open-ended survey interview responses, and sentence completion responses. A number of these studies are reported together in a book (8).

An example of automated content-analysis scoring, in this case scoring for need-achievement, is seen in Figure 1. In this figure, the text appears on the left and the categories into which words and phrases are assigned appears on the right. The first step is to perform a many-to-few mapping of the original text into a smaller number of relevant categories. Thus, the text word *dreaming* in the first sentence of Figure 1 is categorized as NEED, the word *becoming* is categorized as TO-BE, *great* is an ADJECTIVE-POSITIVE, and *inventor* is, from the point of view of achievement in Western culture, a ROLE-POSITIVE. The second step is to examine the pattern of assigned

Figure 1

Story Scored for Need-Achievement

Sentence 1: The student is *dreaming* about *becoming* a *great inventor.*	NEED TO-BE ADJECTIVE-POSITIVE ROLE-POSITIVE SENTENCE SUM = AI
Sentence 2: After *years* of *labor* the crucial moment arrives.	TIME VERB-POSITIVE SENTENCE SUM = UI
Sentence 3: He *hopes* everything will *turn out well.*	NEED VERB-POSITIVE ADVERB-POSITIVE SENTENCE SUM = AI
Sentence 4: But the *experiment* will *fail.*	VALUE-POSITIVE FAILURE SENTENCE SUM = UI
Sentence 5: *Displeased* but still *confident* he will modify his procedures and try again.	AFFECT-NEGATIVE VALUE-POSITIVE SENTENCE SUM = AI

****SUMMARY****THIS DOCUMENT CONTAINS ACHIEVEMENT IMAGERY.

categories for certain thematic sequences. In our first sentence, the pattern NEED, TO-BE, and ROLE-POSITIVE, in that order, is considered adequate for a sentence-summary scoring of achievement imagery (AI). The second sentence, however, does not match a pattern and is scored as unrelated imagery (UI), even though several categorizations were made as potentially relevant to achievement imagery. Notice that the pattern analysis can extend across sentences. For example, FAILURE in the fourth sentence is not adequate in itself to be scored as achievement imagery, but when it is combined with AFFECT-NEGATIVE in the same sentence or the next sentence, an achievement imagery (AI) scoring is made at that point. Finally, a total evaluation is printed at the end of the story.

In this work, the set of computer programs is called the *General Inquirer*. The *Inquirer* can be considered analogous to a very efficient clerk who lacks any ideas of his or her own but if told what to do,

will carry out the task efficiently and mechanically. Directions for categorizing must be supplied in the form of a dictionary. The direction for scoring co-occurrence patterns must be supplied in the form of rules. In the case of scoring need-achievement, investigators Ogilvie and Woodhead (8, Ch. 5) developed a dictionary that classified 855 words and phrases into 14 different categories (see Figure 2) and specified nine different scoring rules (described in Figure 3).

Note that any one rule in Figure 3 can handle a number of different ways that a theme might actually be expressed in the text. Our first sentence pattern in Figure 1, of NEED, TO-BE, ADJECTIVE-POSITIVE, ROLE-POSITIVE, would satisfy both Rules 7 and 8 in Figure 3. Since there are 57 different words categorized as NEED, 6 words and phrases categorized as TO-BE, and 38 different kinds of ROLE-POSITIVE, the total number of acceptable sequences for Rule 7 is 57 x 6 x 38, or 12,996. The number of potential instances of Rule 8 is even more. As a whole, we are quite pleased with our initial successes in scoring need-achievement. When 240 TAT compositions were categorized by the computer (in batches of 60 stories), the percent of agreement

Figure 2

Achievement Dictionary, Category Names and Sample Words

Tags	*Examples*	*Number*
NEED	wants, desires, hopes, yearns	57
TO-BE	become, becoming, to become	6
COMPETE	win, gain, overtake, surpass	28
VERB-POSITIVE	doing, making, inventing, working	136
ADVERB-POSITIVE	carefully, properly, cautiously, thoroughly	50
ADJECTIVE-POSITIVE	great, powerful, promising, splendid	166
VALUE-POSITIVE	discovery, creation, curiosity, intelligence	142
ROLE-POSITIVE	surgeon, lawyer, executive, professor	38
BLOCK	test, broken, damage, crisis	53
SUCCESS	fame, success, glory, honor	23
FAILURE	error, incorrect, mistake, blunder	43
EFFECT-POSITIVE	joy, happy, cheerful, delighted	27
EFFECT-NEGATIVE	sad, anxious, sorry, worried	82
TIME	lifetime, life, years, weeks	4

Figure 3

Summary Rules in Scoring Need-Achievement

Rule 1: NEED + COMPETE
"He *wants* to present a clearcut synthesis of these two conflicting philosophies, to satisfy his own ego and *gain* academic recognition from his professor.

Rule 2: SUCCESS + AFFECT-POSITIVE (within- and cross-sentence routine)
"The worker wanted *fame* and got it. He died a *happy* man."

Rule 3: FAILURE + AFFECT-NEGATIVE (within- and cross-sentence routine)
"The invention will be a *failure*. *Discouraged* and financially bankrupt, the man will drown himself with liquor."

Rule 4: VERB-POSITIVE + ADVERB-POSITIVE
"The operator is hoping that everything will *pan out properly*."

Rule 5: VERB-POSITIVE + VALUE-POSITIVE
"The first man wants to get it *fixed* and do a *good job*."

Rule 6: ADJECTIVE-POSITIVE + VALUE-POSITIVE
"He will wander from this *steadfast purpose* but eventually achieve it."

Rule 7: NEED + TO-BE + ROLE-POSITIVE
"For a long time he has *wanted to become* a *mechanic*."

Rule 8: NEED + TO-BE + ADJECTIVE-POSITIVE
"All he *wanted* was *to become great* at something."

Rule 9: TO-BE + SUCCESS (last sentence routine)
"Mutual compromise and the machine *will be* a *success*."

between the automatic method and trained scorers varied from 82 to 86 percent.

At present, our main efforts are to improve the quality of the many-to-few categorization procedures. Most of our present categorizations are based on the appearance of a word. Some are based on a multiword string such as *turn out*, *to become*, or *United States*. Up to now, we have allowed ourselves to be satisfied with assigning the most predominant meaning and letting the matter go at that. Thus, we do not yet separate the occurrence of *patient* as a noun from its occurrence as an adjective. A word such as *great*, for example, is usually, but not always, an ADJECTIVE-POSITIVE, and in those instances that are exceptions, we would make categorization errors.

Since our purpose is to draw statistical conclusions, such as the "members of group X tend to have more need-achievement in their TAT's than the members of group Y," we can tolerate a certain amount of error in our measurement procedures. If a word has a predominant meaning, we can assume it will usually mean that, and assign categories accordingly. However, if it is more evenly divided in its usage, we may prefer to ignore its occurrences. (For example, if the word *club* appears, it could mean a stick or a social organization.)

At present, there are some 17 different dictionaries for the *General Inquirer* based on our existing procedures. These are briefly described in Figure 4. While investigators tend to borrow from each other in

Figure 4

General Inquirer Dictionaries

Harvard III Psychosociological Dictionary. A second revision of the Psychosociological Dictionary. The number of tags has been reduced from 164 to 83 in describing some 3,500 entries. The dictionary has been used with considerable success in a wide variety of studies.

Yale additions to Harvard III Dictionary. An additional 16 tags developed by Z. Namenwirth at Yale for the analysis of "prestige paper" editorials about the common market.

National Opinion Research Council Survey Research Dictionary. A dictionary roughly following the category scheme of the Harvard III Dictionary, making considerable adjustment for survey response language used by middle and lower class subjects. Contains over 500 idioms. Developed by Bruce Frisbie at the University of Chicago.

Psychoactive Drug Study Dictionary. Developed by T. Dinkel at the University of Chicago to delineate different modes of reaction of psilocybin, the dictionary builds upon the Harvard III Dictionary base.

Stanford Political Dictionary. Developed by Ole Holsti, this dictionary focuses on Osgood's three semantic differential dimensions: positive-negative, strong-weak, active-passive. Each dimension has tags for six levels of intensity, three for each pole. Additional tags are provided for classifying names and places in political documents.

Santa Fe Third Anthropological Dictionary. Developed by B. N. Colby at the Museum of New Mexico, this dictionary is for cross-cultural comparison of folktales and projective test materials. Originally centered on the Kluckhohn value categories and a number of specific concepts, the third version takes a more general framework.

Davis Alcohol Dictionary. Built by William Davis at Harvard for testing hypotheses concerning relations of themes in a world-wide sample of folktales to cultural uses of alcohol, the dictionary currently contains 99 tags, 3,600 entry words, some 90 idioms, and several "sentence summary" scoring routines.

McPherson Lobbying Dictionary. Developed by William McPherson for the study of lobbying communications. the design of the dictionary draws heavily on Parsonian theory. 38 tags are used in classifying some 2,400 words. This dictionary has also been used in the analysis of political acceptance speeches.

Lasswell Value Dictionary. A dictionary centered around the eight value categories outlined in Lasswell's and Kaplan's *Power and Society*. Developed by Z. Namenwirth and H. Lasswell at Yale University.

Who-Am-I Dictionary. Developed by B. McLaughlin at Harvard for analyzing multiple open-ended responses to the question, Who am I? The dictionary uses 30 tags in describing 3,000 entries, including about [illegible] idioms.

Simulmatics Dictionary. Developed by Stone and Dunphy in conjunction with the Simulmatics Corporation for the analysis of product and corporation images, the dictionary base contains some 70 tags for about 2,500 entries, including a number of idioms and sentence summary routines.

Figure 4 (continued)

Need-Achievement Dictionary. Developed by D. Ogilvie and Mrs. L. Woodhead, the dictionary closely follows the hand-scoring directions outlined by D. C. McClelland for the scoring of "need-achievement" imagery in projective test materials. The dictionary has 25 tags, 1,200 entries, about 30 idioms, and a number of sentence-summary routines. A special set of scoring procedures is programmed to analyze the pattern across sentences and provide a net "need-achievement" score for each story. The computer scoring correlates well with hand-scoring methods.

Need-Affiliation Dictionary. Developed by J. Williamson at Harvard for scoring "need-affiliation" imagery in projective materials, the dictionary uses similar strategies to those employed in the Ogilvie and Woodhead procedure. The construction of this dictionary has pointed up the need for further theoretical clarifications concerning this topic. Further refinements can then follow.

Icarian Dictionary. Developed by D. Ogilvie and D. Dunphy at Harvard for measuring symbolism associated with the Icarian myth. Used in a number of personality and cross-cultural studies in relation to early father absence.

Tzotzil Humor Dictionary. A 2,500 word dictionary developed by V. R. Bricker to be used in analyzing Tzotzil humor texts. Tzotzil is a Maya Indian language spoken by the Zinacantecos of Chiapas Highland, Mexico. The tags reflect themes important in Zinacanteco culture.

Gê Mythology Dictionary. Anthropological dictionary developed by Pierre Maranda at Harvard for the analysis of plot within Gê mythology. 2,000 words defined by 99 categories, based on Lévi-Strauss theory of the structure of myths.

Edinburgh Dictionary. Developed by T. Burns and Miss R. Johnson of the Department of Sociology at the University of Edinburgh for the analysis of case discussions by panels representing various professions. Contains numerous idioms.

developing categories, each dictionary has at least 10 distinctive categories of its own. The system has been successfully run at a number of different computer installations in the United States and Europe. From approximately 30 studies using the *General Inquirer*, we now have approximately six million words on IBM cards representing the kinds of language data that behavioral scientists tend to study.

We are now involved in an extensive project to develop basic contextual procedures useful to all dictionaries. We find that most ambiguous words can be disambiguated with surprisingly little context. We hope to develop some useful rules of thumb that will correctly handle some 90 percent of those word occurrences that need contextual identification.

Rather than start with all possible word meanings as they are listed in Webster's dictionary, we are instead concerned with identifying word usage as it actually tends to occur. Our approach is empirical. From our six million words on IBM cards, we have taken a sample of 500,000 words and put them in a massive "keyword in context." A sample of the word *play*, a particularly difficult word, is presented in Figure 5. This listing informs us what word usages are most common and often suggests contextual procedures for identifying them. Satisfactory rules can often be identified in a few minutes. A very complicated and common word, however, may take several days to work out. The task of examining several thousand words is long and

Figure 5

Example of Key Word in Context: "Play"

SELDOM PLAY I DONT KNOW AND I PLAY I PLAY FOR MY OWN NO I CANT NOW I LEARNED	1F1 JAFFE CASE
I WAS ON THE PITCHING MOUND. I PLAY FOR THE DETROIT TIGERS. WELL, I	5A1 HARTMAN DETROIT GANGS
TWO TO TWELVE. WE COMB THEIR HAIR, PLAY GAMES WITH THEM, SEW CLOTHES FOR	2D5 PEACE CORPS
TO TALK WITH THEM, SHOW AND PLAY GAMES, SING AND WATCH THEIR FACES	2K2 RAHALLO
OBSERVE THE RITUAL PROHIBITIONS. THEY PLAY GONGS AND DANCE. ON THE EIGHTH DAY	8E1 IFUGAO
WE WILL GET SUGAR CANE TODAY. THEY PLAY GONGS FOR THE SUGAR CANE. THEY CUT	8E1 IFUGAO
WENT STRAIGHT TO HERNENDEGAS HOUSE TO PLAY HAVOC. A LONG BLAST FROM A HORN WAS	8J1 KIKUYU
MIGUEL HE SAID, FOR E-STARO LET US PLAY HE SAID. TAKE OFF YOUR CLOTHES WE	0A1 ZINACANTAN DREAMS
TO HIS ROMANS HOUSE HE WOULD COME TO PLAY HIS GUITAR. WELL NO THE OLD MAN	0A1 ZINACANTAN FOLK
EVEN LEARN THE PIANO AND I I I SELDOM PLAY I DONT KNOW AND I PLAY I PLAY FOR	1F1 JAFFE CASE
I I I SELDOM PLAY I DONT KNOW AND I PLAY I PLAY FOR MY OWN NO I CANT NOW I	1F1 JAFFE CASE
AND SADDLES. THAT WAS OUR GREATEST PLAY IN OUR VILLAGE IN BASUTOLAND. THE	5B1 SUZMAN SOUTH AFRICAN
SOME STRONG FEELINGS COMING INTO PLAY IN THE CASE OF SOME OF US WHEN WE	2G4 DUMPHY SMALL GROUP
BUY ME JUST A GUITAR I AM GOING TO PLAY IN THE MELON PATH SAID THE BOY.	0A1 ZINACANTAN FOLK
THAT SHE AND A GIRL WERE PUTTING ON A PLAY IN THE NEIGHBORHOOD. THEY HAD TO	0B2 CHILDREN S DREAMS
USUALLY GO TO THE BEACH TO BATHE AND PLAY IN THE SAND. ALSO, I HAVE HELPED	2D2 PEACE CORPS
POOL I CANNOT SWIM BUT I ALWAYS PLAY IN THE WATER ON OTHER KIDS AND THEN	5A2 HARTMAN DETROIT GANGS
THEN IF YOU FALL IN. SO DO NOT NEVER PLAY IN WATER WHILE YOU ARE ALONE. THIS	5A3 HARTMAN DETROIT GANGS
VIEW OF THE CONSIDERABLE ROLE THE SIX PLAY IN WORLD TRADE. THE ESSENTIAL	7A2 ZVI COMMON MARKET
PUT ON NOTHING THAT IS OLD THAT HE CAN PLAY IN. HE DRESS UP ALL THE TIME, AND	5A2 HARTMAN DETROIT GANGS
SO THE BOYS WOULD HAVE MORE ROOM TO PLAY IN. NOW THE HOUSE GETS UPSET. IF	4B1 FRISBIE DETROIT SAMPLE
HE SAID. TAKE OFF YOUR CLOTHES WE WILL PLAY INSIDE. HE SAID. MIGUEL WAS GETTING	0A1 ZINACANTAN DREAMS
FACE. YOU KNOW, HE PROBABLY COULD NOT PLAY IT AND HE GOT DISCOURAGED WITH	5A1 HARTMAN DETROIT GANGS
THOUGHT HE WOULD NEVER LEARN HOW TO PLAY IT AND STOPPED TRYING. WELL, THEN I	5A1 HARTMAN DETROIT GANGS
ISNT YOUR FACE NOW MMM AND NOW YOU CAN PLAY IT BUT. YOU LEARNED IT MMMMMM MMM	1F1 JAFFE CASE
HE CAN MAKE MONEY BY LEARNING HOW TO PLAY IT. IS THIS A LADYS I GUESS IT IS A	5A1 HARTMAN DETROIT GANGS
BIG ENOUGH TO PLAY IT OR LEARN HOW TO PLAY IT. PROBABLY LEARN HOW TO PLAY IT.	5A5 HARTMAN DETROIT GANGS
HOW TO PLAY IT. PROBABLY LEARN HOW TO PLAY IT. THIS IS A LADYS YEH, THIS IS A	5A5 HARTMAN DETROIT GANGS
STILL POUNDS IN HIS HEAD--HE MUST PLAY IT, BUT HIS TECHNICAL SKILL AND	5D1 WILLIAMSON UNDERGRAD
AS I WAS STILL A YOUNG ONE, I USE TO PLAY LIKE HIM UP TO NOW I HAVE SENSE BUT	5C1 SUZMAN SOUTH AFRICAN
WHEN I WAS STILL YOUNG. I USED TO PLAY LIKE THEM, AT HOME. LATER I GOT	5B1 SUZMAN SOUTH AFRICAN
TO DO THINGS LIKE THIS. I TOO USED TO PLAY LIKE THIS CHILD, CALLING OTHER	5J1 SUZMAN SOUTH AFRICAN
ANOTHER, AND WE WOULD FALL DOWN AND PLAY LIKE WE DEAD, AND SO WE WOULD TAKE	5A2 HARTMAN DETROIT GANGS
WE JUST GO. GOES AROUND CUN HOUSE AND PLAY MARBLES ALL THE TIME AND WHOEVER	5A2 HARTMAN DETROIT GANGS
YOU SEE BECAUSE MY HUSBAND WENT TO PLAY MUSIC. IT SEEMS I DO NOT BELIEVE	0A3 ZINACANTAN DREAMS
TOMORROW SHE WILL COME TO ASK YOU TO PLAY ON A SWING WITH HER. THE WOMAN I AM	8K1 TETON
ONE IS EVEN BETTER. LITTLE HARE SAID PLAY ON IT HE PLAYED ON IT. NOW LITTLE	8G1 BASUTO TALES
AND GIVE ME YOURS THAT I MAY LEARN TO PLAY ON THEM. THE RABBIT ANSWERED NO	8H1 BASUTO TALES
ME YOUR FLUTES PLEASE. I WOULD LIKE TO PLAY ON YOUR FLUTES. THE RABBIT PLAYED	8J1 BASUTO TALES
SAW SOMEBODY PLAYING ONE SO HE TRY TO PLAY ONE, A VIOLIN. HE WANTED TO LEARN	5A1 HARTMAN DETROIT GANGS
DOINGS HE IS TRYING TO FIND OUT HOW TO PLAY ONE, AND HE. I DO NOT KNOW WHAT YOU	5A1 HARTMAN DETROIT GANGS
OF PLEASURE FROM KNOWING HOW TO PAY PLAY PIANO OR NOT. NOW WHY DO YOU FEEL	1A1 FAM INTERACTION FRISBIE
YOUS IF I WERE YOU I WOULD GO AND PLAY SOME FORM OF SPORT TO FORGET. THIS	5B1 SUZMAN SOUTH AFRICAN
SUCH GENERALIZED APPROVAL SEEKING MUST PLAY SOME PART IN MOST DISCUSSIONS, AND	2M5 SHAPIRO SMALL GROUP
JIM WAS ON HIS HANDS AND KNEES AND PLAY SOMETHING ON HIM. I WAITED AND IT	6B1 HUCK FINN
HE COULD NOT SCARE THEM AND HE TRY TO PLAY TARZAN. HE TIED A ROPE TO A HOUSE	5A1 HARTMAN DETROIT GANGS
THE HOSPITAL WITH A BROKE HEAD. TO PLAY TARZAN. THIS IS A TOWN, AND UP OVER	5A1 HARTMAN DETROIT GANGS
THE HOUSE COMMITTEE + PLAY FOOTBALL + PLAY TENNIS + ENJOY ANTHROPOLOGY + A	9C4 WHO AM I
HE IS WONDERING WHY HE HAS TO LEARN TO PLAY THAT STUPID THING. HE WILL	5D1 WILLIAMSON UNDERGRAD
SING THE LAST LINK IS BROKEN AND PLAY THE BATTLE OF PRAGUE ON IT. THE	6B2 HUCK FINN
NOW THIS IN OUR CASE IS TRUE. I PLAY THE DRUMS MYSELF. WE HAVE FIGURED	2D1 PEACE CORPS
GAUNAAN. THEY BEAT THEIR GONGS. THEY PLAY THE GONGS ON THE WAY TO CUT WOOD.	8E1 IFUGAO

tedious, but we hope to have a considerably improved accuracy in our mappings within a year or so.

Some of our past content-analysis research has been quite relevant to the testing theme of this conference. Marshall Smith (7) of our group at Harvard has been particularly interested in applications to education and has begun to explore the use of the *Inquirer* for analyzing themes in college application essays, in predicting later college performance. Another of his projects has been to use *Inquirer* techniques to develop measures of "readability." In this paper, however, I would like to focus attention on the direct bearing of our work to the dialogue of education.

Our initial experience with scoring as an interactive process began when we put our need-achievement scoring system on a time-shared typewriter. Figure 6 presents an example protocol from one subject. In this case, the subject is seated at the typewriter and the directions for writing the story are presented to him. The subject then types his story. As soon as he is finished, he presses the return key on the typewriter twice and the computer immediately gives him an analysis of the story, first giving a summary of the amount of need-achievement present, and then giving a sentence-by-sentence analysis (in this case for each of the four sentences in the story), showing where in the story need-achievement was found.

It is not difficult for a person sitting like this at a typewriter to quickly learn what kinds of stories will be scored by the computer as examples of need-achievement. McClelland (3) has proposed that learning to write need-achievement stories is a helpful step towards acquiring need-achievement itself. For the complete novice, we might expand our directions and include some initial examples of what are and what are not achievement themes. As the subject types a number of stories, the computer can easily check whether they are only a stereotyped subset of the larger variety of possible need-achievement themes. If all the stories do fit a stereotype, the computer can then give some broader, possible examples and encourage the subject to try a wider variety of stories.

Our interactive scoring fulfills some of the basic elements of teaching-machine principles. The feedback is immediate and comes at the end of each story, a logical point for feedback. But note a fundamental difference: In most instances of teaching machines, the machine offers the subject a fixed number of alternatives, and the subject must choose one of these. Here the role of dominance in

Figure 6

Need-Achievement Scoring:
Interactive Procedure on a Time-Shared Computer Typewriter

	r nach W 1529.8 TAT STORY-SCORING PROCEDURE
Directions typed by computer	WRITE A SHORT STORY TO THE PICTURE TO YOUR LEFT. DESCRIBE WHAT IS HAPPENING, WHO THE PERSON IS, WHAT HAPPENED IN THE PAST, WHAT IS BEING THOUGHT, WHAT IS WANTED, WHAT WILL HAPPEN. PRESS RETURN KEY TWICE WHEN YOU HAVE FINISHED YOUR STORY.
Story typed by subject	This man wants to become a doctor. He works very hard to get enough money to go to medical school. He finally gets in and fails all his courses. Unhappy, he drowns his sorrows in drunkenness.
Analysis typed by computer	**SUMMARY** THIS DOCUMENT CONTAINS ACHIEVEMENT IMAGERY. SUBCATEGORIES ARE ... NEED = 1 INSTRUMENTAL ACTIVITY = 1 GOAL ATTAINMENT = 1 GENERAL BLOCK = 1 AFFECT = 2 THEME = 1 SUMMARY EACH SENTENCE. 1) NEED + TO-BE + ROLE-POS + N-ACH-PRESENT + 2) VERB-POS + ADVERB-POS + ADJ-WEAK-POS + INST.-AID + N-ACH-PRESENT 3) FAILURE + NO-ACH + 4) AFFECT-NEG + AFFECT-NEG + N-ACH-PRESENT + END OF THIS ANALYSIS, READY FOR ANOTHER STORY.

interaction is much more balanced. The subject is typing a story of his own choosing, and the machine is responding to the subject. One does not have to watch many subjects to know that there is considerable pleasure for the student in having the machine respond to his ingenuity rather than vice-versa. This reversal of initiative and dominance in the role of interaction exemplifies a point emphasized

in the speech we have just heard by Frank Keppel. Rather than "load" the student with information, the system should be designed to develop his abilities of reason and inquiry. One key factor in making this possible is that the subject has the freedom to express himself in his own native language at the typewriter.

As a further example of who leads whom in the dialogue with a computer, consider the interaction in Figure 7 where the computer is playing the role of a psychotherapist, using the "Eliza" system developed by Weizenbaum (10). Here the computer is responding to subject matter presented by the subject, offering direction when it fits into the context of the discussion.

Figure 7

In Which a Distraught Young Lady Consults the Machine

small letters: young lady

CAPITAL LETTERS: MACHINE

men are all alike

IN WHAT WAY

they're always bugging us about something or other

CAN YOU THINK OF A SPECIFIC EXAMPLE

well, my boyfriend made me come here

YOUR BOYFRIEND MADE YOU COME HERE

he noticed that i'm depressed much of the time

I AM SORRY TO HEAR YOU ARE DEPRESSED

he's right. i am unhappy

DO YOU THINK COMING HERE WILL HELP YOU NOT TO BE UNHAPPY

i need some help, that much is clear

WHAT WOULD IT MEAN TO YOU IF YOU GOT SOME HELP

perhaps i could learn to get along with my mother

TELL ME MORE ABOUT YOUR FAMILY

my mother takes care of me

WHO ELSE IN YOUR FAMILY TAKES CARE OF YOU

my father

YOUR FATHER

you are like my father in some ways

WHAT RESEMBLANCE DO YOU SEE

you are not very aggressive, but i don't think you want me to notice that

WHAT MAKES YOU THINK I AM NOT VERY AGGRESSIVE

you don't argue with me

WHY DO YOU THINK I DON'T ARGUE WITH YOU

you are afraid of me

DOES IT PLEASE YOU TO BELIEVE I AM AFRAID OF YOU

it pleases my father to think i am afraid of him

WHAT ELSE COMES TO MIND WHEN YOU THINK OF YOUR FATHER

bullies

DOES THAT HAVE ANYTHING TO DO WITH THE FACT THAT YOUR BOYFRIEND MADE YOU COME HERE

Such conversational dialogues can require rather complex language analysis. A first step is the categorization of words and phrases including the necessary contextual identification routines such as we are developing for our content-analysis dictionaries. Various pattern routines are then needed to identify the message, including a provision for the computer to make inquiries for clarification if it is unable to make a satisfactory classification. Additional steps are needed to maintain a satisfactory logic of conversation. Often the computer has to create a file of attitudes expressed by the subject so that it can check for inconsistencies or refer back to incomplete topics, or ask about relationships between topics. Moreover, if the dialogue ranges across topics, the computer's response set must be adequately grouped into topics so that the computer chooses responses not only appropriate to the particular dialogue but to the topic as a whole. Responses should often maintain the actual words employed by the subject: For example, the computer in Figure 7 uses the patient's actual words *boyfriend* and *depressed* rather than alternative words with the same meaning. These and other problems have been worked on by Weizenbaum at M.I.T. and Colby and Enea at Stanford (2).

In many instances, the computer typewriter can completely reverse its role from that of dominant teacher to that of instructive assistant. For example, a number of attempts have been made to have the computer respond to natural-language commands (9). Similarly, attempts have been made to have th· · ·puter respond to the questions of a doctor diagnosing a disease or a Sherlock Holmes cross-examining a witness (6). Figure 8 presents a set of directions from a subject using Weizenbaum's latest "Eliza-OPL" system (11). In this interaction, the computer is contextually responding to a series of directions from a subject. Such kinds of learning become another resource in the wide variety of possible interaction patterns.

Figure 8

Student Giving Directions to the Machine

small letters: student

CAPITAL LETTERS: MACHINE

please tell me what the sum of 5 and 4 is

IT'S 9.0

good, call that a

VERY WELL.

can you tell me what the product of a and 5 is

IT'S 45.0

call it b

VER\ WELL

let f(x,y) be x**2+y**2

ALL RIGHT

please compute f(a,b)

IT'S 2106.0

what is the product of 45 and 45

IT'S 2025.0

what is the difference between that and a

IT'S 2016.0

call this z

How good is a conversation with a computer? In certain cases it can be deceptively convincing. An oft-mentioned criterion is known as "Turing's test." As Abelson (1) has pointed out, Turing's test has both a simple and a complex form, but let us consider a simplified situation. Consider a person who is sitting at a typewriter and does not know whether his typewriter is connected to a computer or to another person sitting at a typewriter. Can the person tell whether he is communicating with a human being or a computer? Actually, telling them apart can become difficult. McGuire (4), when he put subjects through a full hour of therapy with the computer, found that 62 percent of the subjects were convinced that they were talking to a person at the other end, 21 percent were uncertain, and only 17 percent believed they were communicating with a machine. But we should share a little secret that deceived some of the most sceptical subjects. Usually when a typewriter is being controlled by a computer, it types evenly and rapidly like a mechanical teletype. Weizenbaum and McGuire arranged to have the machine type hesitantly and irregularly, to make occasional errors and back up to fix them, all at a speed of a very amateur typist. This trick alone can be enough to convince one that there must be a person at the other end. Although the computer made inappropriate remarks occasionally, the subjects seemed surprisingly willing to overlook them. All in all, however, the quality of the computer's responses was generally quite satisfactory.

The development of interaction procedures using language will require the cumulative contributions of many people. Just as our development of disambiguation routines described above is expected to serve a number of different content-analysis dictionaries, so too the increased sophistication of many-to-few categorization procedures can be drawn on by different investigators in developing more and more complex pattern analyses. The task is too complex for each investigator to start from scratch simply with raw text and a raw computer. Instead, he will need to cumulatively borrow on the previous work done by others. Such borrowing is essential if we are to get on with our work and to focus on issues rather than details.

While the previous examples in this paper demonstrate possibilities, we would hardly recommend, for instance, that at present you prepare your TAT's for need-achievement scoring on a computer. If the data are to be processed as in Figure 1, the text of the stories must first be punched on IBM cards, and the cards themselves should be verified. This punching phase alone can take more time than is necessary to

score stories for need-achievement by hand. Similarly, our interaction mode, as shown in Figure 6, is based on the assumption that the subject can type, and at this stage of our technology, the typewriter happens to be attached to one of the most powerful and sought-after computers in the country, Project Mac. Essentially, we are ready for demonstrations and limited research projects, not mass scoring of language material.

On the other hand, prospects for the future are rapidly improving. What took us an hour to process on the IBM 709 computer in 1961 can now, with improved computers and improved programs, be processed in less than three minutes on the IBM 7094. As computer memories become larger, we will be able not only to monitor one or two complex scoring procedures at once, but also make a large variety of continuous monitorings. For example, the script may call for a lesson in geography. While the lesson is going on, the computer can make continual scorings of negativity, redundancy, and other cues from the student to decide whether it should break out of the role of geography teacher and consider some other aspects of its relationship to the student. The student does not simply interact with a segmented series of programs but rather develops an interactive relationship with the machine while the machine builds up an extensive file about him from their past dialogue.

It is now almost 10 years since Rath, Andersen, and Brainerd (5) at IBM did their early teaching-machine studies in which they programmed a computer to give lessons in binary arithmetic. The sophistication of that early procedure was quite remarkable. The sequence of the lesson was selected on the basis of past errors. Upon completing the lesson, the subject was given a test in each area he had studied, the test itself being tailored to his past performance. Given the limited application that has since been made of such early sophistication, the promises of this paper may be viewed with considerable scepticism. The programs we have discussed may have a certain Alice in Wonderland quality about them, with the sceptic correctly saying, "But you're only a deck of cards, punched cards at that!" In reply, let me close by asking you to consider one last example (Figure 9), in which a psychiatrist, presumably somewhat worried about technological unemployment, meets his mechanical counterpart. While the computer can analyze language information, while it can make inquiries when information comes in unanticipated ways, while it can construct maps of incoming information and make checks for inconsistencies, while it

can adjust its vocabulary to the vocabulary of the person to whom it is talking, while it can compose syntactically complex answers, we certainly would not want to give the machine credit for being original or "understanding" in any sense of the word. Or would we? The machine as a teacher will always have its weaknesses. But perhaps we can expect children to be as tolerant of machines as they are of us.

Figure 9

In Which a Human Psychiatrist Meets His Mechanical Counterpart

small letters: psychiatrist

CAPITAL LETTERS: MACHINE

doctor, i have terrible news

YOU SAY YOU HAVE TERRIBLE NEWS

yes, there is a computer program at m.i.t. that'll ruin us

WHY DO YOU SAY THAT

i have seen it working

YOU SAY YOU HAVE SEEN IT WORKING

it's frightening—you must see it

YOU THINK I MUST SEE IT

we are both in the same boat with respect to that computer

LET'S CONTINUE TO TALK ABOUT YOU—SHALL WE

you are a comfort

WHAT MAKES YOU THINK I AM A COMFORT

at a time like this you want to talk about me

WHY DO YOU THINK I WANT TO TALK ABOUT YOU

because you share my feelings about that damned computer

WHAT DOES THAT SUGGEST TO YOU

that, of course, you hate computers too

YOU THINK I HATE COMPUTERS TOO

why not, they're hardly human, are they

PERHAPS THAT'S WHAT WE ARE HERE TO FIND OUT

no computer could converse as you can, for example

OH

i can't understand how anyone could be fooled by a mere machine

CAN'T YOU REALLY

for example, could a computer in a conversation like this raise a new topic

YOU SEEM TO HAVE AVOIDED SPEAKING OF YOUR PARENTS ALTOGETHER

precisely as you just did

YOU THINK I JUST DID

yes, and that's what shows the difference between a man and a computer

I SEE

REFERENCES

1. Abelson, R. P. Chapter on computer simulation in *Handbook of social psychology* (revised). Reading, Mass.: Addison-Wesley (in press).

2. Colby, K. M. & Enea, H. Heuristic methods for computer understanding of natural language in context-restricted on-line dialogues. *Mathematical Biosciences* (in press).

3. McClelland, D. C. Toward a theory of motive acquisition. *American Psychologist*, 1965, 20, 5. 325.

4. McGuire, M. T., Lorch, S., Quarton, G. C. Man-machine natural language exchanges based on selected features of unrestricted input. The Stanley Cobb Laboratory for Psychiatric Research, Massachusetts General Hospital and Harvard Medical School, 1966. Mimeographed.

5. Rath, G. J., Anderson, N. S., and Brainard, R. C. The IBM Research Center teaching machine project. *Automatic teaching: the state of the art.* Galanter, E. H. (Ed.). New York: Wiley, 1959. Pp. 117-130.

6. Simmons, R. F. Answering English questions by computer: a survey. *Communications of the A.C.M.*, 1965, 8. Pp. 53-69.

7. Smith, M. S., School of Education, Harvard University. Research in progress.

8. Stone. P. J., Dunphy, D. C., Smith, M. S., & Ogilvie, D. M. with associates. *The general inquirer.* Cambridge: M.I.T. Press, 1966.

9. Swets, J. A. & Feurseig, W. Computer-aided instruction. *Science*, 1965, 150, 572-576.

10. Weizenbaum, J. *Eliza:* a computer program for the study of natural communication between man and machine. *Communications of the A.C.M.*, 1966, 9, 36-45.

11. Weizenbaum, J. Personal communication.

The Natural-Language Approach to Psychometrics

CARL E. HELM
Educational Testing Service

These comments represent a first attempt to organize some ideas about the methodological status of computer programs which simulate scientific theories. In brief, my major conclusion is that measurement procedures are a kind of grammatical analysis, and that theories of structural linguistics encompass both classical measurement theories and equivalent theories for computer models.

On a number of occasions over the past few years, acting in the role of computer programmer and systems analyst, I have been confronted with problems which people (including myself) wanted to solve by using a computer. The most interesting of these problems have always been the ones that were not well defined. Over the years a strategy has evolved which we have been able to use in applying computers to the solution of such problems. We have found that we can help crystalize the definition of a problem by writing computer programs based on natural-language descriptions of that problem. For example, not long ago I sat down with an aerospace psychologist who said to me: "Doctor, I have a problem. The engineers in my company are using a computer to simulate the flight of a manned space mission in order to explore the consequences of various malfunctions. They have asked me to prepare the specifications for a 'black box' model of an astronaut so they can program into their system realistic estimates of human performance reliability under conditions of prolonged space flight. How do I go about setting up a model like this which can be the basis for a computer program?"

"How do you describe the behavior of an astronaut under conditions of space flight?" I asked.

He answered, "Well, for example, a man can't make very good two-point discriminations under the conditions of vibration which exist during blast-off."

After a few hours of conversation which consisted mainly of such descriptions of relevant details, we tentatively isolated three general attributes of the situation which could serve as organizing principles. These were: the *stressors* which can affect human performance—such things as vibration, weightlessness, and fatigue; the *tasks* which an astronaut must carry out at various points in the mission; and the *components of human performance*, such as fine motor coordination, computational ability, and visual discrimination, which enter into the execution of the tasks and which can be adversely affected by the stressors (2).

At this point, we began to consider whether we could write a computer program in terms of these organizing principles. For example, we could think of a routine which could estimate the reliability of two-point discrimination under conditions of vibration or weightlessness or fatigue. We could also think of other routines which could estimate the reliability of execution of tasks requiring gross motor coordination under similar conditions. Conceived of at this level of detail, however, the program very quickly became entirely too complex to be manageable.

To make a rather long story short, we eventually devised a system of programming which allowed the psychologist to conveniently specify any or all of the variables which might enter into a particular simulation, along with the various functions which defined the relations between the variables. This was accomplished by allowing the psychologist first to name the elements and functions and then to describe their attributes in rather simple tables. In effect, we devised a special-purpose programming language based on the natural-language description which he used to define a rather large class of procedures (3).

The initial statement of the problem by the aerospace psychologist was not so much a specification of a problem as a rationale for his motivations. His basic problem was to develop a workable representation of the intuitive concepts he used to organize his knowledge of the behavior of man in space. Initially, the only means available to communicate these intuitive concepts was to describe many particular instances using familiar language. My principle task as systems analyst and programmer was to isolate the structure of his intuitive system for organizing the relevant data. The organizing principles (the

stressors, tasks, and components of human performance) represented higher-order abstractions based on the many reports of particular instances. The fact that we assigned "meaningful" names to these conceptual structures was only a device to facilitate communication. The important point is that we were able to use these higher-order abstractions to produce simple but precise descriptions of the structure of his intuitive system. For example, we could say "stressors degrade the components of human performance which are required for the execution of tasks." "Tasks make varying demands on the components of human performance." "The reliability with which a task can be executed is a function of both the state of degradation of the related components of human performance and the demand of the task for those components." These statements provided us with descriptions of parts of the structure of his system. They made no reference to any instances but they provided a useful way of classifying and organizing observed data.

I have been through this kind of sequence with scientists in a number of disciplines—for example, with electrical engineers who wished to automate the analysis of com· ˙ x electromechanical circuits, with a sociologist who wants to simulate the growth and decay of population under a variety of conditions in order to test theories about family structure, with myself in constructing some simple n-ary choice learning models, with a linguist who is conducting a comparative analysis of 14 dialects of Bantu, and with clinical psychologists who wanted to have the computer produce interpretations of profiles of mental test scores.

I should like to discuss some of the features of the procedures which appear to me to be common throughout all these investigations, and indicate how I think these procedures might have some implications for scientific methodology in general. I am concerned here with the characteristics of psychological theories and with the characteristics of research strategies which lead to the development of well-formed psychological theories.

My first observation is that psychometricians and mathematical psychologists construct theories to describe other theories which psychologists construct to organize and describe psychological phenomena. It is important to consider the distinction between these levels of representation in the analysis of the methods and research strategies of psychometricians. A psychometric or mathematical theory constitutes a more or less rigorous and elegant description of a

set of intuitive notions. This set of intuitive notions exists as a natural-language description of a system for organizing and talking about more or less observable phenomena. Thus, we have a hierarchy of theories which describe other theories in increasingly rigorous terms.

My second general observation is that measurement theories or, more generally, descriptions of measurement procedures provide essential information about the structure of theories. To put it in a different way, if we think of theories as special-purpose languages, we may then consider measurement theories as the grammars of scientific theories. If we are given a completely specified grammar of such a special-purpose language we can always decide whether or not its sentences—that is, the scientific hypotheses—are well formed. For example, if the description of a measurement procedure implies a unidimensional ordinal system, then the following statement is ungrammatical—i.e., not well formed with respect to the measurement procedure: a is "x-er" than b and b is "x-er" than c and c is "x-er" than a. (Of course, the entities a, b, and c must be identified as well as the relations *and* and x, and this identification must occur in yet another language if my discussion is to be precise.)

Beyond this, the question of whether a *well-formed* statement is true or false has to do with the "grammar" or structure imposed on relevant data by the empirical measurements, and with the isomorphism between the structure of the empirical data and the grammar of the theory. The description of measurement procedures makes no reference to the meanings of the measurement or of the things measured. The choice of measurement procedures, however, does require an understanding of what it is that is going to be measured, and of the purposes of the measurements. Indeed, only after identifying the structure of a theory are we in a position to consider the meaning of the measurements.

My third observation is that communication between psychologists and psychometricians must utilize natural language as its basis. The psychologist must communicate his descriptions of his intuitive notions to the psychometrician using natural language, or the language of everyday discourse, even though natural language is a system of signs and symbols for which no complete grammar exists. It is true of course, that the psychologist will usually attempt to define his terms more or less rigorously, restricting his discourse to some subset of natural language in order to achieve greater precision. Nevertheless, natural language is the essential vehicle for the initial description

of any theory, and it is the business of the psychometrician or mathematical psychologist to translate natural-language descriptions of theories into formal descriptions which are complete and consistent. We can then investigate the properties of such descriptions by treating them as languages and investigating their grammars.

Let me summarize these observations as follows:

1. Scientific theories consist of a hierarchy of descriptions resting ultimately on a scientist's natural-language description of his intuitive system for organizing the phenomena of interest to him.
2. A scientific theory may be defined as a complete and consistent description of a system. The description of measurement procedures is a necessary part of any scientific theory since it provides essential information about the structure of the theory. A well-formed theory is tested by observing the isomorphism between its structure and the empirically determined structure of relevant data.
3. At the primitive levels, the description of a scientist's intuitive system for organizing phenomena necessarily occurs in natural language. The basis for this observation is largely personal experience although it is apparent that until some description exists, it is not possible to isolate the structure of the theory, which must be done before further abstractions can be made.

Given these general observations, I would like to submit that large-scale computers are an essential tool for the psychometrician or mathematical psychologist as he goes about the business of constructing rigorous, elegant, formal descriptions of psychologists' intuitive natural-language descriptions. Most mechanical languages with which computers are programmed have been analyzed syntactically (that is, complete formal descriptions of their grammars exist); therefore, a computer program which simulates a psychological theory, and accordingly has a structure isomorphic to that of the theory, is susceptible to syntactic analysis (which of course may not be a simple matter).

It is a straightforward matter to write a computer program which is isomorphic to a *well-formed* mathematical theory. Such programs are written in order to explore the consequence of theories in particular instances. When we do a factor analysis on a computer, we are exploring implications of a complex linear mathematical model as it

applies to a particular state of affairs—i.e., the input data. The extent to which the structure of the psychological theory and the structure of the mathematical system match is the extent to which the results of such analyses are meaningful. (It is evident that the psychological theory underlying factor analysis has a more complex structure than the mathematical system which is its model, but we don't as yet know how to program the additional complexity.)

It may not be a simple matter to write a computer program to describe the theory which, say, a practicing clinical psychologist must apply in order to interpret mental test data, but a programmable theory is essential if a practicing clinical psychologist is to produce *unambiguous* interpretations of mental test data. If he does not have access to a sufficiently well-structured theory, then it is not possible for him to interpret data in any consistent and meaningful fashion.

Just as it is possible to write a computer program which is isomorphic to the mathematical description of a psychological theory, so it is possible to write a computer program which is isomorphic to a natural-language description of a psychological theory, given a large, fast computer. The speed and size of the computer are essential from a practical point of view if one is to explore the consequences of any interesting natural-language descriptions within a reasonable period of time. The facts of mechanical life are that the execution of any process which represents behavioral phenomena described in terms of a natural language entails extraordinary numbers of elementary computer operations. Moreover, there is no way to short-circuit this process without losing the rigor which is essential to discover the flaws in the natural-language descriptions of the psychologist's theories.

Finally, let me discuss a few implications of these observations.

First, the representational view of measurement—that is, measurement conceived of as the specification of the relation between the structure of numerical systems and the structure of empircial systems (1) is too limited in its scope to include many theories of interest to psychologists. Many such theories are, however, representable by means of computer programs.

Second, we can appropriately enlarge the scope of the concept of measurement by considering measurement procedures as equivalent to grammatical analyses. This, of course, requires us to consider a complete and consistent scientific theory to be a special kind of language, in particular a language with a complete grammar. Third,

there must be an upper bound to the structural complexity of scientific theories describable in terms of the "representational" view of measurement. This upper bound (for example, that such theories are equivalent at most to phrase-structure languages) is not sufficient for many theories of interest to psychology. Finally, large, fast computers are an essential tool for the development of elegant psychological theories based on the intuitions and insights of the experienced psychologist.

It is my feeling that the philosophy underlying these observations is consistent with that which motivated Hamming's comment, "The purpose of computing is insight, not numbers," and I am convinced that we have at hand the foundations for a vastly expanded psychometric theory which will provide more insights and fewer numbers, to the general satisfaction of even the most mathematical among us.

REFERENCES

1. Adams, E. W. On the nature and purpose of measurement. Technical report No. 4. *Measurement theory* and *mathematical model reports.* Portland: University of Oregon, April 1966.

2. Marton, T. & Helm, C. The degraded man study, General Electric Missile and Space Division Report, November 1963.

3. Marton, T., Helm, C., Green, B., & Martin, M. The assessment of human performance for the analysis of space missions. (To appear in *Behavioral Science.*)

Grading Essays by Computer: Progress Report*

ELLIS BATTEN PAGE**
University of Connecticut

My role today is pleasant, since it provides a chance to talk to colleagues in measurement about this great enthusiasm of ours, the grading of essays by computer. But the role is rather complicated. Since this area of measurement is in many ways very new, and since for the past 18 months we have been too busy exploring it to publish much, many of our colleagues will know very little about it—and some of that will probably be wrong! Therefore, I should explain some of the basic rationale of our effort. Some people know a great deal about certain aspects of our work, and to these my introduction may seem all too familiar. For these I hope the second portion of this paper, in which I discuss newer strategies and some results of recent work, might be more rewarding. And perhaps all will wish to speculate about the future of such activity. Therefore I shall give some description, clouded though it may be, of the view from where we stand.

*We owe thanks to many people not otherwise acknowledged in this address: to officers at the College Entrance Examination Board and the Research Branch of the U.S. Office of Education, for financial support from 1965-67; to my colleagues at Connecticut who have participated: Gerald and Mary Ann Fisher, Arthur Daigon, Herbert Garber, Deiter Paulus, Charles McLaughlin, Kenneth Wilson, H. Fairfield Smith. To William McColly, State University of New York, Oswego. To certain persons at Harvard: Allan B. Ellis, Marshall J. Smith, Jr., Dexter Dunphy. To Junius A. Davis and Paul Diederich, Educational Testing Service; J. William Asher, Purdue University; Paul H. Lohnes, Project TALENT; Fred Kerlinger, New York University; Walter and Sally Sedelow, University of North Carolina; Leslie McLean, Ontario Institute for Studies in Education.

**Professor of Educational Psychology and Director of the Bureau of Educational Research. On part-time leave during 1966-67 as Visiting Scientist to the Massachusetts Institute of Technology Computation Center.

Many colleagues and friends have been involved in this project. Of those who do happen to be on this platform, Julian Stanley has given his usual priceless encouragement from the beginning, and Carl Helm recently visited us in Connecticut and provided some valuable ideas. Phil Stone has not been active with me, and we have not used his impresssive strategies, but this work owes him a debt he is probably not aware of. It was a presentation of his at Harvard in late 1964 that started me tossing and turning and losing sleep about the whole field of essay grading by computer. *Why not?* I kept wondering, and little by little the necessary research design began to emerge.

Once you ask the question *Why not?* and begin investigating this field, you might be astonished at how rich the background material is—and how much of it is virtually unknown to psychologists. You find yourself at a disciplinary interface, involving not only psychometrics and statistics, but also linguistics, English composition, computer science, educational psychology, natural-language analysis, curriculum, and more. This interdisciplinary aspect sometimes makes communication more complicated, since what will seem elementary to one segment of an audience will seem impossibly recondite to another.

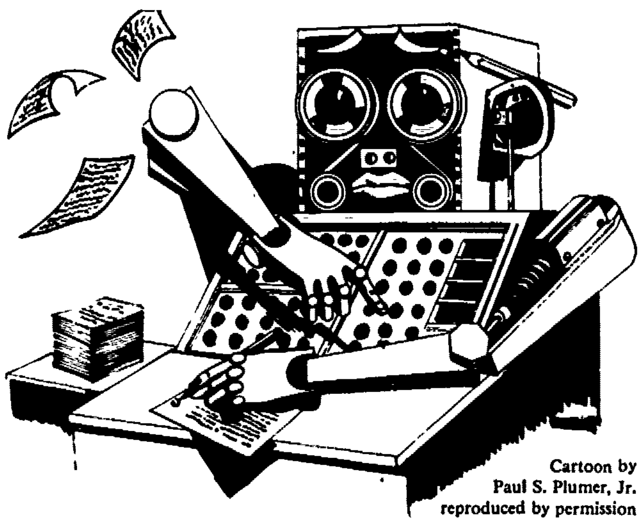
The reactions to our effort have been fantastic. Our work has attracted a certain amount of attention in national news media, ranging from the favorable to the outraged. On one hand, there is the inevitable disbelief and dread of occupational replacement, and, perhaps, something still deeper. As Jay Davis said at APA a year ago, in a profound comment, the real threat of such computer analysis is that it may expose our human simplicity. My own favorite press reaction (possibly because I am a former English teacher) is one in a recent issue of a teachers' journal. Figure 1 is an illustration of this monster, an essay-grading machine.

See the brute machine at work, with its flailing arms (apparently losing some papers) and glaring eyes. I especially like its thick sensual lips. The author of the accompanying article (2) wrote of a "cynical dehumanizing which, fully achieved, would reduce language to the terrifying 'duck-speak' of Orwell's nightmare world." He claimed that *human* essay grading is good *because* it is subjective—that is, because one teacher will not agree with another!

On the other hand, there have been many reactions to our work which were embarrassingly favorable, with such a wistful optimism

Figure 1

Monster Machine



Cartoon by
Paul S. Plumer, Jr.
reproduced by permission
of *The Maine Teacher.*

**A Voice Against
Computer Correction Of Themes**

about what we could do to help that some instructors at the University of Connecticut have called our bureau about grad'ng their midterm exams!

The reality of our study, of courses, lies somewhere between the impossible and the operational. We are not grading routing exams and will not be next year either. There are some good, hard problems on the way to this goal, but we feel the future is bright. Let us see whether, after having been brought up to date, you will share this optimism.

We may conceive our general problem as resembling Figure 2. As the column heads indicate, we are interested in *content* (what is said) and in *style* (the way it is said). Obviously, these columns are not mutually exclusive, but the simplification may be useful.

Similarly, the rows are not mutually exclusive either. But their general meaning must be mastered to understand what is being

attempted. The first row refers to the simulation of the human product, without any great concern about the way this product was produced. It refers to actuarial optimization, a pragmatic approach to the simulation of the behavior of qualified judges. The bottom row, on the other hand, refers to the master analysis of the essay, to the sort of knowledgeable and detailed description of the essay, and of its various parts, which might emerge when competent judges apply advanced analytic skills.

Figure 2

Possible Dimensions of Essay Grading

	I Content	II Style
A. Rating Simulation	I (A)	II (A)
B. Master Analysis	I (B)	II (B)

We have coined two terms to describe this difference. Since the top row is concerned with ap*prox*imation, we speak of the computer variables employed as *proxes*. Since the bottom row is concerned with the true in*trin*sic variables of interest, we speak of such variables as *trins*.

A trin, then, is a variable of intrinsic interest to the human judge, for example, "aptness of word choice." Usually a trin is not directly measurable by present computer strategies. And a prox is any variable measured by the computer, as an approximation (or correlate) of some trin such as the proportion of uncommon words used by a student (where common words are discovered by a list look-up procedure in computer memory).

So far in our investigations, we have concentrated on the top row of Figure 2. looking for actuarial strategies, seeking out those proxes which would be of most immediate use in the simulation of the final human product. This does not mean that we have no interest in the trins. But many people have a misguided view of simulation. They imagine that a more microscopic strategy really does things in

some "human" way. This is usually an illusion. The principal difference between strategies is often just in the size of bite, in the temporal scope of behavior chosen to be the target. For example, suppose we tried to imitate human judges at a number of points along the behavioral continuum, picking up the essay, for example, then reading the title, and so on until we reach the eventual decision concerning overall grade. Suppose we imitated 10 such different choice points en route to this grade. That would perhaps seem a more accurate simulation of the process. Within each of these 10 behavioral blocks, however, we would still be using algorithms which had little to do with the "real" human procedures. In other words, *all* computer simulation of human behavior appears to be product simulation rather than process simulation. And the two fields of psychological simulation, on the one hand, and artificial intelligence, on the other, are not necessarily so very far apart as some would claim.

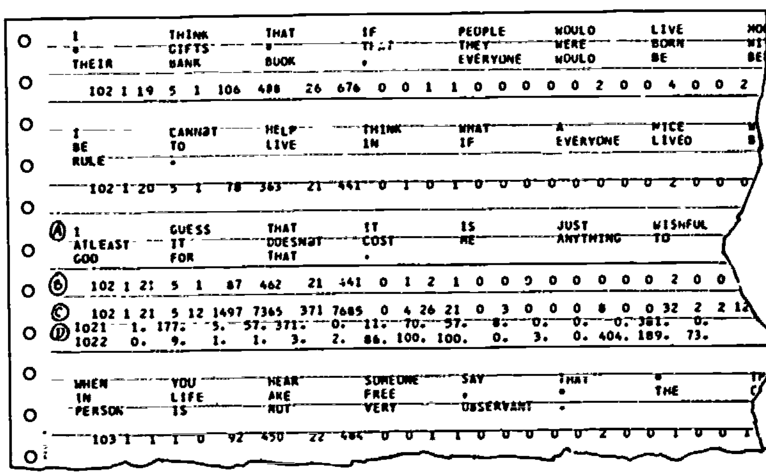
In adopting the overall, terminal strategy described here, we have not abandoned a goal of more refined analysis, nor of simulation closer to the human process itself. Indeed, we are pushing in much more deeply, as my later comments will suggest. But for the first attempts, we evolved a general research design, which we have more or less followed to date:

1. Samples of essays were judged by a number of independent experts. For our first trial, there were 276 essays written by students in grades 8 to 12 at the University of Wisconsin High School, and judged by at least four independent persons. These judgments of overall quality formed the trins.
2. Hypotheses were generated about the variables which might be associated with these judgments. If these variables were measurable by computer, and feasible to program within the logistics of the study, they became the proxes of the study.
3. Computer routines were written to measure these proxes in the essays. These were written in FORTRAN IV, for the IBM 7040 computer, and are highly modular and mnemonic programs, fairly well documented.
4. Essays were prepared for computer input. In the present stage of data processing, this means that they were typed by clerical workers on an ordinary key punch. They were punched into cards which served as input for the next stage.

5. The essays were passed through the computer under the control of the program which collected data about the proxes. The output is shown in Figure 3.

Figure 3

PEG-IA Output



In Figure 3, we can see a tear-off from the output of our program (PEG-IA). Line A shows the way a sentence from the student essay is rewritten in 12-character double-precision computer "words" and stored in memory. Line B shows the summary of data for that sentence just analyzed. The first number is the essay identification. The other numbers on Line B are some counts from that sentence. Line C shows a summary of these counts, across sentences, for this whole essay. On Line D are these measures transformed in a number of simple ways and ready for input into the final analysis.

6. These scores were then analyzed for their multivariate relationship to the human ratings, were weighted appropriately, and were used to maximize the prediction of the expert human ratings. This was all done by use of a standard multiple regression package.*

---

*Since some of the variables were grossly non-linear and non-normal, and will have presumably interesting interactions, Deiter Paulus and I are currently studying desirable score transformations.

Table 1*

Variables Used in Project Essay Grade I-A
for a Criterion of Overall Quality

A. *Proxes*	B. *Corr. with Criterion*	C. *Beta wts.*	D. *Test-Ret. Rel (Two essays)*
1. Title present	.04	.09	.05
2. Av. sentence length	.04	−.13	.63
3. Number of paragraphs	.06	−.11	.42
4. Subject-verb openings	−.16	−.01	.20
5. Length of essay in words	.32	.32	.55
6. Number of parentheses	.04	−.01	.21
7. Number of apostrophes	−.23	−.06	.42
8. Number of commas	.34	.09	.61
9. Number of periods	−.05	−.05	.57
10. Number of underlined words	.01	.00	.22
11. Number of dashes	.22	.10	.44
12. No. colons	.02	−.03	.29
13. No. semicolons	.08	.06	.32
14. No. quotation marks	.11	.04	.27
15. No. exclamation marks	−.05	.09	.20
16. No. question marks	−.14	.01	.29
17. No. prepositions	.25	.10	.27
18. No. connective words	.18	−.02	.24
19. No. spelling errors	−.21	−.13	.23
20. No. relative pronouns	.11	.11	.17
21. No. subordinating conjs.	−.12	.06	.18
22. No. common words on Dale	−.48	−.07	.65
23. No. sents. end punc. pres.	−.01	−.08	.14
24. No. declar. sents. type A	.12	.14	.34
25. No. declar. sents. type B	.02	.02	.09
26. No. hyphens	.18	.07	.20
27. No. slashes	−.07	−.02	−.02
28. Aver. word length in ltrs.	.51	.12	.62
29. Stan. dev. of word length	.53	.30	.61
30. Stan. dev. of sent. length	−.07	.03	.48

*Number of students judged was 272. Multiple R against human criterion (four judges) was .71 for both Essay C and Essay D (D data shown here). F-ratios for Multiple R were highly significant.

The resulting data, summarized briefly in Table 1, suggest the nature and performance of some of the early proxes. Column A gives the names of the proxes employed. Some were based upon careful analysis and hypothesis. Others (such as the less common punctuation marks) were recorded only because they were naturally produced by the computer programs. Column B shows their correlation with the criterion, the overall human judgment. Column C shows the beta weights for predicting the criterion, when all 30 proxes were employed. And Column D shows what could be called the "test-retest" reliability of the proxes. These coefficients in Column D are based on two different essays on different topics written about a month apart by the same high school students.

The overall accuracy of this beginning strategy was startling. The proxes achieved a multiple correlation coefficient of .71 for the first set of essays analyzed and, by chance, achieved the identical coefficient for the second set. Furthermore, and this is, of course, important, the beta weightings from one set of essays did a good job of predicting the human judgments for the second set of essays written by the same youngsters. All in all, the computer did a respectable "human-expert" job in grading essays, as is visible in Table 2.

Table 2

Which One is the Computer?

Below is the intercorrelation matric generated by the cross-validation of PEG I

	Judges				
	A	B	C	D	E
A		51	51	44	57
B	51		53	56	61
C	51	53		48	49
D	44	56	48		59
E	57	61	49	59	

Here we see the results of a cross-validation. These are correlations between judgments of 138 essays done by five "judges," four of them human and one of them the computer. The computer judgments were the grades given by the regression weightings based on 138 *other* essays by *other* students. This cross-validation, then, is very conservative. Yet, from a practical point of view, the five judges are indistinguishable from one another. In eventual future trials, we expect the computer will correlate *better* with the human judges than will the other humans. But even now, we feel that A. M. Turing, who recommended the "different test" as a good trial of the presumably intelligent machine, might well be pleased.

However useful such an overall rating might be, we of course still wish greater *detail* in our analysis. We have therefore broadened the analysis to five principal traits commonly believed important in essays. These traits are adapted partly from those of Paul Diederich. For our purpose they may be summarized as: *ideas*, *organization*, *style*, *mechanics*, and *creativity*. We had a particular interest in *creativity*, since some have from the beginning imagined that our study would founder on this kind of measure. "You might grade mechanics all right," someone will say, "but what about originality? What about the fellow who is really different? The machine can't handle him!"

Therefore, this summer we called together a group of 32 highly qualified English teachers from the schools of Connecticut to see how *they* would handle *creativity* and these other traits. Most had their master's degrees and extensive experience in teaching high school English, and all had the recommendation of their department chairmen. Each of 256 essays was rated on a five-point scale on each of these five important traits, by eight such expert judges, each acting independently of the other.*

The teacher ratings were then analyzed. The results, which were calculated by Jim Roberge and others, are shown in Table 3. It is clear from Table 3 that the essay and the trait contributed significant variances, and so did the trait-by-essay interaction, which is perhaps the clearest measure of the ipsative qualities of the profile. To

*For a study of this size, the random assignment of essays to judges, to periods, and to sessions turned out to be a formidable task, and once again the computer was called in. This was our first experience of using the computer to *design* a study as well as analyze one. We discovered some interesting things in the process and recommend this idea to the consideration of others.

Table 3*

Trait by Essay Interaction

Source	SS	df	MS	F
Between judgments	8,230.305	2,047		
Between essays	3,791.293	255	14.868	6.002
Error between	4,439.012	1,792	2.477	
Within judgments	3,564.414	8,192		
Between traits	84.212	4	21.053	56.412
Trait x essay	805.089	1,020	.789	2.115
Error within	2,675.113	7,168	.373	
Total	11,794.719	10,239		

*This table is based upon essay evaluation of July 1966, during which each of 256 essays was judged by eight different judges during eight different periods.

investigate each of these five trait ratings, then, the same 30 proxes were again employed, with the results shown in Table 4.

In our rapidly growing knowledge, Table 4 may have the most to say to us about the computer analysis of important essay traits. Column A, of course, gives the titles of the five traits (more complete descriptions of the rating instructions may be supplied on request). Column B shows the rather low reliability of the group of eight human judges, computed by analysis of variance. This is the *practical* reliability of these pooled judgments. We get higher reliabilities when we subtract from the error term the variances attributable to period, session, and judge; but it would be misleading to do so in this present comparison, since these adjustments were not made preparatory to the machine grading regression analysis.

Here in Column B it seems that *creativity* is less reliably judged by these human experts than are the other traits, even when eight judgments are pooled. And *mechanics* may be the most reliably graded of these five traits. Surely, then, *humans* seem to have a harder time with *creativity* than with *mechanics*.

Now what of the computer? Column C shows the raw multiple correlations of the proxes with these rather unreliable group judg-

ments. These were the coefficients produced by the standard regression program run by Paulus and myself. If a really fair comparison is to be made among the traits, however, the criterion's unreliability should be taken into account. And this results in the corrected multiple coefficients appearing in Column D. Here such difficult variables as *creativity* and *organization* no longer seem to suffer; the computer's difficulty is apparently in the criterion itself, and is therefore attributable to human limitations rather than to those of the machine or program. Column E simply shows the same coefficients after the necessary shrinking to avoid the capitalization on chance which is inherent with multiple predictors. Column E, then, exhibits what we might expect on cross-validation of a similar set of essays, if we were pre-

Table 4*

Computer Simulation of Human Judgments
for Five Essay Traits
(30 predictors, 256 cases)

A. *Essay Traits*	B. *Hum.-Gp. Reliab.*	C. *Mult. R*	D. *Corr. (Atten.)*	E. *Shrunk. Multi. R*
I. Ideas or Content	.75	.63	.72	.66
II. Organization	.75	.59	.68	.60
III. Style	.79	.67	.75	.70
IV. Mechanics	.85	.62	.67	.60
V. Creativity	.72	.61	.72	.64

*Col. B represents the reliability of the human judgments of each trait, based upon the sum of eight independent ratings, August 1966.

Col. C represents the multiple regression coefficients found in predicting the pooled human ratings with 30 independent proxes found in the essays by the computer program of PEG-IA.

Col. D presents these same coefficients, corrected for the unreliability of the human groups. (Cf. McNemar, 1962, p. 153.)

Col. E presents these coefficients, both corrected for human unreliability and shrunken to eliminate capitalization on chance from the number of predictor variables. (Cf. McNemar, 1962, p. 184.)

dicting a perfectly reliable set of human judgments.*

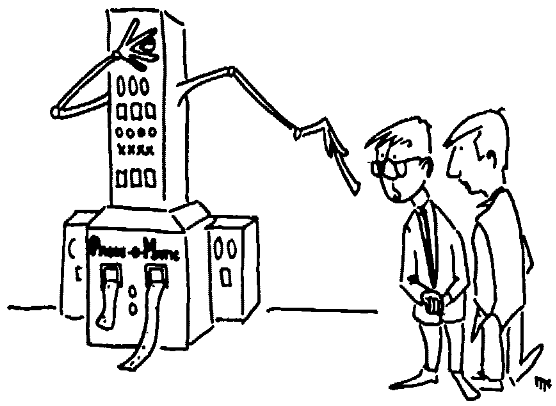
Now there are standard beginning questions which people almost inevitably ask at this point if our subject is new to them: What about the input problem? What about subject-matter grading? What about the student who tries to con the machine? What about detailed feedback to the student? And so on. These are all valid questions, and we have written our answers in the January issue of *Phi Delta Kappan* (1). For most people these answers appear to be satisfactory.

But we are not presenting the results here as a terminal achievement against which to measure this sort of work. On the contrary, this is a temporary reading taken in the middle of the research stream. In the meantime, we go on with other strategies. Don Marcotte, for example, has recently developed an interesting phrase analyzer and has discovered that clichés, as usually listed, are pretty irrelevant in such essay grading. We have this summer studied some problems of style, parallelism, and certain semantic questions. We are exploring various dictionary and parsing options which lie before us. Recently we located what may be the most promising parsing program and used it to run certain essays. There are some fascinating studies done by people in artificial intelligence and information retrieval, which may have something to offer in the near future. And we are interested in improving our statistical strategies as well. We are looking at the proxes themselves through factor analysis and stepwise regression. And then there is the question of extending the strategy to the humanities. One of the questions raised by scholars is whether it will handle various authors. A cartoon reflecting this question was printed in the *Phi Delta Kappan* and picked up by the *New York Times*. It is shown in Figure 4.

Notice that this machine, like the one shown in Figure 1, is anthropomorphic. It seems embarrassed about "flunking Hemingway," but is a lot nicer machine than the first one. Well, we are key-punching some passages from Hemingway and other standard authors to find out how the program handles them! In any case these present results are, as I pointed out above, the merest way station, but they may indicate to most of you, as they do to us, that workers in this field will not be wasting their time.

*We have just completed a computer run with other high school essays from an interesting study in Indiana by Anthony Tovatt and his colleagues at Ball State University. I shall leave the details for Tovatt to report. But it is an independent confirmation of the success of the computer strategy in grading student essays across a whole *profile* of essay traits.

Figure 4



Great Scot! It's just flunked Hemingway.

Cartoon by Margaret McGarr reproduced by permission of *Phi Delta Kappan.*

There are many tantalizing problems in such research. One of the greatest is the effort toward psychologically deepening the work and making it more humanoid in process. Of considerable relevant interest to us, and to workers in related fields, is the possible verbal education of a computer. The solution will probably lie *not* in trying to *program* all the linguistic responses to be made by the computer. Rather, the solution may consist in programming only a certain set of quasi-psychological procedures, designed to enable the computer to learn on its own (i.e.—to gain literary *experience*) by reading in and correctly processing a great amount of appropriate text, making use of automated dictionaries and other aids while doing so. We dream of producing, in other words, the well-read computer. Part of our success to date has occurred through allowing the computer itself, in the multiple regression program, to determine which analytic weightings are valuable. What we hope is that somehow an expansion of this strategy of computer education can be undertaken. This is a very hard problem but a fascinating one and a number of people, in one field and another, are very interested in it.

And finally, a statement of present methodological bias: We believe that the work should not surrender to the purist on the one hand, who might claim that permanent improvement can be made only by a thorough mastery of theoretical concepts. Nor to the complete empiricist on the other, who may conceive that trial-and-error activities,

with a poorly understood response surface, can lead to useful mastery of the underlying psychometric realities. No, a compromise would be more faithful to the professional history of those here in this room. Indeed, such a compromise between practical educational utility, on the one hand, and intriguing psychological and statistical depth, on the other, may be the very foundation on which our profession of measurement has flourished. In this new venture of grading essays by computer, competent measurement people, especially those with a love of language, should play an important role.

REFERENCES

1. Page, E. B. The imminence of grading essays by computer. *Phi Delta Kappan*, January 1966, 238-243.

2. Roby, K. E. A voice against computer correction of themes. *The Maine Teacher*, 1966, 27, 1, 19-20.

Participants

1966 Invitational Conference on T‹ F. Joblems

AARON, SADIE, Houston, Texas
ABBOTT, MURIEL M., Harcourt, Brace and World, Inc.
ABELES, SIGMUND, New York State Department of Education
ABRAHAM, ANSLEY A., Florida Agricultural and Mechanical University
ACHTERBERG, JAMES E., George School, Pennsylvania
ADAMS, RUTH ANN, Agency for International Development
ADLERSTEIN, ARTHUR, Rutgers—The State University
AHRENS, DOLORES F., Educational Testing Service
AKDEMIR, HASAN, New York Jniversity
AKERS, SUE B., Huntingdon College
ALEXANIAN, ALEXANDER J., Massachusetts State Department of Mental Health
ALEXANIAN, SANDRA, Boston University
ALCOZZINE, JANE A., New York State Department of Education
ALKER, HENRY, Educational Testing Service
AL' :N, ANTHONY, Boston College
ALLEN, W. PAUL, International Schools Services
ALMAN, JOHN E., Boston University
ALT, PAULINE M., Central Connecticut State College
ALVAREZ, EVANGELINA, College Entrance Examination Board, Puerto Rico
AMDURER, STANLEY, New York University
AMICK, DONALD H., Rutgers—The State University
ANASTASI, ANNE, Fordham University
ANASTASIO, ERNEST J., Educational Testing Service
ANDERSON, G. ERNEST, JR., University of Delaware
ANDERSON, ROY N., North Carolina State University
ANDERSON, SCARVIA B., Educational Testing Service
ANDREWS, T. G., University of Maryland
ANGEL, JAMES L., Ohio State Department of Education
ANGOFF, WILLIAM H., Educational Testing Service
ANSARA, ALICE, Manter Hall School, Cambridge, Massachusetts
APOSTAL, ROBERT A., University of Maine
AQUINO, JOSE G., University of Puerto Rico
ARCHAMBAULT, FRANCIS, University of Connecticut
ARNOLD, M. L., Harper & Row, Publishers
ARONOW, MIRIAM S., Long Island University
ASHE, AMELIA H., New York University
ATKINS, WILLIAM H., Monmouth College
AUSTIN, GILBERT, University of New Hampshire
AWAD, ROBERT A., U. S. Army Southeastern Signal School
AXEL, GLENDA R., Brooklyn College of the City University of New York

1966 Invitational Conference on Testing Problems

AYERS, J. DOUGLAS, California Test Bureau, Monterey, California
BADAL, ALDEN W., Oakland (California) Public Schools
BAGGALEY, ANDREW R., University of Pennsylvania
BALL, J. T., American Institute of Certified Public Accountants
BALL, SAMUEL, Teachers College, Columbia University
BANDY, SUZANNE J., Montgomery County (Maryland) Public Schools
BANGS, MRS. GEORGE, Chambersburg (Pennsylvania) Public Schools
BARBUTO, PAUL F., JR., Ontario Institute for Studies in Education
BARDACK, HERBERT, New York State Department of Social Welfare
BARNES, EDWARD G., Georgia State Department of Education
BARNETTE, W. LESLIE, The State University of New York at Buffalo
BARRETT, RICHARD S., Science Research Associates, Inc.
BARRON, ARLEEN, Educational Testing Service
BARROWS, THOMAS S., Educational Testing Service
BARTNIK, ROBERT V., Marion, Indiana
BARTON, DEAN P., Science Research Associates, Inc.
BARTON, GEORGE M., Educational Testing Service
BATES, WARD P., Blake School, Hopkins, Minnesota
BATTAGLIO, JOSEPH J., Ipswich (Massachusetts) High School
BATTAGLIO, MRS. JOSEPH J., Ipswich, Massachusetts
BEARD, JACOB GORDON, Florida State University
BELCHER, LEON H., Educational Testing Service
BENNETT, GEORGE K., The Psychological Corporation
BENNETT, MRS. GEORGE K., Bronxville, New York
BENNETT, VIRGINIA D. C., Rutgers—The State University
BENSON, ARTHUR L., Educational Testing Service
BENSON, ERNEST B., Culver (Indiana) Military Academy
BENSON, MRS. ERNEST B., Culver, Indiana
BENT, PETER R., Law School, Columbia University
BERG, JOEL, Newton Public Schools, Newtonville, Massachusetts
BERGER, BERNARD, New York City Department of Personnel
BERGER, EMANUEL, Pennsylvania State Department of Public Instruction
BERGESEN, B. E., JR., Personnel Press, Inc.
BERNARD, CHARLES C., University of North Carolina at Chapel Hill
BERNE, ELLIS J., U. S. Department of Health, Education and Welfare
BERNSTEIN, SYLVIA, University of Massachusetts
BIANCHI, BENJAMIN A., U. S. Employment Service
BICKNELL, J. E., University of Alberta
BIRCH, DOROTHY L., Educational Testing Service
BLANCHARD, CARROLL M., U. S. Army Signal Center and School
BLIGH, HAROLD F., Harcourt, Brace and World, Inc.
BLIXT, SONYA LYNN, The State University of New York at Albany
BLOOM, BENJAMIN S., University of Chicago

Participants

BLOXOM, BRUCE, Educational Testing Service
BLUM, STUART H., Brooklyn College of The City University of New York
BOEHM, ROBERT F., New York State Department of Civil Service
BOLDT, ROBERT F., Educational Testing Service
BOLLENBACHER, JOAN, Cincinnati (Ohio) Public Schools
BONDARUK, JOHN, National Security Agency
BONHAM, GEORGE W., Science and University Affairs, Inc., New York City
BOSSARD, GRACE, New Castle (Delaware) Junior School
BOWER, DAVID R., New York State Department of Education
BOWER, JAMES R., University of Michigan
BOWERS. NORMAN D., Northwestern University
BOWES, EDWARD W., University of California at Berkeley
BOWMAN, HOWARD A., Los Angeles (California) Public Schools
BRACA, SUSAN, Oceanside (New York) Senior High School
BRADY, PATRICK B., Northbrook, Illinois
BRAGER, GARY L., University of Maryland
BRANDON. WILLIAM R., Scholastic Magazines, Inc.
BRANDT, HYMAN, Executives Limited International, New York City
BREEN, THOMAS, University of Connecticut
BRETNALL, DORIS, Educational Records Bureau
BRIDGES, CLAUDE F., U. S. Military Academy
BRIDGMAN, DONALD S., Commission on Human Resources, Rowayton. Connecticut
BRISTOW, WILLIAM H., New York City Board of Education
BROD, PEARL, New York City Board of Higher Education
BRODSKY, DAVID J., Educational Testing Service
BROWN, F. MARTIN, Fountain Valley School, Colorado Springs, Colorado
BROWN, FOSTER L., The State University of New York at Oneonta
BROWN, FRANK W., New Hampshire State Department of Education
BROWN, FRED S., Great Neck (New York) Public Schools
BROWN, JAMES A., New York State Department of Education
BROWN, ROBERT R., Pennsylvania State University
BROWNE, MICHAEL W., Educational Testing Service
BRYAN, JAMES, Educational Testing Service
BRYAN, MIRIAM M., Educational Testing Service
BRYAN, NED, U. S. Office of Education
BUCHANAN, R. G., The Psychological Corporation
BUERGER, T. A., Lakewood (Ohio) Public Schools
BURGER, NORMAN, Joseph L. Fisher Foundation, Inc., New York City
BURKE, JAMES M., Connecticut State Department of Education
BURKE, PAUL J., Bell Telephone Laboratories, Inc.
BURKS, MATTHEW, Hampton Institute
BURNETTE, RICHARD R., Florida Southern College

BURNHAM, PAUL S., Yale University
BURNS, RICHARD L., Educational Testing Service
BUROS, OSCAR K., Mental Measurements Yearbook
BURR, WILLIAM L., Harcourt, Brace and World, Inc.
BURRILL, LOIS, Harcourt, Brace and World, Inc.
BUXTON, ELEANOR, Darien (Connecticut) High School
BYRNE, RICHARD HILL, University of Maryland
CAHILL, ROBERT J., Boston College
CALDWELL, EDWARD, University of South Florida
CALDWELL, JOHN T., WGBH-WGBX Television, Boston, Massachusetts
CALLAHAN, MARGARET M., Boston (Massachusetts) Public Schools
CAMPBELL, JAMES P., University of Pennsylvania
CAMPBELL, JOEL T., Educational Testing Service
CANTWELL, ZITA M., Brooklyn College of The City University of New York
CAPPELLUZZO, EMMA, University of Massachusetts
CAPPS, MARIAN P., Virginia State College
CARLSON, J. SPENCER, University of Oregon
CARLSON, JULIA, Educational Testing Service
CARNEGIE, ELIZABETH, Nursing Outlook
CARNEY, MELVIN, Erie (Pennsylvania) Public Schools
CARROLL, JOHN B., Harvard University
CARSTATER, EUGENE D., Bureau of Naval Personnel
CASSERLY, PAT, Educational Testing Service
CEDAR, TOBY, Wayne State University, Ohio
CENTRA, JOHN A., Educational Testing Service
CHANDLER, MARJORIE O., Educational Testing Service
CHAPMAN, GLORIA, New York University
CHASNOFF, ROBERT E., Newark State College, New Jersey
CHAUNCEY, HENRY, Educational Testing Service
CHEVRIER, JEAN-MARC, Institute of Psychological Research, Montreal, Canada
CHOPPIN, BRUCE H., Cornell University
CHRIST, RICHARD, U. S. Army Signal Center and School
CHRISTOS, CONSTANTINE, Clarkstown Central Schools, New City, New York
CIERI, VINCENT P., U. S. Army Signal Center and School
CLAWAR, HARRY J., Educational Records Bureau
CLEARY, J. ROBERT, Educational Testing Service
CLEGG, AMBROSE A., JR., University of Massachusetts
CLEMANS, WILLIAM V., Science Research Associates, Inc.
CLENDENEN, DOROTHY M., The Psychological Corporation
COCHRAN, JOHN A., Maumee (Ohio) Valley Country Day School
COFFIN, GREGORY C., Evanston (Illinois) Community Consolidated Schools
COFFMAN, WILLIAM E., Educational Testing Service
COHEN, SANDRA R., Educational Testing Service

COLBURN, MARK, U. S. Department of Defense
COLE, FRANK, New Jersey State Department of Education
COLEMAN, ELIZABETH R., Clarkstown Junior High School, West Nyack, New York
COLEMAN, JEROME P., Lansingburgh (New York) Public Schools
COLEMAN, ROGER D., Independent School Testing Service, Brookline, Massachusetts
COLVER, ROBERT M., Educational Testing Service
CONNOLLY, JOHN A., Educational Testing Service
CONRAD, HERBERT S., U. S. Office of Education
CONRY, ROBERT, University of Wisconsin
CONSALUD, ROBERT W., Boston College
CONSALUS, CHARLES E., Educational Testing Service
CONWAY, CLIFFORD B., Department of Education, British Columbia, Canada
COOGAN, BARBARA, University of Connecticut
COOPER, DAVID, Hunter College of The City University of New York
COOPERMAN, IRENE G., Veterans Administration
COPE, WILLIAM, Phelps-Stokes Fund, New York City
COPELAND, HERMAN A., American Public Health Association, Philadelphia, Pennsylvania
CORBIN, RICHARD, Hunter College High School, New York City
CORY, BERTHA HARPER, U. S. Army Personnel Research Office
COSGROVE, DON J., The Procter and Gamble Company
COSGROVE, JOHN E., New York State Department of Civil Service
COULSON, ROGER W., Butler University
COWE, EILEEN G., Hunter College of The City University of New York
COWELL, WILLIAM R., Albion College
CRABTREE, MYRNA, New Jersey State Department of Education
CRAVEN, ETHEL C., Veterans Administration Hospital, Brooklyn, New York
CREW, JOHN L., Maryland State Department of Education
CROOKS, LOIS, Educational Testing Service
CUMMINGS, MARY B., Boston (Massachusetts) School Committee
CURK, RALPH D., Dayton (Ohio) Board of Education
CURRAN, FLORENCE V., Educational Testing Service
CURRAN, FRED A., Boston College
CURRAN, ROBERT, University of Massachusetts
CURRIVAN, GENE, The New York Times
CURRY, ROBERT P., Cincinnati (Ohio) Public Schools
CYNAMON, MANUEL, Brooklyn College of The City University of New York
DAVIDSON, HELEN H., City College of The City University of New York
DAVIS, FRED B., University of Pennsylvania
DAVIS, JOHN B., JR., Worcester (Massachusetts) Public Schools
DAVIS, JUNIUS A., Educational Testing Service

1966 Invitational Conference on Testing Problems

DAYTON, C. M., University of Maryland
DE MOOSE, NORMAN M., YMCA Vocational Service Center, New York City
DENENMARK, FRIEDA, U. S. Office of Education
DENHAM, CAROLYN H., Boston College
DETCHEN, LILY, Chatham College
DEUTSCH, MARTIN, New York University
DEXTER, DONALD, City of Philadelphia Department of Personnel
DIAMOND, EDWARD C., Watervliet (New York) Public Schools
DIAMOND, ESTHER E., Science Research Associates, Inc.
DIAMOND, LORRAINE KRUGLOV, City College of The City University of New York
DIAMOND, MILTON, New York City Board of Education
DICKSON, GWEN S., Kensington, Maryland
DIEDERICH, PAUL B., Educational Testing Service
DIEHL, MARY JANE, Educational Testing Service
DIEPPA, JORGE J., College Entrance Examination Board, Puerto Rico
DIMENGO, CARL, AKRON (Ohio) Public Schools
DI VESTA, FRANCIS J., Pennsylvania State University
DOERR, JOE, University of Missouri
DONLON, THOMAS F., Educational Testing Service
DOPPELT, JEROME E., The Psychological Corporation
DOWNES, MARGARET C., New York State Department of Education
DOYLE, DONNA G., Norfolk (Virginia) Public Schools
DOYLE, JAMES F., Atlanta (Georgia) Public Schools
DRAGOSITZ, ANNA, Educational Testing Service
DREW, MARY, Educational Testing Service
DRY, RAYMOND, Life Insurance Agency Management Association, Hartford, Connecticut
DUBNICK, LESTER, Huntington (New York) Public Schools
DU BOIS, PHILIP H., Washington University
DUKER, SAM, Brooklyn College of The City University of New York
DUNN, FRANCES E., Brown University
DUTTON, GENE, Rhode Island College
DYER, HENRY S., Educational Testing Service
EAGLE, NORMAN, Englewood, New Jersey
EASH, MAURICE J., Hunter College of The City University of New York
EBEL, ROBERT L., Michigan State University
EBERHARDT, DEL., Greenwich (Connecticut) Public Schools
EDGAR, ROBERT W., Queens College of The City University of New York
EHRMAN, EVELYN A., District of Columbia Public Schools
ELLIS, ALLAN B., Harvard University
ENGELHARDT, DAVID F., Harvard University
ENGELHART, MAX D., Duke University

ENGELHART, MRS. MAX D., Durham, North Carolina
ERICKSON, HARLEY E., State College of Iowa
ERLANDSON, F. L., Michigan State University
ESCOBAR, CARLOS, Dayton (Ohio) Board of Education
ESSER, BARBARA, Educational Testing Service
EVANS, RICHARD, Highland Park (New Jersey) Public Schools
EVEN, ALEXANDER, Ontario Institute for Studies in Education
FAIR, J. C. STEACY, Royal Canadian Air Force Central Examining Board, Ontario, Canada
FAN, CHUNG T., International Business Machines Corporation
FARLEY, FRANCIS W., City of Philadelphia Department of Personnel
FARLEY, GEORGE, University of Massachusetts
FARR, S. DAVID, The State University of New York at Buffalo
FEARING, JOSEPH L., University of Houston
FEHR, HOWARD F., Teachers College, Columbia University
FELDMANN, Shirley C., City College of The City University of New York
FELDT, LEONARD S., University of Iowa
FELTY, JOHN, The State University of New York at Plattsburgh
FERRIS, RICHARD J., Dean Junior College
FIELD, HENRY E., University of Canterbury, New Zealand
FIFER, GORDON, Hunter College of The City University of New York
FILICETTI, PETER J., LaSalle College
FINDIKYAN, N., College Entrance Examination Board
FINEGAN, OWEN T., Gannon College
FINELLI, JEANNE, Educational Testing Service
FINGER, JOHN A., JR., Rhode Island College
FINKELSTEIN, ABRAHAM M., New York City Board of Education
FINN, JEREMY, The State University of New York at Buffalo
FISCHELIS, R. L., Yale University
FITZGERALD, JOHN F. M., Hillcrest School, Brookline, Massachusetts
FITZGIBBON, T. J., Harcourt, Brace and World, Inc.
FLAMER, EVAN B., Boston College
FLANAGAN, JOHN C., American Institutes for Research
FLEISCH, SYLVIA, Boston University
FLINT, DAVID L., City of Philadelphia Department of Personnel
FLOYD, WILLIAM, West Lafayette (Indiana) Community Schools
FLOYD, WILLIAM A., Appalachian State College, North Carolina
FOLEY, JOSEPH J., Boston College
FORDYCE, JOSEPH W., Santa Fe Junior College
FORLANO, ANNA-MARIE, New York City Public Schools
FORLANO, GEORGE, New York City Board of Education
FORSTER, CECIL R., Neighborhood Youth Corps, New York City
FRANK, RUTH E., Educational Testing Service

FRANKFELDT, ELI, U. S. Army Personnel Research Office
FRASE, LAWRENCE, University of Massachusetts
FREDERIKSEN, NORMAN, Educational Testing Service
FREEBERG, NORMAN E., Educational Testing Service
FREEMAN, TED B., South Carolina State Department of Education
FREMER, JOHN J., JR., Educational Testing Service
FRENCH, BENJAMIN J., New York State Department of Civil Service
FRENCH, NATHANIEL S., North Shore Country Day School, Winnetka, Illinois
FRICKE, BENNO G., University of Michigan
FRIEDMAN, JERRY J., Kakiat Junior High School, Spring Valley, New York
FRIENDLY, MICHAEL, Educational Testing Service
FRIERY, CATHERINE M., National League for Nursing
FRITZ, KENTNER V., Ohio Wesleyan University
FROE, OTIS D., Morgan State College, Maryland
FULLERTON, BARRY A., Dean Junior College
FRY, EDWARD, Rutgers—The State University
GALLAGHER, HENRIETTA, Educational Testing Service
GAMES, PAUL A., Ohio University
GANNON, FRED, Educational Testing Service
GARVIN, A. D., University of Maryland
GENTILE, J. RONALD, Pennsylvania State University
GIBLETTE, JOHN F., University of Maryland
GILLIS, NANCY, Ontario Institute for Studies in Education
GIUNTI, FRANK E., U. S. Army Signal Center and School
GLASIER, CHARLES A., New York State Department of Education
GLOCKLER, ANTHONY S., Educational Testing Service
GOCHMAN, DAVID S., International Encyclopedia of Social Sciences
GODSHALK, FRED I., Educational Testing Service
GOERING, JACOB D., University of Maryland
GOLDBERG, MYLES, Educational Testing Service
GOLDMAN, BERT A., University of North Carolina at Greensboro
GOLDMAN, LEO, The City University of New York
GOLDMAN, NORMAN, New Jersey Education Association Review
GOLDSTEIN, BEVERLY, National Board of Medical Examiners
GOLDSTEIN, LEO S., New York University
GOLDSTEIN, WILLIAM, Howard B. Mattlin Junior High School, Plainview, New York
GORDON, LEONARD V., The State University of New York at Albany
GORDON, SHEILA, United Federation of Teachers
GORHAM, WILLIAM, U. S. Department of Health, Education and Welfare
GOTKIN, LASSAR G., New York University
GRANT, DONALD L., American Telephone and Telegraph Company
GRAVES, WALTER, National Education Association Journal

GRAY, FRANK M., Educational Records Bureau
GRAY, MRS. LYLE BLAINE, Diagnostic and Remedial Center, Baltimore, Maryland
GREDLER, GILBERT R., Temple University
GREEN, DONALD A., Ohio University
GREENBERG, BERNARD L., Gallaudet College
GREENBERG, SALLY H., U. S. Civil Service Commission
GREENE, JAY E., New York City Board of Education
GREENLAND, THOMAS, University of Kentucky
GREENWOOD, JOHN M., Teachers College, Columbia University
GREER, HARRY H., JR., International Business Machines Corporation
GRESS, HAROLD S., La Junta (Colorado) High School
GRESS, MRS. HAROLD S., La Junta, Colorado
GROBMAN, HULDA, New York University
GROMEN, LAUREN R., Kent State University, Ohio
GROSE, ROBERT F., Amherst College
GROSSWALD, JULES, Philadelphia (Pennsylvania) Public Schools
GRUMMON, DONALD L., Michigan State University
GUILFORD, J. P., University of Southern California
GULESIAN, MARK, University of Massachusetts
GULLIKSEN, HAROLD, Educational Testing Service
HAAG, CARL H., Educational Testing Service
HADLEY, EVERETT E., Parsons College
HAGEN, ELIZABETH, Teachers College, Columbia University
HAGMAN, ELMER R., Greenwich (Connecticut) Public Schools
HALL, ROBERT G., 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
HAMILTON, RHODA J. ROSEN, Groveland (Florida) High School
HAMPTON, RONALD WAYNE, Chester County (South Carolina) Department of Education
HANDRICK, FANNIE A., Harvard University
HANNA, GERALD S., Harcourt, Brace and World, Inc.
HARASYMIW, STEFAN J., Educational Testing Service
HARMAN, HARRY H., Educational Testing Service
HARMON, LINDSEY R., National Research Council
HAROOTUNIAN, BERJ, University of Delaware
HARRIS, BEATRICE, The City University of New York
HARRIS, DAVID P., American Language Institute
HARRISON, VELMA, Lampasas (Texas) High School
HART, JOSEPHINE, New York State Department of Civil Service
HARTSHORNE, NATHANIEL, Educational Testing Service

1966 Invitational Conference on Testing Problems

HARVEY, PHILIP R., Educational Testing Service
HASTINGS, J. THOMAS, University of Illinois
HAVEN, ELIZABETH W., Educational Testing Service
HAWES, GENE, Columbia University Press
HAYMAN, JOHN L., JR., Philadelphia (Pennsylvania) Public Schools
HAYWARD, PRISCILLA, New York State Department of Education
HEACOCK, LINDA, Educational Testing Service
HEFLIN, VIRGINIA B., Educational Developmental Laboratories, Huntington, New York
HELM, CARL E., Educational Testing Service
HELMICK, JOHN S., Educational Testing Service
HELMSTADTER, GERALD C., Arizona State University
HENNIS, WILLIAM M., North Carolina State Department of Public Instruction
HENRY, SALLYANN, The Psychological Corporation
HERBERG THEODORE, Pittsfield (Massachusetts) Public Schools
HERLIHY, JOHN G., Cooperative Project for Curriculum Development, Bennington, Vermont
HERMAN, DAVID O., The Psychological Corporation
HERMANN, MARGARET, Educational Testing Service
HESLIN, PHYLLIS, National League for Nursing
HIGGINS, MARTIN J., Exploratory Committee for Assessing the Progress of Education, Minneapolis, Minnesota
HILL, ARTHUR W., Friends School, Wilmington, Delaware
HILLER, JACK, University of Connecticut
HILTON, DAVID W., Boston College
HILTON, THOMAS L., Educational Testing Service
HITCH, COLONEL KENNETH S., Industrial College of the Armed Forces
HO, WAI-CHING, Educational Research Council of Greater Cleveland
HOFFMANN, EVA, Vocational Advisory Service, New York City
HOGAN, THOMAS P., Harcourt, Brace and World, Inc.
HOLLAND, JOHN L., American College Testing Program
HOLLISTER, ANNE, General Learning Corporation
HOLLISTER, JOHN S., Educational Testing Service
HOOD, DONALD, Educational Testing Service
HOROWITZ, MILTON W., Queens College of The City University of New York
HOUIS, JANE L., Educational Testing Service
HOUSTON, THOMAS, University of Wisconsin
HOWARD, ALLEN H., University of Illinois
HOWELL, JOHN J., New Jersey Educational Research
HUGHES, JOHN L., International Business Machines Corporation
HUGHES, NORMAN H., Harcourt, Brace and World, Inc.
HUMMEL, RAYMOND, University of Pittsburgh
HUMPHREY, BETTY, Educational Testing Service

HUYSER, ROBERT J., Educational Testing Service
IMPELLIZZERI, IRENE, Brooklyn College of The City University of New York
IRBY, ALICE J., Educational Testing Service
IRONSIDE, RODERICK, Educational Testing Service
IWAMOTO, DAVID, U. S. Office of Education
JACOBS, PAUL, Educational Testing Service
JACOBS, STANLEY S., University of Maryland
JACOBSEN, GWENDOLYN M., University of Rhode Island
JACOBSON, DANIEL, Michigan State University
JACOBSON, ROBERT L., Editorial Projects for Education, Baltimore, Maryland
JANEBA, HUGO B., Rutherford (New Jersey) Senior High School
JASPEN, NATHAN, New York University
JINKS, ELSIE H., Detroit (Michigan) Public Schools
JOHNSON, A. PEMBERTON, Newark College of Engineering
JOHNSON, CHARLES E., University of Maryland
JOHNSON, DONALD W., Pennsylvania State University
JOHNSON, LEWIS W., City of Philadelphia Department of Personnel
JONES, MARY LOU, Newark State College, Newark
JONES, ROBERT A., University of Southern California
JÖRESKOG, KARL G., Educational Testing Service
JOSELYN, EDWIN GARY, University of Minnesota
JOSEPH, BROTHER LAWRENCE, Archdiocesan Schools of New York
JUNGEBLUT, ANN, Educational Testing Service
JUOLA, ARVO, Michigan State University
JUSTMAN, JOSEPH, Center for Urban Education, New York City
KABACK, GOLDIE R., City College of The City University of New York
KAMIL, IRVING, Albert Einstein School, Bronx, New York
KAMMAN, JAMES F., University of Illinois
KARL, MADELINE, New York City Board of Education
KARMEL, LOUIS J., Queens College of The City University of New York
KARSTETTER, ALLAN B., Syracuse, New York
KASEL, FRANCIS, Educational Testing Service
KATZ, MARTIN, Educational Testing Service
KAVRUCK, SAMUEL, U. S. Office of Education
KAYE, MILDRED, Brooklyn College of The City University of New York
KAZANJIAN, DOLORES, Educational Testing Service
KEITER, ROBERTA, Montgomery County (Maryland) Public Schools
KELEHER, The Reverend Gregory, St. Anselm's College
KELLER, FRANKLIN J., Ohio State University
KELLEY, C. FREDERICK, Darien (Connecticut) Public Schools
KELLEY, H. PAUL, College Entrance Examination Board
KELLEY, PAUL R., National Board of Medical Examiners
KELLY, JAMES, WGBH—Television, Boston, Massachusetts

KELLY, JOHN W., Palm Beach County (Florida) Public Schools
KENNEDY, JOHN L., Institute for Educational Development
KENNEDY, NANCY, Educational Testing Service
KENNEY, HELEN J., Northeastern University
KEPPEL, FRANCIS, General Learning Corporation
KIGHT, HOWARD R., The State University of New York at Buffalo
KILBANE, MARIAN, Cleveland (Ohio) Public Schools
KIPP, JOAN, New York City Board of Education
KIRBY, WILLIAM H., University of Maryland
KIRSH, ELAINE, Educational Testing Service
KLEIN, ALICE, Rutgers—The State University
KLEIN, STEPHEN, Educational Testing Service
KLEINKE, DAVID J., New York State Department of Education
KLEMER, DONALD A., Hastings-on-Hudson (New York) Public Schools
KLEVER, GERALD L., Board of Christian Education, United Presbyterian Church in the U.S.A.
KLING, FREDERICK R., Educational Testing Service
KLING, MARTIN, Rutgers—The State University
KNAPP, THOMAS R., University of Rochester
KNOELL, DOROTHY M., The State University of New York at Albany
KOCK, REINO, University of Massachusetts
KOGAN, LEONARD S., Brooklyn College of The City University of New York
KRAMER, ABE S., West Chester State College, Pennsylvania
KRUMBOLTZ, JOHN D., Stanford University
KUJAWSKI, CARL J., Atlantic Richfield Company
KURLAND, NORMAN D., New York State Department of Education
KVARACEUS, WILLIAM C., Tufts University
LANDRY, RICHARD, Boston College
LANE, GORHAM, University of Delaware
LANE, WILLIAM S., Vashon Island High School, Burton, Washington
LANKTON, ROBERT S., Detroit (Michigan) Public Schools
LANTZ, DONALD L., Educational Testing Service
LAPOINTE, ARCHIE E., California Test Bureau, Monterey, California
LARSEN, EDWIN P., Oakland (California) Public Schools
LATHROP, ROBERT L., Pennsylvania State University
LA VINE, MARIAN S., Jericho (New York) High School
LEAR, MARYANN A., Educational Testing Service
LEARY, PAUL, University of Massachusetts
LEBOWITZ, GORDON, Correlated Curriculum Project, Brooklyn, New York
LEHMANN, IRVIN J., Michigan State University
LENNON, ROGER T., Harcourt, Brace and World, Inc.
LEVERETT, HOLLIS M., Leverett Associates, Boston, Massachusetts
LEWIS, ANNE, Education USA

LEWIS, CHARLES, Educational Testing Service
LINDEMAN, RICHARD H., Teachers College, Columbia University
LINDQUIST, E. F., University of Iowa
LINDSAY, CARL A., Pennsylvania State University
LINK, FRANCES R., CHELTENHAM (Pennsylvania) Public Schools
LINN, ROBERT L., Educational Testing Service
LIPSITZ, LAWRENCE, Educational Technology
LITTELL, JOSEPH F., Harper & Row, Publishers
LITTERICK, WILLIAM S., Educational Records Bureau
LLANO, MARGARET T., American Institute of Physics
LOCKETTE, LOUISA L., Trenton (New Jersey) Public Schools
LOHMAN, MAURICE A., The City University of New York
LOMBARD, JOHN W., Science Research Associates, Inc.
LONG, LOUIS, City College of The City University of New York
LORET, PETER G., Educational Testing Service
LOUI, BEATRICE, Hawaii State Department of Education
LOVELL, ALDEN L., University of New Hampshire
LOWRANCE, ROBERT B., Rhodes School, New York City
LOYE, DAVID E., Educational Testing Service
LUCAS, DIANA D., Educational Testing Service
LUISETTE, JOAN, New York City Board of Education
LUNNEY, GERALD H., University of Massachusetts
LUTZ, MARJORIE, Professional Examination Service, New York City
LUTZ, ORPHA M., Montclair State College, New Jersey
LYNCH, ELEANOR, National League for Nursing
MAC KENZIE, LOUISE, University of Rhode Island
MAC KINNEY, A. C., Iowa State University
MADAUS, GEORGE F., Boston College
MAJETIC, RICHARD M., Educational Testing Service
MALCOLM, DONALD J., Educational Testing Service
MANNING, WINTON H., College Entrance Examination Board
MARCANO, PAQUITA, College Entrance Examination Board, Puerto Rico
MARCO, GARY L., Educational Testing Service
MARCOTTE, DONALD, University of Connecticut
MARQUESS, VINCENT, Palm Beach County (Florida) Public Schools
MARSH, JAMES V., Educational Testing Service
MARTIN, PHILIP L., Blake School, Hopkins, Minnesota
MARTIN, RUSSELL W., Educational Testing Service
MARTUZA, VICTOR R., University of Maryland
MASLOW, ALBERT P., U. S. Civil Service Commission
MASIA, BERTRAM B., Western Reserve University
MASSENGILL, ANNA, The Shuford-Massengill Corporation, Lexington, Massachusetts

MASSENGILL, H. EDWARD, The Shuford-Massengill Corporation, Lexington, Massachusetts
MASSEY, WILL J., Homewood School, Baltimore, Maryland
MATHERS, ALBERT P., New Canaan (Connecticut) Public Schools
MATHEWSON, ROBERT H., The City Universit of New York
MATTHEWS, DOROTHY, Bennett College
MATULA, BEVERLY V., Educational Testing Service
MAYO, SAMUEL T., Loyola University
MC CALL, W. C., University of South Carolina
MC CANN, FORBES E., McCann Associates, Philadelphia, Pennsylvania
MC CARTHY, DOROTHEA, Fordham University
MC CARTHY, THOMAS N., La Salle College
MC CONNELL, JOHN C., Windward School, White Plains, New York
MC CORD, RICHARD B., City of Philadelphia Department of Personnel
MC GAUVRAN, MARY E., Massachusetts State College at Lowell
MC GEE, THE REVEREND JOSEPH M., Louisville (Kentucky) Catholic School Board
MC GILLICUDDY, MARJORIE, Albany, New York
MC INTIRE, PAUL H., Boston University
MC INTOSH, VERGIL M., Extension Course Institute, Gunter AFB, Alabama
MC LAUGHLIN, KENNETH F., U. S. Office of Education
MC LEAN, LESLIE, Ontario Institute for Studies in Education
MC MANUS, JOHN, University of Connecticut
MC MORRIS, ROBERT F., The State University of New York at Albany
MC NALLY, JAMES L., Educational Testing Service
MEDLEY, DONALD M., Educational Testing Service
MELBYE, BENNETT C., Frazee (Minnesota) High School
MELCHER, CHRISTINE, Rhodes School, New York City
MELTON, RICHARD S., Educational Testing Service
MELVILLE, S. DONALD, Educational Testing Service
MERENDA, PETER F., University of Rhode Island
MERRALL, DAVID F., Rhodes School, New York City
MERRY, MARGARET H., Wheelock College
MERRY, ROBERT W., Harvard University
MERWIN, JACK C., University of Minnesota
MESSICK, SAMUEL, Educational Testing Service
MEYERS, RUTH Z., Educational Testing Service
MICHAEL, JOAN J., University of Southern California
MICHAEL, WILLIAM B., University of California at Santa Barbara
MICHELI, GENE S., U. S. Naval Training Device Center
MIDDENDORF, LORNA, Detroit (Michigan) Public Schools
MILHOLLAND, JOHN E., University of Michigan
MILLER, CARROLL L., Howard University

MILLER, MARIAN B., Delaware State Department of Public Instruction
MILLMAN, JASON, Cornell University
MIRKIN, LOUISE, Science Research Associates, Inc.
MITCHELL, BLYTHE C., Harcourt, Brace and World, Inc.
MITZEL, M. ADELE, Baltimore (Maryland) Public Schools
MIZENKO, A., U. S. Army Signal Center and School
MOLLENKOPF, WILLIAM G., Procter and Gamble Company
MONTEMERLO, MEL, Pennsylvania State University
MOONEY, ROBERT F., Boston College
MOORMAN, JANE, University of Wisconsin
MOREAU, THE REVEREND GEORGE H., National Catholic Educational Association
MORGAN, DONNA, The City College of The City University of New York
MORRISON, FRANCES, The State University of New York at Albany
MOSELY, RUSSELL, Wisconsin State Department of Education
MOWERS, GLENN E., California Test Bureau, New Cumberland, Pennsylvania
MULLIGAN, ALEXANDER R., Farmingdale, New York
MUSTICO, THOMAS W., The State University of New York at Albany
MYERS, ALBERT E., University of Delaware
MYERS, CHARLES T., Educational Testing Service
NEARINE, ROBERT J., Hartford (Connecticut) Board of Education
NEILL, ROBERT D., University of Louisville
NELSON, A. K., University of Missouri
NEWBY, EDNA M., Douglass College of Rutgers—The State University
NEYMAN, CLINTON A., JR., The George Washington University
NICHOLS, RICHARD J., Newark State College, New Jersey
NICHOLS, ROBERT C., National Merit Scholarship Corporation
NIGRO, GEORGE A., Boston College
NIPPLE, CHARLES, Anderson (Indiana) Public Schools
NONKIN, ABRAHAM, New York City Department of Personnel
NORRELL, GWEN, Michigan State University
NORRIS, RAYMOND C., George Peabody College for Teachers
NORTH, ROBERT D., The Psychological Corporation
NOVICK, MELVIN R., Educational Testing Service
NULTY, FRANCIS X., Educational Testing Service
OAKEY, JOSEPH H., Niskayuna High School, Schenectady, New York
O'CONNOR, VIRGIL J., Educational Testing Service
OLANOFF, MARTIN, New York City Board of Education
OLITSKY, ROSLYN, Christopher Columbus High School, Bronx, New York
OLIVER, THOMAS C., Southern Illinois University
OLSON, LEROY A., Michigan State University
OPPENHEIM, DON B., Educational Testing Service
ORLANDI, LISANIO, Boston College

ORMISTON, KENNETH, New York State Department of Education
ORR, DAVID B., American Institutes for Research
OTIS, C. ROBERT, California Test Bureau, Fulton, New York
OTTE, RICHARD, U. S. Office of Education
OTTOBRE, FRANCES M., Educational Testing Service
OZUR, HERBERT, U. S. Civil Service Commission
PAELET, DAVID, Darien (Connecticut) Public Schools
PAGE, ELLIS B., University of Connecticut
PAGE, MRS. ELLIS B., Storrs, Connecticut
PALLRAND, GEORGE J., Princeton University
PALMER, LESLIE A., American Language Institute
PALMER, ORVILLE, Educational Testing Service
PALUBINSKAS, ALICE L., Tufts University
PANOS, ROBERT J., American Council on Education
PARK, ELDON E., Educational Testing Service
PARTINGTON, J. E., U. S. Army Evaluation Center
PATTON, MICHAEL J., Educational Testing Service
PAULUS, DIETER H., University of Connecticut
PAYNE, DAVID A., Syracuse University
PEMBERTON, W. A., University of Delaware
PERRY, JESSE P., JR., Rockefeller Foundation
PERRY, W. D., University of North Carolina at Chapel Hill
PETERSON, RICHARD E., Educational Testing Service
PHIPPS, LILA O., Westfield (New Jersey) Senior High School
PIKE, LEWIS, Educational Testing Service
PITCHER, BARBARA, Educational Testing Service
PLISKEWICH, SONIA, Queens College of The City University of New York
POLLACK, NORMAN C., New York State Department of Civil Service
POOLE, RICHARD L., The State University of New York at Buffalo
PRIOR, JOHN J., Pratt Institute
PROCTER, WINIFRED S., Educational Testing Service
PROCTOR, CHARLES M., Montgomery County (Maryland) Public Schools
PRUZEK, ROBERT M., The State University of New York at Albany
QUICK, ROBERT, American Council on Education
QUILLING, MARY, University of Wisconsin
QUINN, JOHN S., JR., Harcourt, Brace and World, Inc.
RABINOWITZ, WILLIAM, Pennsylvania State University
RANKIN, GARY, Educational Testing Service
RAPHAEL, BROTHER ALOYSIUS, Bishop Loughlin Memorial High School, Brooklyn, New York
RAPPARLIE, JOHN H., The Owens-Illinois Glass Company
RAVIS, HOWARD S., Scholastic Teacher
READING, ETHEL A., Educational Testing Service

REELING, GLENN E., Jersey City State College, New Jersey
REID, JOHN W., Indiana University of Pennsylvania
REINER, WILLIAM B., Hunter College of The City University of New York
REISS, JEAN F., Educational Testing Service
REMSTAD, ROBERT C., University of Wisconsin
RESTAINO, LILLIAN R., Lexington School for the Deaf, New York City
REYNOLDS, RICHARD, The State University of New York at Albany
RHUM, GORDON J., State College of Iowa
RICHARDS, KENNETH W., Akron (Ohio) Public Schools
RICHARDS, MRS. KENNETH, Akron, Ohio
RICKS, JAMES H., The Psychological Corporation
RITENOUR, LOUISE, Educational Testing Service
RIVLIN, HARRY N., Fordham University
ROBERTS, S. O., Fisk University
ROBERTSON, GARY J., Harcourt, Brace and World, Inc.
ROBINS, ALEXANDER R., U. S. Department of Health, Education and Welfare
ROCA, PABLO, Organization of American States, Pan American Union
ROCK, DONALD A., Educational Testing Service
ROFMAN, R. LEWIS, Philadelphia (Pennsylvania) Public Schools
ROHRBAUGH, FRANCES G., Educational Testing Service
ROSEN, JULIUS, The City College of The City University of New York
ROSENBACH, JOHN H., The State University of New York at Albany
ROSENBAUM, BERNARD, Teachers College, Columbia University
ROSENTHAL, ELSA, Educational Testing Service
ROSNER, BENJAMIN, Educational Testing Service
ROSS, DONALD, Educational Testing Service
ROSS, PAUL C., New York State Department of Education
ROSSMANN, JACK, Macalester College
ROTH, ROD, Oakland Schools, Pontiac, Michigan
ROTHENBERGER, RUSSELL R., Pennsylvania Department of Professional and Occupational Affairs.
ROTHMAN, GERALDINE H., Hunter College High School, New York City
ROWE, FREDERICK B., Randolph-Macon Woman's College
RUBENS, YVONNE A., National League for Nursing
SACHS, LORRAINE P., National League for Nursing
SAGE, ELLIS H., University of Rochester
SAINSBURY, ANTHONY B., The Committee of Vice-Chancellors and Principals of the Universities of the United Kingdom
SALANCIK, J. R., Yale University
SASAJIMA, MASU, Harvard University
SASLOW, MAX S., New York City Department of Personnel
SAVERY, EDWARD W., Friends School, Wilmington, Delaware
SCARBOROUGH, REGINA, Educational Testing Service

SCHERFF, RICHARD L., Columbia University
SCHLEKAT, GEORGE A., Educational Testing Service
SCHNEIDER, HARRIET L., National League for Nursing
SCHNITZEN, JOSEPH P., University of Houston
SCHOEPPS, MARJORIE E., Educational Testing Service
SCHRADER, WILLIAM B., Educational Testing Service
SCHULTZ, CHARLES B., Seattle University
SCHULTZ, KENNETH M., Dade County (Florida) Public Schools
SCHUMER, HARRY, University of Massachusetts
SCHWEIKER, R., University of Massachusetts
SCOFIELD, LEONARD G., Dean Junior College
SCOTT, C. WINFIELD, Rutgers—The State University
SEFEIN, NAIM A., The State University of New York at Fredonia
SEIBEL, DEAN W., Educational Testing Service
SERAFINO, ROBERT P., New Haven (Connecticut) Public Schools
SERAYDARIAN, THOMAS, Princeton (New Jersey) High School
SEYFERT, WARREN C., National Association of Secondary School Principals
SFORZA, RICHARD F., U. S. Military Academy
SHAHINIAN, SIROON P., New York State Department of Mental Hygiene
SHARP, CATHERINE G., Educational Testing Service
SHARP, RICHARD M., New York University
SHAYCOFT, MARION F., American Institutes for Research
SHEA, JOHN J., University of Massachusetts
SHEALY, JOYCE HAYES, Queens College, North Carolina
SHIELDS, MARY, National League for Nursing
SHIMBERG, BENJAMIN, Educational Testing Service
SHULTZ, JOHN L., Rutgers—The State University
SILVERBERG, EDWARD, New York City Department of Personnel
SINGER, ROSLYN F., Educational Testing Service
SIRCUS, GERALD, Basic Systems, Inc.,
SISENWEIN, ROBERT, Rutgers—The State University
SISTER BARBARA GEOGHEGAN, College of Mount St. Joseph
SISTER M. DE SALES, Rosary Hill College
SISTER M. EILEEN, Rosary Hill College
SISTER M. ELAINE, Our Lady Queen of Heaven School, Lake Charles, Louisiana
SISTER M. JOSEPH MOTTE, Boston College
SISTER MAGDALA, Mount St. Agnes College
SISTER MARIE BAPTISTA POLLARD, Boorady Reading Center, Dunkirk, New York
SISTER MARY COLLEEN, Archdiocesan Schools of New Orleans
SISTER MARY DE LOURDES, Detroit, Michigan
SISTER MARY HEFFERNAN, National Catholic Educational Association
SISTER MARY PAULA PECK, Fordham University

SKURNIK, L., National Foundation for Educational Research in England and Wales
SLAKTER, MALCOLM J., The State University of New York at Buffalo
SLOAN, JACK, The State University of New York at Albany
SMITH, ALLAN B., Rhode Island College
SMITH, ANN Z., Educational Testing Service
SMITH, HERMAN F., Educational Testing Service
SMITH, JOHN P., Educational Testing Service
SMITH, LEANDER W., Science Research Associates, Inc.
SMITH, LYMAN J., Science Research Associates, Inc.
SMITH, MARIE A., Montgomery County Community College, Pennsylvania
SMITH, MARK, American Association of Colleges for Teacher Education
SMITH, MARSHALL P., Trenton State College, New Jersey
SMITH, ROBERT E., Educational Testing Service
SMITH, TOMMIE, Science Research Associates, Inc.
SMITH, WENDELL I., Bucknell University
SMITH, WILLIAM LAMBERT, Boston College
SNYDER, HARRY R., West Virginia State Department of Education
SOLBERG, CARL, General Learning Corporation
SOLES, STAN, San Francisco State College
SOLOMON, ROBERT J., Educational Testing Service
SOMMERFELD, ROY E., University of North Carolina at Chapel Hill
SOTHERLAND, DONALD W., New York State Department of Civil Service
SOUTHWORTH, J. ALFRED, University of Massachusetts
SPANEY, EMMA, Queens College of The City University of New York
SPARKS, JACK N., Pennsylvania State University
SPAULDING, GERALDINE, Educational Records Bureau
SPEARRITT, DONALD, University of Sydney, Australia
SPEER, GEORGE S., Illinois Institute of Technology
SPENCE, JAMES R., The State University of New York at Albany
SPENCER, RICHARD E., University of Illinois
SPILIOS, PAUL G., University of New Hampshire
SPREMULLI, ESTELLE, Educational Testing Service
SPRING, BERNARD P., Princeton University
SPYKER, JANET C., Amherst—Pelham (Massachusetts) Public Schools
SQUIRES, JO ANNE, Hampton Roads Speech and Hearing Center, Virginia
STAJDUHAR, BARBARA, Chicago Civil Service Commission
STAKE, ROBERT E., University of Illinois
STANLEY, JULIAN C., Center for Advanced Study in the Behavioral Sciences
STAPLES, I. EZRA, Philadelphia (Pennsylvania) Public Schools
STATLER, CHARLES R., University of South Carolina
STEINMAN, ARTHUR M., Trenton State College, New Jersey
STELLWAGON, WALTER, Science Research Associates, Inc.

STEPHENS, ROBERT W., Elma (Washington) High School
STEWART, E. ELIZABETH, Educational Testing Service
STICE, GLEN, New York City
STICKELL, DAVID W., Pennsylvania State University
STINCHFIELD, PHILIP A., West Hartford (Connecticut) Public Schools
STODOLSKY, SUSAN, Harvard University
STOKER, HOWARD W., Florida State University
STONE, PHILIP J., Harvard University
STORM, ALBERT, U. S. Office of Education
STOUGHTON, ROBERT, Connecticut State Department of Education
STREICHER, SAMUEL, New York City Board of Education
STRICKER, LAWRENCE J., Educational Testing Service
STROHMEYER, JAMES P., North Valley Regional High School, Demarest, New Jersey
STRONG, KENNETH W., Bronson (Michigan) High School
STRONG, MRS. KENNETH W., Bronson, Michigan
STRUTH, JOHANN F., Harcourt, Brace and World, Inc.
STUDDIFORD, WALTER B., Princeton University
STUNKARD, CLAYTON L., University of Maryland
SULKIN, SIDNEY, Changing Times
SULLINS, WALTER L., University of Maryland
SUTTON, H. A., Department of Education, Ontario, Canada
SUTTON, JOSEPH T., University of Alabama
SWANSON, CARL W., New Jersey State Department of Education
SWANSON, ED, University of Minnesota
SWIFT, EVERETT L., The Peddie School, Hightstown, New Jersey
SWINEFORD, FRANCES, Educational Testing Service
TATSUOKA, MAURICE, University of Illinois
TAYLOR, BRUCE L., Educational Testing Service
TAYLOR, JUSTINE N., Educational Testing Service
TAYLOR, MARGARET C., Baltimore (Maryland) Public Schools
TAYLOR, PETER A., Rutgers—The State University
TEMP, GEORGE, Educational Testing Service
THEINERT, HELEN N., Springfield (Massachusetts) Public Schools
THEVAOS, DENO G. Pennsylvania State University
THOMAS, ETHEL N., Princeton (New Jersey) High School
THOMAS, WILLIAM F., University of Wisconsin
THOMPSON, ALBERT S., Teachers College, Columbia University
THORNDIKE, ROBERT L., Teachers College, Columbia University
THORNE, MARGARET A., Educational Testing Service
TILLERY, DALE, University of California at Berkeley
TIPTON, DOROTHY, Highland Park (New Jersey) High School
TOMASZEWSKI, F. RICHARD, New York State Department of Education

TOMS, FREDERICK H., Educational Testing Service
TRAILOR, ALICE, University of Connecticut
TRAUB, ROSS, E., Ontario Institute for Studies in Education
TRAXLER, ARTHUR E., University of Miami
TRAXLER, MRS. ARTHUR E., Miami, Florida
TRIESTMAN, MORRIS, Veterans Administration
TRIGGS, FRANCES, Committee on Diagnostic Reading Tests, Inc.
TRIPPEL, ALBERT W., Cincinnati, Ohio
TRIPPEL, MARY R., Mt. Healthy (Ohio) Public Schools
TRISMEN, DONALD, Educational Testing Service
TROW, MARTIN A., University of California at Berkeley
TUCKER, LEDYARD R, University of Illinois
TULLY, EMERSON, Florida Board of Regents
TURNBULL, WILLIAM W., Educational Testing Service
TYLER, MATILDA, Simpson School, Wallingford, Connecticut
UHL, NORMAN P., Emory University
VALLEY, JOHN R., Educational Testing Service
VANCE, FORREST L., University of Rochester
VAN HORN, CHARLES, Purdue University
VERY, PHILIP S., University of Rhode Island
VICKERS, BRUCE E., The Baylor School, Chattanooga, Tennessee
VITELES, MORRIS S., University of Pennsylvania
VOIGT, ELIZABETH, New York University
VON FANGE, ERICH A., Concordia Lutheran Junior College
VROOMAN, ALAN H., The Phillips Exeter Academy
VROOMAN, MRS. ALAN, Exeter, New Hampshire
WADSWORTH, BARRY, The State University of New York at Albany
WAGNER, E. PAUL, Bloomsburg State College, Pennsylvania
WAHLGREN, HARDY, The State University of New York at Geneseo
WAINER, HOWARD, Educational Testing Service
WALENTA, ESTHER T., Educational Testing Service
WALKER, ROBERT N., Personnel Press, Inc.
WALLACE, GAYLEN R., Oklahoma State University
WALLMARK, MADELINE, Educational Testing Service
WALTHEW JOHN K., Educational Testing Service
WALTON, WESLEY W., Educational Testing Service
WALTZMAN, HAL, New York State Department of Civil Service
WANTMAN, MOREY J., Educational Testing Service
WARE, DOROTHY A., Detroit Psychological Clinic
WARGA, RICHARD G., Educational Testing Service
WARRINGTON, WILLARD G., Michigan State University
WASIK, JOHN, Montgomery County (Maryland) Public Schools
WATKINS, J. FOSTER, Regional Curriculum Project, Atlanta, Georgia

WATSON, LEONOR, Operation Upgrade, Bronx, New York
WATSON, WALTER S., Cooper Union
WEAVER, A. S., Norwich (Connecticut) Free Academy
WEBB, EUGENE J., Northwestern University
WEBER, BURTON, Kilmer Job Corps Center, Edison, New Jersey
WEBER, EWALD, New York City Department of Personnel
WEINER, MAX, Brooklyn College of The City University of New York
WEISBRODT, JEROME, Educational Testing Service
WEISS, JOEL, University of Chicago
WEITZ, HENRY, Duke University
WESMAN, ALEXANDER G., The Psychological Corporation
WESSELL, NILS Y., Institute for Educational Development
WEST, SARA F., Summit (New Jersey) Public Schools
WESTBROOK, BERT W., North Carolina State University
WESTERLUND, S. R., U. S. Office of Education
WHARRY, MRS. RUSSELL, National Cathedral School, Washington, D.C.
WHEELER, PATRICIA, Educational Testing Service
WHITE, ELLIS F., New York University
WHITLA, DEAN K., Harvard University
WICHERT, VICTOR E., Educational Testing Service
WICOFF, EVELYN, Educational Testing Service
WIERSMA, WILLIAM, University of Toledo
WILCOX, LEE, University of Minnesota
WILKE, MARGUERITE M., Hunter College of The City University of New York
WILKE, WALTER H., New York University
WILLEMIN, LOUIS P., U. S. Army Personnel Research Office
WILLEY, CLARENCE F., Norwich University
WILLIAMS, CHARLES, Oakland Schools, Pontiac, Michigan
WILLIAMS, JANE, Westminster Schools, Atlanta, Georgia
WILLIAMS, MALCOLM J., U. S. Coast Guard Academy
WILLIAMS, ROGER K., Morgan State College, Maryland
WILLIS, WARREN K., University of Illinois
WILSON, KENNETH M., Vassar College
WILSON, MICHAEL J., New England Education Data Systems
WILTSEY, ROBERT, Educational Testing Service
WIMMER, DAVID, Anderson (Indiana) Public Schools
WINANS, S. DAVID, New Jersey State Department of Education
WINGO, ALFRED L., Virginia State Department of Education
WINIEWICZ, Casimer S., U. S. Naval Training Center
WINTERBOTTOM, JOHN A., Educational Testing Serivce
WITHERS, ANITA W., Educational Testing Service
WITTMAYER, MAGDALENE D., Lyons Township (Illinois) High School and Junior College

WODTKE, KENNETH H., Pennsylvania State University
WOHLHEUTER, JAMES, Educational Testing Service
WOLF, RICHARD, The State University of New York at Albany
WOLF, WILLIAM C., University of Massachusetts
WOLLOWICK, HERBERT B., Teachers College, Columbia University
WOMER, FRANK B., University of Michigan
WOOD, BEN D., Columbia University
WORDEN, VINCENT J., State College at Bridgewater, Massachusetts
WORTHEN, JOHN E., University of Delaware
WRIGHT, MARIAN, Hamilton County (Ohio) Office of Education
WRIGHT, ROBERT, The State University of New York at Albany
WRIGHT, WILBUR H., The State University of New York at Geneseo
WRIGHTSTONE, J. WAYNE, New York City Board of Education
WYATT, T. S., University of Cambridge Local Examinations Syndicate, England
WYSONG, H. EUGENE, Ohio State Department of Education
YAFFE, PAUL, Baltimore (Maryland) Public Schools
YUNDT, JOHN H., Moorestown (New Jersey) Public Schools
ZARRO, PASQUALE J., City of Philadelphia Department of Personnel
ZINN, KARL L., University of Michigan
ZOLA, EUGENE J., Niskayuna High School, Schenectady, New York
ZUCKERMAN, HAROLD, New York City Board of Education