DOCUMENT RESUME

ED 045 725 TM 000 305

AUTHOR Findley, Warren G.
TITLE Measurement and Research in the Service of Education.
INSTITUTION Georgia Univ., Athens. Research and Development Center in Educational Stimulation.
SPONS AGENCY Office of Education (DHEW), Washington, D.C. Cooperative Research Program.
REPORT NO Ccc-Pap-10
PUB DATE Jun 70
CONTRACT CEC-6-10-061
NOTE 18p.; Invited address to Division D, Measurement and Research Methodology, American Educational Research Association, Minneapolis, Minnesota, March 1970

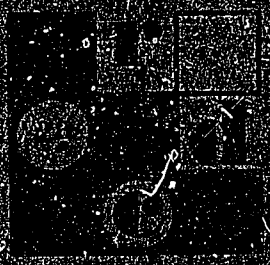
EDRS PRICE EDRS Price MF-$0.25 HC-$1.00
DESCRIPTORS Educational Innovation, Educational Policy, Educational Problems, *Educational Research, Educational Testing, *Evaluation, Instrumentation, International Education, Longitudinal Studies, *Measurement Techniques, National Surveys, *Research and Development Centers, Research Methodology, Research Utilization, *Statistical Analysis

ABSTRACT

This running commentary cites examples to show that measurement and research in education, since Binet, have been used to solve school problems. Current problems cited stem from post-World War II acceptance of the goal of educating "all the children of all the people." Compensatory education, peer-tutoring vs. competitive scrambling, mastery learning, retention vs. dropout, and socioeconomic segregation by tests, require study. Background surveys are cited that define the scope of the problems. Longitudinal studies are recommended that feed back evidence of need and of success and failure en route to terminal evaluation. Multivariate analysis is advanced as essential methodology; person-environment interactions should be studied, not just controlled, if individualization is a goal of the teaching-learning process. Major research and development organizations need to be based on an interactive model, rather than a linear model, to speed action and guide research. Innovations and their evaluation should be demanded and supported locally, not just by outside funds. ERIC is hailed for dissemination. New instrumentation is demanded for assessing the young, the disadvantaged, and foreign students; examples of adaptations are given. Forceful graphic presentation is described. Emphasis on the magnitude vs. the inferential certainty of findings is recommended. (Author)

MEASUREMENT AND RESEARCH IN THE PROVINCE OF EDUCATION

Wilfred C. Hadley



Research and Development Center in Educational Stimulation
University of Georgia Athens, Georgia

Occasional Paper No. 10

MEASUREMENT AND RESEARCH IN THE SERVICE OF EDUCATION

Warren G. Findley
Research and Development Center

RESEARCH AND DEVELOPMENT CENTER IN EDUCATIONAL STIMULATION
University of Georgia
Athens, Georgia
June, 1970

This paper was prepared as part of the activities of the Research and Development Center in Educational Stimulation, University of Georgia, pursuant to a contract with the United States Department of Health, Education and Welfare, Office of Education, under the Provisions of the Cooperative Research Program.

Center No. 5-0250 Contract No. OE 6-10-061

Measurement and Research in the Service of Education*

Warren G. Findley

Research and Development Center in Educational Stimulation

University of Georgia

The history of measurement and research in education, and the development of the methodology of these fields, is the history of efforts to solve real problems faced by educators. The rate at which the solutions offered by research workers were adopted and the extent to which they were applied have been questioned. But the history of application has not always been as so often pictured, a matter of reluctant adoption long after results were found.

The classic example is the work of Binet (Peterson, 1925). Binet was not commissioned to build an intelligence test or to advance the concept of an IQ with its imputed innateness and stability. Rather, he was set the task of helping schoolmen in Paris pick out children unable to keep up with the majority of their peers in mastering school learning as then presented in organized classes in school. He solved the problem by conceiving and developing a mental age scale useful in evaluating children's skill in essential mental processes. The scale was applied immediately and directly, and with refinements by others was later made the basis for assigning slow-learning children to special classes or institutions in the United States as well.

In the process, someone else (Stern, 1914) conceived the IQ as a ratic of mental age to chronological age that had considerable stability under prevailing conditions in school and society, a phenomenon that gave credibility to the notion that an innate measureable quality had been found.

---

*Invited address to Division D, Measurement and Research Methodology, American Educational Research Association, Minneapolis, Minnesota, March 4, 1970.

Army Alpha, developed to meet the very real problem of rapidly classifying recruits in World War I, the first time this country had faced the implications of universal conscription, was seized upon as the model not only for group measurement of "intelligence" among school children but for a scheme of achievement testing that hopefully was to undergird a new "science of education." Grade scores joined age scores as indices of mental measurement and were applied to successively larger segments of the public school population. Measurement instruments were applied with gusto in school surveys and elaborate concepts of educational quotients (educational age over chronological age) and achievement quotients (educational age over mental age) were advanced. Underachievement and overachievement were born (Thorndike, 1963). Underachievement consisted of doing less than one was capable of doing, while overachievement meant, of course, doing more than one was capable of! (Findley, 1959). How widely these concepts were adopted may be debated. Perhaps it was a mark of good sense of schoolmen not to be completely won over to these new concepts of the educational research community.

We have ever since been trying to qualify and clarify these concepts in the light of subsequent research. Factor analysis chiefly destroyed belief in the essentially unitary character of the IQ, and the Iowa and Chicago studies of the 1930's challenged its stability (Stoddard, 1943). The individual mental ability test continues to provide a basic index of mental retardation although it does not unequivocally point to special class placement for all children who score low.

Terman (1925-47), of course, had explored the implications of the upper end of the ability scales, culminating in his longitudinal studies. But neither he nor others generated an instructional model to follow, unless it was "ability grouping," a practice that has yet to live up to its promise of simplifying the instructional task by putting together those most alike in learning ability. Applied subject by subject and with due regard to more specific aspects of individual differences, it shows some promise. In the instructional field, research on the teaching of reading up to now has chiefly shown some clues to success in teaching middle class children, but no confident unanimity regarding method (Chall, 1967).

Meanwhile, in a market of scarcity for higher education, group tests like the College Entrance Examination Board's Scholastic Aptitude Test have demonstrated their usefulness in the selection of those most likely to master the academic curriculum of colleges, especially when used jointly with previous school records. Incidentally, now that "middle class" has become a term for special environmental advantages over the "disadvantaged," we generally forget that it was such tests that helped the "middle class" child of a generation or two ago to compete successfully with the "upper class" child of that day for selection to succeed in the academic world that had until then been largely reserved for a socioeconomic elite.

Progress was made during the 1930's and 1940's, considerably through the leadership of Ralph Tyler (1934; Smith and Tyler, 1942) in improving measures of school learning. The objectives of instruction became the objectives of test construction. Tests were designed to measure ability to apply knowledge and the highly adaptable multiple-choice format was proven capable of measuring applications as well as simple factual knowledge. As this trend developed, the distinctions between "intelligence," "scholastic aptitude," and "achievement" began to blur. All merged as "developed abilities" and it became clear that in so-called aptitude and intelligence tests we were measuring various kinds of achievement with a view to prediction (Wesman, 1948, 1956).

So, I submit, ours has been an eagerness for findings applicable to the problems of education. It now becomes appropriate to delineate the educational problems we see surfacing in our new period of uneven affluence.

After World War II, it became popular to talk about educating "all the children of all the people" (Educational Policies Commission, 1944). Compulsory attendance age had become 16 or higher everywhere. More recently state after state has moved toward universal accessibility of at least junior college education. Yet these advances bring into clearer focus new goals of even-handed availability of educational proficiency to all. What of those with backgrounds that do not permit them to take full advantage of new "equal" opportunities? What of the cumulating trend that has made the more affluent and more educated parents translate their favorable socioeconomic situation into educational "advantage" for their children? What agreement are we reaching on what constitutes equal educational opportunity for " all the children of all the people"?

Several approaches are being tried that show promise. We have already mentioned wide accessibility to higher education. This is largely a matter of money, and our states are appropriating it. A second trend is toward earlier education of children. Funds have not flowed so directly here, but neither have we reached agreement on what would help most.

Compensatory education for disadvantaged is another approach, but this is not clearly defined as to what it consists of. Paraprofessionals offer much promise, both for help to the children and for adult status. Male models are needed. Desegregation is approved, but presents problems of implementation even over and above the issue of residential segregation. Actually, socioeconomic segregation generates quite as strong a feeling as ethnic separation, although it does not stand out so sharply and visibly.

One very promising trend is tutoring (Cloward, 1967), by peers or by older children. When the older children are chosen from those who stand "below grade" among their peers, an especially fortunate situation prevails. Studies already show the tutors gain even more than the tutored. And why not? For once, these tutors were elevated from being ones everyone deplored or pitied to ones who could help someone less able. What's more they could see their success. Their pupils obviously had learned something. And they had brought this about. We who are teachers should be the first to understand.

It works with peers, too. Please pardon this personal anecdote. It goes back to my elementary school days and age 11. In the small town in which I grew up, spring meant sandlot baseball. It also meant a problem of finding teams to play. One team consisted of boys we would today characterize as living "on the other side of the tracks". We entered such games knowing that, if we won, they would rough us up to get even and that, if we lost, they would beat us up to show who was boss.

I remember particularly one boy who usually worked me over. Let us call him Frankie because that was his name. One day in seventh grade, we were having "Music" in the junior-senior high school auditorium where the rows of seats were close together and made passing difficult. Frankie came in last and tromped hard and incentionally on my left foot. Without pausing to reflect, I stuck out my right foot, tripping Frankie and sending him sprawling down the row. The teacher easily diagnosed the situation and sent us both to the principal's office.

The principal stayed in his office for a half hour while we waited under the watchful eye of his secretary. I suspect the principal could have come out sooner, but he was a wise old bird even then. (Later he became President of the American Educational Research Association as Dr. J. Cayce Morrison). I do not remember which one of us had brought his arithmetic book with him, but before the half hour was up, I had somehow gotten around to teaching Frankie some arithmetic he did not understand. Thereafter he never beat me up. And I was launched on my career as an educator, if not an educational psychologist!

I tell this story in a light vein, but it has a significant point. At least one boy "from the other side of the tracks" had found that the boy who always succeeded in arithmetic without seeming effort and made him feel he was just out to shine for himself, not only could but would share his skill because, indeed, he did make "hard" arithmetic make sense! How often do we give opportunity for this kind of mutually beneficial experience to take place in our competitive school atmosphere? ("Cheating" is the term applied to sub rosa efforts to help in what is often the only situation offered for such collaboration, but which the teacher has meant for a test!)

A corollary to this, in a way, is the new emphasis on mastery learning and its measurement. (Bloom, 1968; Carroll, 1963). I am teaching a class this quarter in statistics. The students are being taught the same material and taking the same tests as ever. But they are being given extra time to study the content under direction and are taking parallel tests - midterm and final. Already, the C's are practically wiped out, the 2 D's are, and several B's have become solid A's. Teachers with criterion-referenced tests can do this at almost any school level. It is perhaps easiest at the higher levels--college as contrasted with elementary school. But it can be done in elementary school too. I remember an eighth-grade arithmetic teacher who found ways to give us the extra time we needed to learn in class so that we never had homework. Programs of computer-assisted instruction and other programed approaches must be incorporated, wherever efficient, into our armamentarium of teaching-learning procedures to individualize instruction in some working compromise with the merits of humane personal attention to children's maturing.

If we are to address ourselves to the issues of what I have called the "even-handed availability of educational proficiency to all," what are the necessary conditions of methodology of research and measurement?

First, we need major surveys to define the scope of our problems. We have been enjoying the fruits of John Flanagan's conception of Project TALENT since 1960 (Flanagan, et al, 1962a, 1962b, 1964, 1966). Wave after wave of new and extended data have shown us what each of four successive classes of national samples of measured high school students have become and what factors in their backgrounds and experience have been most determinative in producing these developments. It may be argued that successive new waves are required, that the data of the 1960's are unequal to the facts of the 1970's. More about that shortly.

We have been treated to other studies, cross-sectional in their origins and current status, but capable of extension and replication at suitable intervals. The study of "Equal Educational Opportunity" directed by James Coleman (1966) has sought in the learning experiences of younger children an insight into the determinative factors in later productivity and happiness. Fortunately, for those dissatisfied with the imperfections of this "crash" study conducted in response to an act of Congress the data are there to be reworked by newer and more refined procedures. McPartland (1969a, 1969b) for example, has reworked the data on 5,075 Negro ninth-graders from New England and the Middle Atlantic states to show the predominant effect of presence or absence of white classmates on the achievement of these blacks. Using more elaborate methods of regression analysis developed by Mood too late for the original study, Mayeske and associates (1969), have explored the cumulative interaction effect between student background variables and the schooling prevalent today that leads initial advantage or disadvantage to be enhanced with time. Walker and associates (1967) have used

Baltimore City and County data to assess the cumulative impact of residence in a resegregating environment on ethnic-socioeconomic disadvantage.

The two-volume International Study of Achievement in Mathematics (Husén, 1967) used a cross-cultural sample from twelve cooperating nations to show the effects of differing national patterns of education upon mastery and interest. Lessons learned in presenting the findings of this major international comparative study will permit high-lighting even more effectively the effects of policies in extending universal education and of teaching students in inclusive or separate institutions in ways a study within a single national system would be limited by the pervasiveness of national patterns. Already data are gathered on science, reading comprehension, English and French as second languages, and civic education.

What more is in prospect for us as we face projection of our further research needs? The National Assessment of Progress in Education (Womer, 1970) is busily refining criterion-referenced measures of achievement in ten areas of school learning, to be administered in three-year cycles to national samples of schoolchildren ages 9, 13, and 17, and to adults 25-35. Data by region, community size, and ethnic group are being obtained. Successive triennial determinations will serve to chart progress in the gross national product of education as no past survey has done. Incidentally, our worthy purposes of national and international assessment are having to give attention to possible conflict of school populations for testing. Liaison has been established to prevent testing the same children to produce uncertain effects, while at the same time taking advantage of a common storehouse of sampling procedures. In both studies the possibility remains of follow-up studies of adult samples drawn from the same universes.

If these national and international surveys are to furnish our quality control assessments, what other requirements are to be set? Certainly, longitudinal studies of more limited problems are a must. We may well note the special value of Project TALENT in providing the directional value of following individuals long enough to do more than presume lasting effects. It is significant to note that the six-year Denver study (McKee and Brzeinski, 1966; Brzeinski et al. 1967) of the effects of beginning the teaching of reading in kindergarten at age five was supported on condition that no report of claims for its effectiveness would be made until the children had completed the six years of systematic effort. It is even more important to note that the notable gains over contemporaries who began such study at age six were no greater than the advantage over those who started at age five and were admitted to regular first grades after a single year of special initial instruction. This needs to be borne in mind in the present attraction to performance contracts. Let the contracts cover a span of years sufficient to demonstrate lasting effects. The first year's gain is not from a push at the top of a toboggan, but a first boost up a mighty tall tree of knowledge to climb.

It should be a first order of business to mount substantial learning projects beginning at ages 5, 4, 3 or even earlier, using different instructional plans and assumptions, and to follow the effects to age 11 at least, as the Denver study did. We do not know just what to do at every point, but we know enough to start and can revise and revamp plans for successive groups. We can use sufficiently game-like procedures to stimulate as well as appraise progress from early stages. And the youngsters will be happy withal. Let no one who subjects his children to TV deny this early education. Let us put to the empirical test of lasting effects the kind of early childhood education that has been practiced effectively but without systematic evaluation, in stimulating children since the Iowa studies of the 1930's and even earlier.

Another byword of the day is "multivariate". Multiple criteria need to be used to evaluate the outcomes of the teaching-learning process by our computer-enhanced multivariate methods of analysis because the teaching-learning process is multivariate in its inputs and treatments as well as its outputs. Those with misgivings about the side effects of early educational stimulation or any other innovation need to be invited to apply their measures of such effects in a multivariate design in which they are enabled to interact with the main effects of learning. Simple discriminant analysis between two groups has been powerful for sometime. Multivariate discriminant analysis promises to do for multiple groups what factor analysis has done for individual differences. Alternative differentiation in completion of courses and course sequences in college, for example, can be explored for most significant relations to student background data.

Factor analysis, or component analysis if you will, merits application sequentially to find best solutions in terms of the meaningful parameters of a substantive field (Findley, 1969). The variety of programs now available for extraction of factors (principal axes, R^2, minres), for rotations (varimax, quartimax, equamax), as well as choice of numbers of factors, need to be exploited to the full to make greatest sense out of known intercorrelations.

A prevailing theme in the current discussion of significance of differences is the relative importance of the magnitude and the certainty of a finding. Long before the day of our refined small sample statistics Percival Symonds (1930) proposed bi-serial r as a measure of magnitude of differences between means. Today we need to pay attention to the magnitude of differences between our adjusted means as well as to their inferential certainty. We need, in other words, to be as much concerned with Type B error as with Type A. In interpreting data to the less sophisticated, but even to ourselves, we would do well to consider a substitute for the word "significance". What about the "reliability" of a difference, or even better the "dependability" of a difference as a statement of the certainty of a finding while discussing the magnitude for its size, importance, or whatever.

This paper will not have failed of its purpose if it merely brings to serious attention the excellent discussion of "Education's Challenge to Psychology: The Prediction of Behavior from Person-Environment Interactions" by James V. Mitchell, Jr. (1969) in the latest issue of the Review of Educational Research. In fact, all responsible for the refreshing new viewpoint in this latest in a long series of triennial issues devoted to the Methodology of Educational Research are to be congratulated. In his chapter Mitchell challenges educational psychology, as Cronbach has before, to consider basic interactions between individual differences and environmental variables rather than to partial out or control individual differences to get a better grasp on other variables and their interrelations. Mitchell discusses three different interaction systems. First, he challenges us to expand and extend our multivariate procedures to cope with the aptitude-method interactions involved in teaching. Washburne and Heil's (1960) three categories of teachers and four categories of learners immediately spring to mind. Pupils were strivers, conformers, waverers, or rebellers, while teachers were either controlling, turbulent, or fearful. As I recall it, strivers and conformers performed well for any kind of teacher, while waverers and rebellers did better for controlling teachers than for turbulent ones. Fearful teachers were too unsure for anyone, so the controlling teachers showed to general advantage, turbulent teachers having a slight advantage only in teaching science. Mitchell goes on to cite cost benefit issues to be met in deciding at what point of specificity to attempt to individualize instruction to accommodate these differences. One is here reminded of Gage's (1967) concept of the good teacher as one with a "spray" effect, being not quite all things to all pupils, but at least alternately sensitive and responsive to one, then another to achieve continuity of attention and learning for all. The ultimate challenge appears to be to find multivariate analysis procedures equal to the complexity of the task.

Mitchell turns second to the verbal interaction analysis of Flanders which he finds too simplistic for the complexities of pupil-teacher interactions although more than a good start in analyzing them.

Mitchell finally turns to the various measures of campus climate of Pace, Stern, Astin, Holland and Pervin. These depend variously on student perceptions of environmental press. Here the challenge would appear to be a multivariate concept of subcultures to correspond to the diversity characteristic of large college populations.

An issue engaging much attention at this convention, as it has nationally for sometime, is the effective relating of research and development in the educational enterprise. Certainly we need much of both if curriculum development and longitudinal studies in depth are to be projected. Both are extensive and costly, hence require the long-time commitment of teams of research workers in established centers. Dr. Gallagher in his keynote remarks cited the ambiguity that bothered the OECD critics in the U.S.O.E.'s planning outline. They did not understand the difference, he felt, between the linear sequential character of the outline and the fundamentally interactive pattern of operations. It has always seemed to me,

since I heard him state it this way two years ago, that Frank Chase (1968) made the essential relationship abundantly clear when he said:

> "I incline to the view that research and development may be considered as an entity and not simply as a combination of some research with some development. It may be thought of primarily as a set of interrelated processes for dealing with problems in the context of the systems or situations in which they arise. It leads to the modification of existing systems for more effective performance and/or construction of new subsystems for performance of specified functions. The search is not so much for a perfect solution or product, but for the best that can be devised through the use of existing knowledge and technologies. Simultaneous processes of research and invention are employed, however, to increase the working capital of applicable knowledge and technology. The research, therefore, is development-relevant or motivated whether it is used to improve understanding of phenomena, to contribute to the solution of identified problems, or to test the effects of operations. The development in turn is research-informed, or guided, though not research-limited. Research is essential to systematic and continuing extension of the knowledge base on which development rests; and development constantly poses new problems which require research. At its best development often out-marches research by imaginative theoretical constructions and inventions; but as it does so, it gives new impetus to research and counts on the latter to regulate the pace for the health of the systems, societies, and individuals concerned."

The linear model of research, planning. development, evaluation, diffusion and utilization, or the shorter catechism of research, development, evaluation, and dissemination are good checklists of a fundamental sequence to bear in mind, but if the interactive relation is not explicitly spelled out, the linear sequence takes hold of too many minds. The model of industry with its large separate departments is even more distracting. No research and development center can long endure thus divided. In fact, no research and development center can flourish on campus without an additional function of training recognized as an integral part of its operations. Universities are training institutions equally with their research emphases and need to find a way to relate research and training. Lacking this, mission-oriented departments of government, with their built-in needs to change direction and emphasis, place disproportionate strains on sponsored programs in institutions of higher learning.

Education surely needs research and development. Public education especially needs guided innovation. But there is a need for continuity and for service to the public that argues for special continuing agencies of government with programs that transcend administrations. It is interesting that the new Review of Educational Research has a chapter on the Politics of Education (Kirst and Mosher, 1969) as an area of research concentration. The authors and some others seem almost to glory in this discovery. Previous desire for and attainment of insulation from political buffeting are viewed as outdated, naive. Yet children are at stake and require a continuity for growth in strength from their schools that in some measure is akin to the stability enjoyed in an intact family structure in the home. The need for continuity and stability must be balanced against a need for stimulation and change. Just as we are striving for new and improved ways to help children learn through experimentation, we are hearing of their rights and our invasion of their privacy. Schools seeking new and better ways must be the ones to reach out in behalf of their pupils for a continuously improving scheme of teaching, guided and supported by research. Is it too much to hope that initiative will come from local parents and schools on behalf of their children, seeking aid from state and federal leadership through university consultants and full-time staff supported by the local tax dollar as well as state and federal equalization formulas?

In the current debate over research vs. evaluation, I see much merit on both sides. Like Julian Stanley (1969), I see every reason for careful design and terminal judgment based thereon in every study. Such judgment needs also to be as independent as possible of the hopes and fears of the proponents or operators of an innovation under scrutiny. However, there is no question about the value, nay the essential importance, of formative evaluation during the process of the study. Formative evaluation not only provides immediate corrective feedback, but foretells the terminal or summative judgment. I was horrified a year ago to be informed by a usually reliable person that a group had planned to visit a project about which they had read an exciting report only to find that the project had been discontinued. It appears that the report they had read was the final report. Decision to continue the project had had to be made before the results were in and no one was willing to commit himself to continuing the project at that point.

Another anecdote about our earlier preference for terminal evaluation--even follow-up evaluation--is told of one of my colleagues. In the days of project research only, he had the practice of conducting two parallel streams of research in order to maintain a continuing staff rather than face the prospect of assembling, let alone training, a new staff when suitable time had elapsed after completing a project, certifying to its success, and obtaining a new grant to extend his explorations. By having parallel, alternating cycles in the two streams of research, workers could be transfered to Project B_1 while Project A_1 was being written up and Project A_2 was being proposed for funding. I don't believe, as rumor had it, that A-projects began in odd-numbered years and B-projects in even-numbered years, but you get the picture. Formative evaluation, let it be said, places demands on all participants in the project. Substantive specialists

have to embrace evaluation and research specialists have to become suddenly knowledgeable substantively.

To aid this, we require a new emphasis on clear, correct and forceful presentation of the facts and forces in education. An emphasis on graphics, found in too few elementary textbooks in measurement or statistics, is a must. The use of unit frequency polygons, parallel or superimposed, (Educational Testing Service, 1957) tell a story of comparative central tendencies, extent of variability and proportion of overlap that cannot be comprehended at all as clearly in tabular or textual form. When I first met John Ivey, now dean at Michigan State University, I was struck by his title at the University of North Carolina. He was the Director of a Division of Research Interpretation. To this day I remember his effective multi-colored charts; I often borrowed his lightweight easel for carrying charts for presentations.

In another vein, we are being aided in the communication of our results by the relatively new machinery of the ERIC Clearinghouses (U.S. Office of Education, 1970) now numbering 19, for the several substantive areas of education. The U.S. Office of Education, long the burial ground for research it had paid for, had become the patron of the researchers on a large scale, rescuing from obscurity not only government research, but papers and studies done in-house everywhere, but not otherwise available for wide distribution. Especially significant are the procedures developed by the several clearinghouses for making "information analyses", state of the art papers, for distribution to their special clientele.

A final word about instrumentation. A whole field crying for attention is the adaptation of appraisal devices to new populations. In early childhood education we have learned to capture children's interest and turn it toward earlier mastery of basic skills or processes. In doing this, much is made of games or game-like activities. In these activities children strive as we always did, and they learn. It remains for us to systematize observations and records of this learning. I (Findley, 1968) have watched preprimary children play a dart-gun game, shooting paper leaves off a tree drawn on the wall. Each leaf at first contains a single letter--later they carry whole words. In the game as commonly played, turns are taken with the gun and each child scores whenever he knocks off a leaf and can name the letter or word. If he cannot name the letter or word, it does not count and it becomes someone else's turn. Children play such games eagerly with minimum supervision for thirty or forty minutes at a stretch. At regular intervals of a week or two the teacher can put up a "standardized" set of letters and/or words and record the success of the children in successive groups. A more advanced group can be given a more advanced set of letters and/or words at the "test" sessions. Observations to date do not confirm the usual finding of better mastery of these reading fundamentals by girls.

At the present time, I am reviewing a first attempt to develop a reading test especially for ghetto children--in this case, Negro boys. Based on the assumption that the standard reading fare so much enjoyed by

middle class girls and many middle class boys will not command the interest and effort of ghetto children, the stories are about Negro athletes, getting evicted from poor housing, street incidents with police, and the like. I can only hope that this may open the way to measuring early progress en route to the later skill of reading textbook material so essential in school progress. There is good reason to hope. There is evidence that tests involve a matter of acceptance as well as capability. Research has shown, (Machover, 1943) for example, that the same children who are low on reading and arithmetic tests are equally low on abstract reasoning tests that appear meaningless exercises to them. At least they know it would be good to read well and do arithmetic problems, but abstract tests designed to free them from schooling handicaps fail to capture their imagination and effort. It is hopeful that we are attempting tests as well as instructional materials adapted to interests.

Another adaptation (Findley, 1962) practiced in many school systems and in experimental projects is to test children on standardized tests at their previously established reading levels. This insures that children can show what they know in different areas rather than proving on several different tests that they cannot read at the level required by the tests.

Presentation of tests orally has long been a feature of readiness tests, or diagnostic measures. Much can be done to test poor readers at higher language levels they understand when spoken, by reading multiple-choice test questions aloud to them while they follow on a copy before them. If each question is read aloud twice, with the teacher pausing at the end of each question or stem and at the end of successive alternatives printed on separate lines, third-graders and lower can follow the reading and mark their choice of options quite dependably.

Special adaptation is essential for appraising readiness or mastery of foreign students to whom English is a second language (Findley, 1966). Mention has already been made of the prospective value of the International Study of Educational Achievement in this area. But I was reminded once again only two weeks ago when a Korean student in my statistics course asked if he might bring a pocket Korean-English dictionary to the midterm test. I remembered an earlier foreign student, also a Korean, who had the same difficulty with my multiple-choice tests and is now holder of the doctorate and a staff member at AIR. I always allow extra time on my statistics examinations to permit American students to feel unhurried as they carry through the mysterious computations involved in finding a standard deviation. My former Korean student assured me that the extra time permitted him, when necessary, to translate whole multiple-choice exercises into Korean before answering. The fact is that English is so fine a first language because its derivation from the diverse Latin and German roots allows finer nuances of expression than languages that have only a single word for a quality and use a negative prefix to indicate its opposite. With this base we find it helps American students to present alternative choices and require selection of an answer rather than to

offer an open-ended question and put the burden of formulating an answer upon them. For the foreign student it is the other way around. He does better to formulate his answer to an essay question as best he can and put the burden of interpreting his intent on his reader or listener rather than accepting the burden on himself of interpreting a stem or question and four or five alternatives. I remember vividly a Pakistani student a few years back who always entered the final examination in one of our courses one grade higher than he finally received. I had students as well as faculty members come to me spontaneously with accounts of how well he had presented a statistical topic in class only to learn later that he had fallen down on the final examination. Laborious use of context to clarify uncertain terms in American discourse were the price of a second language.

So measurement and research continue to have a rich variety of contributions to offer to American education. The challenge is there for all of us to seize.

Bibliography

Bloom, Benjamin S. Learning for Mastery, Evaluation Comment, Vol. 1, No. 2, May 1968 (Center for the Study of Evaluation of Instructional Programs, 105 Moore Hall, 405 Hilgard, Los Angeles, Calif. 90024). pp. 1-7.

Brzeinski, Joseph E., et al. Should Johnny Read in Kindergarten? NEA Journal 56, 23-25, 1967.

Carroll, John B. A Model of School Learning. Teachers College Record Vol. 64, No. 8, May 1963, pp. 723-733.

Chall, Jeanne S. Learning to Read: the Great Debate. New York: McGraw-Hill, 1967. 372 pp.

Chase, Francis S. Educational Research and Development: Promise or Mirage? Journal of Research and Development in Education, Vol. 1, No. 4, Summer 1968, pp.3-14.

Cloward, Robert D. Studies in Tutoring. Journal of Experimental Education, Vol. 36, No. 1, Fall 1967, pp. 14-25.

Coleman, James S., et al. Equality of Educational Opportunity. Washington, D.C.: U.S. Government Printing Office, 1966. 737pp.

Educational Policies Commission. Education for All American Youth. Washington, D. C.: National Education Association, 1944. 421 pp.

Educational Testing Service. Learning and Teaching in Atlanta's Public Schools, 1955-56. Princeton, N.J.: Educational Testing Service, 1957, 132 pp.

Findley, Warren G. Falacies in the Concept of Overachievement. Vice-Presidential Address, Section Q-Education, AAAS - Chicago, Illinois, December 1959, 7 pp.

Findley, Warren G. The Effect of Adapting Standardized Tests to Reading Level in Grades 4-7 on Interpretation of Achievement. Paper read at AERA, Atlantic City, N.J., February 1962, 8pp.

Findley, Warren G. Selective Admissions of Teachers to Graduate Study. Regional Conference on Special Programs for the Preparation of In-Service Teachers for Admission to Graduate Study. Atlanta, Ga.: Atlanta University, 1966, pp.21-27.

Findley, Warren G. Early Education as Continuous Stimulation. Journal of Research and Development in Education, Vol.1, No. 3, Spring 1968, pp. 46-50.

Findley, Warren G. Use of Sequential Factor Analyses to Clarify Interpretation of Underlying Relations. Paper read at AERA, Los Angeles, Calif., February 1969, 6pp.

Flanagan, John C. et al. Design for a Study of American Youth. Boston, Mass.: Houghton Mifflin, 1962(a) 240 pp.

Flanagan, John C. et al. Studies of the American High School. Project Talent, Monograph No. 2. Pittsburgh, Penna.: University of Pittsburgh, 1962(b) 368 pp.

Flanagan, John C. et al. The American High-School Student. Pittsburgh, Pa.: Project Talent Office, University of Pittsburgh, 1964. 700pp.

Flanagan, John C. and Cooley, William W. One-Year Follow-Up Studies. Pittsburgh, Pa.: University of Pittsburgh, 1966. 347pp.

Gage, N. L. and Unruh, W.R. Theoretical Formulations for Research on Teaching. Review of Educational Research,Vol. 37, No. 3, June 1967, pp. 358-370.

Husén, Torsten, et al. International Study of Achievement in Mathematics, Vols I and II. New York: Wiley, 1967. 672 pp.

Kirst, Michael W. and Mosher, Edith K. Politics of Education. Review of Educational Research,Vol. 39, No. 5, December 1969, pp. 623-640.

Machover, Solomon. Cultural and Racial Variations in Patterns of Intellect; Performance of Negro and White Criminals on the Bellevue Adult Intelligence Scale. T. C. Contributions to Education, No. 875, 1943, 91 pp.; also Teachers College Record, Vol. 45, pp. 52-54, October 1943 (summary).

Mayeske, George W., et al. A Study of Our Nation's Schools. Washington, D.C.: U.S. Government Printing Office, 1969.

McKee, Paul G. and Brzeinski, Joseph E. The Effectiveness of Teaching Reading in Kindergarten. Cooperative Research Project No. 5-0371, Denver, Colo.: Denver Public Schools, 1966. 132pp.

McPartland, James S. The Relative Influence of School and of Classroom Desegregation on the Academic Achievement of Ninth Grade Negro Students. Journal of Social Issues, Vol. 25, No. 3, Summer 1969, pp. 93-102.

McPartland, James S. The Segregated Student in Desegregated Schools: Sources of Influence on Negro Secondary Students. (Final Report.) Baltimore, Md.: The Johns Hopkins University Center for the Study of Social Organization of Schools, 1968. 359pp.

Mitchell, James V.,Jr. Education's Challenge to Psychology: The Prediction of Behavior from Person-Environment Reactions. Review of Educational Research, Vol. 39, No.5. December 1969, pp. 695-721.

Peterson, Joseph M. Early Conceptions and Tests of Intelligence. New York: Harcourt, Brace & World, 1925. 320pp.

Smith, E. R. and Tyler, R. W. Appraising and Recording Student Progress. New York: Harper and Row, 1942. 550pp.

Stanley, Julian C. Principles of Scientific Research Are Not Pertinent to Educational Evaluation? Educational Researcher, Vol. 20, No.5, 1969. pp.8-9.

Stern, William L. The Psychological Methods of Testing Intelligence. Translated by G. M. Whipple. Baltimore, Md.: Warwick and York, 1914. 160 pp.

Stoddard, George D. The Meaning of Intelligence. New York: Macmillan, 1943. 504 pp.

Symonds, Percival M. Comparison of Statistical Measures of Overlapping with Charts for Estimating the Value of Bi-serial r, Journal of Educational Psychology, Vol. 21, November 1930. pp. 586-596.

Terman, Lewis M. Genetic Studies of Genius, Volumes 1-4. Stanford, Calif.: Stanford University Press, 1925-47.

Thorndike, Robert L. The Concepts of Over- and Underachievement. New York: Teachers College Bureau of Publications, 1963. 79pp.

Tyler, Ralph W. Constructing Achievement Tests. Columbus, Ohio: Ohio State University Press, 1934. 110 pp.

U. S. Office of Education. ERIC, Your Guide to Current Educational Information. Washington, D.C.: U.S. Government Printing Office, 1970. 8pp. (Order No.: 1970 0-374-712).

Walker, Dollie, et al. School Desegregation in Baltimore. Baltimore, Md.: The Johns Hopkins University Center for the Study of Social Organization of Schools, 1967. 47pp.

Washburne, Carleton W. and Heil, Louis M. What Characteristics of Teachers Affect Children's Growth? School Review, Vol 68, No.4, Winter 1960. pp.420-428.

Wesman, Alexander G. What Is an Aptitude? Test Service Bulletin No.36. New York: Psychological Corporation, 1948. 2 pp.

Wesman, Alexander G. Aptitude, Intelligence and Achievement. Test Service Bulletin No.51. New York, Psychological Corporation, 1956. 3 pp.

Womer, Frank B. What Is National Assessment? Ann Arbor. Mich.: National Assessment of Educational Progress, 1970. 52 pp.