HERSCHEL T. MANUEL, Ph. D.

Professor of Educational Psychology, University of Texas. Visiting Professor of Education, University of Puerto Rico, Summer, 1954.

Address to the Puerto Rico Psychological Asociation, July 12, 1954.

PROBLEMS OF INTER-LANGUAGE AND INTER-CULTURE MEASUREMENT IN PSYCHOLOGY AND EDUCATION

A Personal Word

I esteem it a high honor to have been invited to address the Puerto Rico Psychological Association on this occasion. The event means much more to me than it can possibly mean to you. I am deeply grateful.

In the organization of the Puerto Rico Psychological Association you have given the world additional evidence of the professional interests and resources of your people. I congratulate you most heartily and extend to you my very best wishes.

As some of you know, I have been coming to this beautiful island at intervals for a number of years. I think I can justly claim a genuine interest in the welfare of its people. I have

many friends here. An increasing number have been students in my classes, and I take great pride in their accomplishments, their leadership, and their productivity. I cannot claim that my work with them has been a substantial factor in their success, for they were already persons of great potentiality before they came to me. Nevertheless, they were with me for a time, and I have experienced a certain identification of my interests with theirs.

The spread of my visits over a period of years has given me an unusual opportunity to witness a socio-economic change of magnitude and rapidity which, I think, would be difficult to match anywhere in the world. Puerto Rico is taking its place as a center of culture and as a community where human welfare is its major aim. It is a lighthouse in an important area of the world, guiding perplexed and troubled peoples toward peace, security, democratic living, and the enjoyment of the fruits of science, art, and religion. Students and professional people from other countries are coming in ever greater numbers to see at first hand what you are doing and to be taught by you.

There are tangible evidences of material progress in extensive new housing, in modern highways, in rural electrification, and in new factories. One needs no statistics to see that the standard of living is much higher than it was only a short time ago. In general the people I see at work and on the street, show less evidence of insecurity and of the strain which comes from the struggle for a precarious livelihood.

Education also is going forward. I note with pride the new buildings which dot the landscape, viewing your new library especially with unrestrained admiration. I find the insular Department of Education, and the staff of the University alert and eager to go forward. I rejoice with you in the graduation of your first class from the Medical School. Being a teacher myself, I look forward as many of you do to the further development of graduate work at the University. I cannot pretend to speak from the standpoint of resources, but I am clear as to the need — particularly in the field of my own specialization, education and psychology. Education is the bulwark of a demo-

cratic society, and psychology is at the very heart of the educative process. In the long run it is education, guided by high ideals, that makes men free. The most precious resources in all the world are human beings, and these resources are developed through education — education which prepares the lowliest person for vocation and leadership, education that develops leaders, inventors, and artists. And it is psychology that tells how human behavior may be controlled. At this crossroads of the Americas it would be tragic to overlook the needs of advanced professional work in education and psychology. I am sure, of course, that these needs are not being overlooked.

Measurement in Psychology and Education

In discussing problems of inter-language and inter-culture measurement in psychology and education, we shall narrow the field by starting with the assumption that the use of tests, examinations, rating scales, inventories, and projective devices is properly called measurement. In accordance with this assumption, we may remind ourselves that psychologists are seeking through measurement just what other scientists are seeking — (1) to render their observations more precise and objetive, (2) to find descriptive units which can be subjected to mathematical treatment, and (3) to find a language through which their observations can be precisely stated, recorded, and communicated to others.

Students of the human organism have access, of course, to all of the units of measurement developed in the other sciences and in the practical affairs of life. Units of time, length, weight, electrical current, temperature, and the like, are just as available in psychology as elsewhere. The difficulties in our field are first those that characterize the delicate structure and the life processes of all the higher biological organisms, and second those that stem from the fact that our primary interest is in behavior, a field that is indescribably complex.

Difficulties of psychological analysis

If all the machines in the world could be built into one giant integrated mechanism doing everything that each of the machines now does, the problem of describing its behavior would still be less complicated and less difficult than the problem of describing human behavior. Such a machine would perform all the kinds of things done by steam-shovels, adding machines, typewriters, taxicabs, printing presses, lawn mowers, meat choppers, grain reapers, cane grinders, conveyor belts, and all the rest. As a matter of fact, these machines are but tools for man himself. What they do are but an extension of his own activities. Now imagine starting out to describe and measure the performance of all the machines in the world, and you have a task less difficult than describing and measuring man's behavior, for he does all the kinds of things which they do, and more. How can you get a meaningful description of a steam shovel, a typewriter, a calculator, and a printing press without describing each in turn? And if you try to describe the activities of all machines in turn, you will find that the task is almost interminable.

General and specific factors

There are certain aspects of the structure and behavior of machines which can be described in common language — axles, levers, wheels, belts, speed, horsepower, and the like. There are other aspects which must be described in terms of the specific thing which the machine does. The steam-shovel may have the same horsepower as a truck, but the steam-shovel digs into the earth and the truck carries things here and there.

In similar fashion there are ways of describing certain aspects of very different human behavior in common terms. Spearman, for example, likened his *g* to the energy which operates machines and found it to run through all intellectual processes. But you will recall the disturbing fact that a complete de-

scription of intellectual processes using Spearman's *g* would require also factors specific to the different activities, to say nothing of emotional processes which would require a still different description. The factor analysts hold out the hope that we may yet find a limited number of terms sufficient for a description of different segments of human behavior.

There is no reason to hope, however, that all of the needs of science and of daily life can be served by a few common descriptive terms. We shall still have to be interested in specific behavior as such — digging dirt, hauling goods, preparing food, and the like when we think of machines; and adding figures, computing interest, reading poetry, choosing a vocation, giving opposites, and the like when we think of human behavior. We need to know not only the horse-power of a steam-shovel but also how effective it is in digging into the earth, not only in general but in specific situations created by changes in the type of earth and the terrain on which it must work. Similarly, we need to know not only the general mental ability of a person and not only the strength of his "primary mental abilities," but also how well he adds a column of figures, how well he spells, how well he reads, how well he drives a car, and how effectively he performs a thousand other activities. In addition we need to know what he *wants* to do and what satisfactions he receives in what he does.

From specific performances to generalization

The problem of measurement in dealing with human behavior is just as complicated as is the description of behavior. Since its purpose is to add precision to description, there is theoretically a problem of measurement wherever there is description. Thus one needs to measure general mental ability and the elemental factors used to describe behavior, and also a multitude of specific activities, which for some purposes cannot be adequately described in terms of their component factors.

It is relatively easy to measure certain aspects of perform-

ance in a specific activity if the activity can be isolated and observed. Thus a reading paragraph can be presented and the time required for reading it recorded. It is easy also to find whether the reader can answer certain questions on the paragraph. The difficulty comes when we attempt to generalize from these specific observations. For example, does the speed with which he reads *this* paragraph represent the speed with which he reads *other* paragraphs? Does his *answer* to the *questions* asked reveal his *comprehension* of this paragraph? Does his comprehension of *this* paragraph represent his comprehension of *other* paragraphs? Does his *performance* reflect his *ability*, or does it reflect other factors?

In psychological and educational measurement we are sometimes interested in the specific performance on a test, as when a teacher asks the pupils to learn to spell a list of ten words and then gives a test to measure whether the pupils have learned these same ten words. Ofter, however, we are not primarily interested in the specific performance itself. We are using it as a sample of some more extensive behavior, as when a teacher gives a test of twenty questions to test the mastery of work covering a whole semester. Or a psychologist obtains reactions to a series of inkblots as a key to personality organization and emotional characteristics. Or he assigns a task —perhaps only the drawing of a man— to obtain evidence of the mental maturity of a child.

Measurement in Different Languages and Cultures

To this point in the paper we have dealt with problems of measurement growing out of the nature of the subject-matter of psychology. These are problems that face all psychologists, and, I think you will agree, they are extremely difficult. With this background we turn to the problems of measurement when those who apply the measures are of different language and culture or when those whose behavior is measured are of different language and culture, or when both the measurer and

the measured are of different language and culture. We shall be concerned primarily with pencil-and-paper tests.

Principle of multiple causation

You will recall that the psychologist is interested in generalizations from the specific test performance. It is at this point that the most perplexing problems of inter-language and inter-culture measurement appear. Thus, a paragraph written in English and dealing with objects and events in a highly industrialized economy could be presented alike to an English-speaking child in Chicago and an Indian child who is learning English in a mission school in a remote village of an undeveloped country. The speed with which each reads the material and his answers to the questions could be recorded. The scores would faithfully show the performance of each on this behavior sample, and the results of both tests would be equally valid as a test of response to a specific situation. Yet the results might have and probably would have meanings which are very different in important respects. The psychologist would guess that even different attitudes toward test-taking might influence the results, to say nothing of more obvious cultural differences. In interpreting the outcome of any test, the psychologist must always be keenly aware of the principle of multiple causation and especially cautious when there are known differences which are likely to influence test results. The most difficult problem in testing is to generalize from specific test performance — i.e., to interpret the scores.

Let us take a further step. Suppose we are interested in comparing the reading abilities of two children who use different languages. Here are two segments of behavior. Each reads. But their reading cannot be tested by the same test. It is like two tabulating-machines, one built to operate on IBM cards and another built to operate on cards with different spacing of punched holes. Both may operate effectively, but there is no way to test them with the same set of test cards. Some way must be found to punch cards with equivalent material, but each

adapted to the machine with which it is to be used. With reference to the children who speak different languages, obviously the material with which they are to be tested must be written in the two different languages; and there rises immediately the problem of equivalence. Nor is that all. If the results are to have the same meaning, the content must represent equivalent selections from the two universes of reading material. To use our machine analogy, one child may be more able to deal with one type of content than with another.

Problem of sampling

The problem of the relation of a sample of behavior to the universe which it represents is a crucial one. In taking intelligence, for example, the universe of test situations which might be presented is indefinitely large and varied. At first thought it might seem that a form-board which could be presented without language would be equally representative of the intelligence of groups representing different cultures. This, however, is not necessarily true. Success with that type of problem might represent a relatively higher accomplishment in one culture than in another.

In a group which is homogeneous with respect to language and culture the equivalence of two tests is commonly studied through administration of the two tests to the same group. If the tests are in different languages, obviously this cannot be done, for it cannot be assumed that members of any group are equally capable in two languages. If the equivalence of tests in different languages is to be established, it must be done through other methods.

Development of Tests

General principles

Now let us suppose that we are members of an international congress which has been called to consider world policies with

respect to the development of tests. What are some of the principles which should guide us? I think that some one would make the point at once that, for the sake of ease in communication, we should have common units of measurement. Assuming that we might be able to create equivalent tests in different languages and cultures, we should express the results in similar units. We should profit from the experience of the physical scientists who find communication much easier if all use units of the metric system rather than the various units found in different regions of the world. It is a challenging problem to find similar common units in the behavioral sciences.

I think that someone might raise the point also that workers in one country should be able to profit from the experiences and inventions of another. That is precisely what has been happening from the time of Binet. Tests developed in one country are translated and adapted for use in another.

Then another member of the group would likely raise the question of inter-language and inter-culture cooperation in the building of new tests. He would probably make the point that a test built with sole thought of its use in one language and culture would likely offer considerable difficulty to the translator and adapter. He might then propose that an international group be brought together to see whether such a group could find suitable content and equivalent language for building at least a few tests which could be used with somewhat the same meaning in different languages and cultures. In this case there would be no translation and adaptation of ready-made tests. New tests would be constructed of items chosen initially to be parallel and equivalent in the different languages and cultures.

The Inter-American Tests

There has been at least one pilot experiment of this kind. The tests now known as the Cooperative Inter-American Tests were developed in parallel English and Spanish editions by native English-speaking and native Spanish-speaking personnel.

Puerto Rico through the Department of Education and the University of Puerto Rico contributed liberally to that experiment. Persons sharing directly in the construction or revision of the tests included a number of your educators and psychologists. Without consulting my notes, I recall Dr. Teobaldo Casanova, Dr. Pablo Roca, Dr. Ismael Rodríguez Bou, Mr. Clyde Fisher, Professor Abigaíl Díaz, Dr. Pauline Rojas and Professor Ramón Ramírez López. Your own Chancellor, Assistant Chancellor, Secretary of Education, Assistant Secretary of Education, former Commissioner Gallardo, and the late and beloved Dean Osuna, and a host of others throughout the Island gave valuable assistance.

Digressing for a moment I want to tell you of an incident which I like to think of as symtomatic of the interest of the Puerto Rican people in education. It illustrates the truth that man cannot live abundantly and well "by bread alone." The Inter-American test materials were constructed initially for a study of the teaching of English in Puerto Rico, but they were printed on the Continent, and it was necessary to ship them to Puerto Rico during the war. Because of enemy submarine activity it was difficult enough to supply the people of the Island with food. Priorities had to be secured for shipments of all kinds and here were literally tons of test material to be shipped. When Dr. Mellado made a personal appeal for a priority to ship the tests, the answer was something like this, "Don't you think the Puerto Rican people need bread more than paper?" But the persistent educator got the priority, and the activities of both educating and feeding the people went forward.

Since it was impossible to evaluate the equivalence of the English and Spanish editions of the Inter-American Tests by administering both editions to the same group, we had to rely on putting into the tests materials which we thought would be equivalent. Thus each item was as nearly the same in both the English and Spanish editions as we could make it. We used our best judgment as to cultural bias and we attempted to use language of the same meaning and difficulty. In vocabulary we had the assistance of word lists in English and Spanish. Then

we sought to choose words which on a difficulty scale would be about as difficult relative to other words in one language as it was in the other. For example, if a Spanish word was tenth in difficulty in a list of sixty words, we thought that the word expressing the same concept in English should be near the tenth position in the English list.

In evaluating this method of test building we may frankly question our success as measured by the criterion of producing scores which have approximately the same meaning. There can be a serious question whether the same items are equally representative of the two universes from which they were selected. The most that we *can* claim is that approximately the same materials are included in the two editions. Thus the Puerto Rican child may take approximately the same test in Spanish or in English, and he may take approximately the same test in Spanish as a Continental child takes in English.

As experience with the Inter-American Tests grows, we are gradually assembling evaluative material and getting suggestions for new research. One thing that puzzles us is the differential reaction of Puerto Ricans and Texans to reading on the one hand and non-verbal material on the other. It might be thought that tests based upon pictures and drawings would give a truer comparison than tests of reading. There is some evidence, not yet conclusive, that language is a better basis for comparison than is non-language pencil-and-paper material. This at least may be said: The reading test showed University of Texas freshmen and University of Puerto Rican freshmen reading at about the same level, the first in English, of course, and the second in Spanish. A wider difference, however, in non-verbal scores suggests the possibility of a cultural difference.

Different programs of test development

Looking toward the future, I see no reason why workers in each language and culture should not develop their own tests for their own purposes. Certainly, in the foreseeable future we

cannot expect inter-language and inter-culture tests alone to meet all the needs of psychology and education. There is a place also for the translated and adapted tests with norms constructed for the new population on which it is to be used. There is every reason to believe that carefully constructed tests can be useful elsewhere in translated and adapted form with proper safeguards as to interpretation — and that needs to be emphasized. The results of the original and the adapted may not be equivalent.

I happen to have a special interest, however, in the test that is especially constructed for inter-language and inter-culture use, and I should like to see international cooperation toward that end. As an immediate objective, I suggest that we try to find in each of a few different fields a series of useful test items which have approximately the same order of difficulty in any group, regardless of language or culture. For the moment we need not be concerned whether they would be more difficult for persons in one social group or in another. Our immediate objective is to find a series of items in which the order of difficulty is independent of language or culture. For example, repeating six digits backwards is presumably more difficult than repeating five digits backwards, no matter what group tries it. I suppose also that a two-step problem is more difficult than a one-step problem in Texas, Puerto Rico, and wherever else it may be tried. If we could find a few scales in which the order of item difficulty would remain constant, we should have a substantial basis for intercommunication, even if one cultural group should find the entire series somewhat easier than another group found it to be.

Helen Eaton in her *Semantic Frequency List* has given us a revealing illustration of the possibility of arranging concepts in order of frequency of use in different languages. Perhaps we might be clever enough to find a hundred concepts and words to express them which could be arranged in an order of difficulty which would be approximately the same in a number of different languages. If that could be done, the list would serve as a common denominator to which different tests could be

related, and we should have a most useful instrument for communicating across linguistic and cultural lines. A similar thing might be done in different areas — perhaps number, space, and others.

A Possible Project for Puerto Rico

In concluding this paper I wish to present a plea for more cooperation among educators and psychologists of different languages and cultures and to point out the peculiar strategic position of Puerto Rico for leadership in this type of activity. I do not have in mind simply the exchange of information and attendance at the same conventions. These are important, to be sure, but we need more. We need to work together on common problems. Again, I do not mean that you should come to help me, or I to help you. That also is important. But the thing I have in mind is that we assemble workers as equals to carry forward research on projects which are of importance to all of them. Building a measuring scale which would be independent of population might be one such project.

With its increasing leadership in the Western Hemisphere, Puerto Rico may well wish to take the initiative in assembling inter-language and inter-culture groups for research on common problems. If so, this would be merely an extension of the important service which it now renders.