Recollections and Reflections

C. Robert Pace

[Editor's Note: The following essay by C. Robert Pace is the first of a new initiative in the Handbook. Several distinguished scholars from other disciplines whose cumulative contributions are seminal to the development of higher education research literature have been invited to write autobiographical essays or memoirs recounting their professional careers and their reflections on the field of higher education. Higher education as a field of study was essentially founded by scholars from other disciplines, and those of us actively engaged in the field today owe them a great debt. We have much to learn from their insights over the past four or five decades.

The authors invited to prepare these essays have been given a free hand in shaping their contributions. They have been asked to focus on two broad themes. First, they were encouraged to share their thoughts on what brought them to higher education as a field and on the major developments and individuals that shaped their careers. Second, they were invited to share their perspectives on the evolution of the field and to offer advice to contemporary scholars on needed research. In essence, they were invited to recount both their "personal" reflections on their careers and their "professional" reflections on the field.

This initial contribution by C. Robert Pace will be followed in subsequent volumes with reminiscences by Wilbert J. McKeachie and Burton R. Clark. We hope you find their essays enjoyable and instructive—JCS.]

In a letter dated January 6, 1996, John Smart invited me to write an autobiographical essay/memoir for inclusion in the Handbook. He wished to begin a series of such essays by senior scholars whose activities have shaped the field of higher education research. I was, of course, honored by this invitation and said yes. Then from time to time over the next several months I wondered what to say, and how to say it. Looking back, I saw that my jobs had been the main influence on my thoughts about higher education—spanning a 60-year period beginning at Minnesota in the 1930s and ending at UCLA in the 1990s.

What follows is a personal recollection, because when I moved from Los Angeles to Arcata in 1993 I left nearly all of my professional books to
the Graduate School of Education at UCLA and many cartons of correspondence, reports, and records of my professional activities for the UCLA archives.

INTRODUCTION TO MEASUREMENT: THE MINNESOTA YEARS

In the fall of 1933 I enrolled as a graduate student in psychology at the University of Minnesota. All of us were introduced to Pavlov's dogs, Thorndike's cats, and Tolman's rats. Many of us also took Donald Paterson's course in individual differences. In the winter quarter one of the courses was Group Mental Tests, taught by Alvin C. Eurich. There were about 60 students in the class. We all took a series of tests—the Army Alpha, Army Beta, American Council Psychological Exam, a mental test developed at Ohio State, a Reading Comprehension exam, and several others. In each case we made a distribution of the scores for the class, computed the mean, standard deviation, percentile ranks, etc., and read the test manual and published norms. This was my first real encounter with measurement.

At the beginning of the spring quarter I went to see Dr. Eurich to ask if he knew of any graduate assistantship that might be available. Dr. Eurich was the assistant director of the University's Committee on Educational Research where several graduate students were employed. A position had just opened. It was to supervise 20 or so clerks who hand-scored objective achievement tests, converted the raw scores to percentile ranks, made item analyses, and prepared other information about test responses to report to the instructors. Most of the tests were from the General College, a new enterprise in the university, where all the courses were new courses, the instructors were borrowed from other university departments, the classes were large, and students' progress was indicated by their performance on comprehensive objective achievement tests. It was an experimental college in every sense, different from all other parts of the university. So, my job with the Committee on Educational Research put me in touch with the latest developments in educational testing and the latest developments in higher education curricula at the university.

Committee on Educational Research

This was a time of great activity in the history of testing. Many types of objective test items were developed—true-false, multiple choice, reverse multiple choice, matching, sequencing, items to measure application of knowledge and principles, interpretation of data, relationships, and more. On all the tests, the instructors received detailed information about students' responses to every item—its difficulty, discrimination, correlation with other items, etc. Such information presumably contributed to the improvement of examinations as well as enabling the instructors to see exactly what knowledge and skills the students had acquired. The connection between testing and teaching was apparent, because the content of test items should reflect the content of the course being taught. The intellectual and creative task of constructing reliable and valid objective test items is one of the best ways of clarifying the objectives of a course.

After about a year the job of supervising test scoring and item analyses was passed on to another graduate student, and that enabled me to work on various research projects in the Committee on Educational Research. Three events were especially important in my education and subsequent career.

In 1935, September 9 to 14, Minnesota held an invitational conference on research in higher education. More than 50 scholars came to this conference. Seminars were held on such topics as curriculum, instruction, objectives, student personnel and guidance, internal organization, and regionalism in higher education. The current state of knowledge was noted, and new research ideas were discussed. The participants were the leading contributors to this research; for example, Herman Remmers at Purdue, Coleman Griffith at Illinois, W.H. Cowley at Ohio State, Earl McGrath at Buffalo, Truman Kelley at Harvard, Fred Kelly at the U.S. Office of Education, and others from Columbia, Yale, Stanford, Chicago, Michigan, Nebraska, and elsewhere. I had been asked by Dr. Eurich to review the research literature on some of the topics and to attend the meetings. A bibliography of more than 250 publications was distributed. In all, the conference was a significant opportunity for me to see the scope of higher education research and get acquainted with some of the people who were doing it.

In 1936 President Coffman told Dr. Eurich that he would like to have the Committee on Educational Research make a survey of the occupational and economic status of Minnesota graduates, especially during the depression years. Between 1928 and 1936 there were about 14,000 bachelor degree recipients whose addresses were available at the alumni office. A one-page questionnaire mailed to each of them simply asked them to list their occupational and educational experience since they received their first degree, to indicate the dates of any unemployment, to indicate how soon after they got their first degree they got their first position, and, with respect to their first position and to their present position, how closely related it was to their field of specialization at the university. I was responsible for managing the survey, analyzing the data, and preparing a report of the results. More than 6,000 questionnaire returns were classified and tabulated—by year of graduation, gender, geographic region, and the different schools and colleges in the university such as engineering, business, education, liberal arts, agriculture, etc. The report of this research, A Follow-Up Study of Minnesota Graduates from 1928 to 1936, was published by the university in 1938; and the gist of it and its significance were briefly summarized.

Out of this picture of the trends in job opportunities and the occupational and financial status of recent college graduates should come a more realistic and sounder conception of the function of higher education. A Bachelor's degree is not an insurance policy against the effects of an economic depression. During 1932 and 1933 jobs for college graduates were harder to find, they were less likely to be in line with the student's preparation, they were at a lower occupational level, and they paid less money. If the value
of a college education is conceived in terms of immediate job getting and money making, then education was of limited value to the depression graduates.

Today, 1996, when many people judge the value of higher education by jobs and money, there is again a need for a more realistic and sounder conception of the function of higher education.

The third activity having long-range influence on my development was my Ph.D. dissertation. The topic was the measurement of attitudes. Psychologists typically inferred attitudes from stated opinions. I developed and validated a test to infer attitudes from stated behavior and further studied the relationship between attitudes and information.

My interest in attitude surveys and opinion polls has continued to the present day. For example, in studies of college students and college graduates over many years I have usually included measures of attitudes, and been concerned with how well attitudes predict action. I joined the American Association for Public Opinion Research in 1948 (it was founded in 1947), and have been a regular reader of Public Opinion Quarterly. How one thinks about questions and responses in higher education opinion surveys can benefit from acquaintance with the public opinion survey literature.

The General College

The General College at the University of Minnesota was established in 1932 to experiment with a new type of educational program, appropriate for students who wanted to come to the university but who were not eligible for admission to the existing colleges and schools because of low grades and aptitude scores, and the college would be for two years rather than four years. The first curriculum was set up on best guesses about the needs of the students. There were ten areas of study—human biology, physical sciences, euthenics, speech and writing, literature and the arts, psychology, current affairs, social problems, economics, and history and government. The course content was not the usual introduction to an academic discipline; the content was selected for its presumed relevance to the students' needs and interests. In 1935, with a grant from the General Education Board of the Rockefeller Foundation, the college established a student personnel service and research staff, and John G. Darley directed a comprehensive survey of General College students—their skills, interests, attitudes, problems, activities, and needs. This reflected the belief that knowing what to teach and how to teach requires knowledge about the students. Beyond a concern about the immediate interests and needs of current students, more long-range needs must also be considered, needs likely to be important as students become young adults facing problems of out-of-school living. So, a second major study was started—a study of former Minnesota students who were now young adults—which was also supported by the General Education Board of the Rockefeller Foundation. Locally we referred to this study as the adult study. In September of 1937 I became the director of it. With my Ph.D. completed in July, and thanks to the faculty mem-

bers who had been directing the study—designing the sample, developing the content, etc.—both having left Minnesota for other jobs, the job opportunity presented itself.

The sample for the study consisted of 1,600 names, 800 men and 800 women, 400 from each of the entering classes of 1924, 1925, 1928, and 1929, drawn proportionately from the four largest undergraduate divisions of the university, and selected at random from alphabetical lists within the four colleges. The content of the proposed questionnaire was organized under four broad areas—personal, home and family, vocational, and social and civic. Within each area the instructors were asked to prepare questions about the activities, needs and problems, attitudes or points of view they thought would be valuable in their teaching. The cooperative process of working and thinking to prepare the questionnaire took more than a year. When I became director, about two-thirds of the questionnaire content had been assembled. My job was to finish it, get it printed and distributed, analyze all the replies, distribute the results to all the faculty members, and write a final report of the study.

The final questionnaire, titled "Building the University of Tomorrow," was very attractively printed and illustrated with photographs and line drawings, easy to read, easy to answer in most parts, had altogether about 1,000 items, and filled a 52-page booklet. The questionnaires were mailed December 2, 1937, and three months later, after two follow-up postcards, a two-page follow-up letter, and two more postcards, nearly 70 percent of the adults who received them had filled them out and returned them. I do not know whether such a high rate of return could be obtained today. But I believe that the main factors accounting for the high return then are still valid and influential. In addition to the attractiveness of the questionnaire, the main factors were that it came from an institution held in high esteem and in which they all had a personal experience, and that the content dealt directly with activities, interests, and concerns in their everyday life.

Because getting a 70 percent return to a 52-page questionnaire may be unique in higher education and survey research, more information about it may be useful to present day researchers. At the beginning of the questionnaire booklet there was a message from Malcolm MacLean, director of the General College. Here is some of that message.

One of the best ways to evaluate education and to plan the education of the future is to discover what former students are doing now and what their experiences have been. That is why we are sending you this questionnaire. We hope you will find it interesting and challenging, and that you will derive as much benefit from the chance to think about your activities, problems, and points of view as we will derive from examining the results which you and 1600 other young men and women may send to us... During the winter and spring, we hope we may have an opportunity to talk personally with many of you who are living in Minnesota, provided, of course, that you are willing. If you are willing to help us further through an interview about some of your activities and problems, will you write your name and address on the following lines?
More than half of the questionnaires were signed, including more than 100 by people who were not even living in Minnesota. From about 260 Twin City residents we interviewed 172. Almost all of them had shown the questionnaire to their friends and had talked about it; said they found it interesting and stimulating; said it took a long time to fill out but was a pleasure; and liked the attractive way it was put together. Nearly 100 of them asked if they could have copies of it to keep.

My advice to questionnaire makers is this: If you really put the time and effort and thought that is needed to construct a good instrument, one that is clear and can be answered from experience, that deals with topics of definite relevance and importance, and that is attractively presented, the chances are fairly good that those who get it will answer it.

From the experience of analyzing and reporting the answers to nearly 1,000 questionnaire items I came to some conclusions about research methods which I thought then and still think today were useful. The sample of names for the survey had been drawn from entering students and so, of course, about half of them subsequently graduated and half did not. One could compare graduates and non-graduates, all of whom were initially qualified for admission to college and all of whom began college. A control group like this may be impossible to get today because a much higher percentage of students eventually graduate, and many of them attend more than one college. In the adult study the questionnaire respondents were divided into eight groups and the same complete analysis of responses was made in each group. The eight groups were identified first between men and women, then for each of these two groups the cases were divided by year of entrance to the university (1923-4 and 1928-9), and then each of those four groups was divided between graduates and non-graduates. Within each group we calculated the percent of adults giving a particular response to every one of the nearly 1,000 items.

One of the advantages in this procedure was to see what general consistency there might be in the responses. Was there a significant difference between graduates and non-graduates on all comparisons or only on one or two? At no time did we ever consider percentages that were based on the total number of respondents. The responses of all groups to all items were reported to all General College instructors and staff members. The detail was in one sense necessary and in another sense overwhelming. One of my contributions to the content of the questionnaire was to include several measures for which normative or other comparative data was available so that scores could be obtained. These included a job satisfaction score, a measure of the economic and cultural status of the home, a liberal-conservative attitude scale, and a measure of general adjustment and morale. Ten years later when I again had a chance to do a follow-up study of former college students nearly all the content of the questionnaire I developed consisted of scales and other sets of items that could be scored.

This is not the place to report the results of the adult study. I will say only that except for occupational level, income, and job satisfaction, the differences between graduates and non-graduates in other aspects of life were not very obvious and frequently non-existent. The story of this adult study—why, how, and with what result—is contained in a book I wrote, They Went to College, published by the University of Minnesota Press in 1941.

FROM MEASUREMENT TO EVALUATION: THE COMMISSION ON TEACHER EDUCATION

I left Minnesota in the fall of 1940, having been offered a post-doctoral fellowship by the Commission on Teacher Education, paid by the General Education Board of the Rockefeller Foundation, to spend at Teachers College, Columbia. The Commission, located in the offices of the American Council on Education in Washington, D.C., was embarked on a cooperative study with colleges and school systems for the improvement of teacher education. In the participating institutions a lot of attention was paid to evaluation. Maurice Troyer on the Commission staff was the head of the evaluation activities.

Much of what we did at Minnesota was evaluation and properly called evaluation. Ruth Eckert was employed to do an overall evaluation of the general college programs, and during my last year at Minnesota I worked with her on several projects. In all major areas of the curriculum students took achievement tests at the beginning and the end of the year and in some areas they took the achievement test a year after they had taken the course to see how much had been retained. There were also measures of reading comprehension and of effective writing, and of attitudes, interests, values, vocational plans, and more. The evaluation of student learning and development at the General College was systematic, comprehensive, and reliable. My particular experience at Minnesota, however, was mainly to carry out a specific project.

With the Commission on Teacher Education, the focus was on providing service to the people in the colleges and schools who wanted to evaluate what they were doing. The role was that of a consultant. In the beginning, at Teachers College, the fellowship allowed me to do some things on my own, so I could get better acquainted with teacher education and improve my understanding of measurement and research. I attended the Foundations course that all students had to take. With Irving Lorge, who invited me to come to his course on measurement, I had many discussions about research and evaluation. For the Commission I constructed an exam for the Foundations course, made visits to several colleges to discuss their evaluation activities—the normal school in Oneonta New York, Syracuse University, and Buffalo State College—and attended a workshop where representatives from all the cooperating colleges discussed their evaluation problems. At the end of the year in New York I was invited to join the Commission staff in Washington, D.C., where for the next two years Maurice Troyer and I handled the evaluation activities of the Commission.
In my files at home I did not find a list of all the colleges that were involved in the Commission’s work. I know that Dr. Troyer and I visited most of them several times, serving as consultants on the evaluation activities they were undertaking. Among the colleges were Ohio State, Wayne State, Nebraska, Chicago, Texas, Milwaukee, Stanford, Furman, Southern Illinois, William and Mary, Michigan State, and Troy Alabama. We organized workshops and conferences, facilitated communication among the colleges, assembled files of evaluation instruments. Among the evaluation techniques were profiles, rating scales, check lists, essays, achievement tests, case studies, guides to classroom observation, diaries, etc. My experience with the Commission enabled me to become acquainted with institutions and individuals all across the country and see the myriad methods and materials being used in evaluation.

For me, the most important lesson I learned was an attitude or philosophy about how education is improved. This point of view was described briefly in the book Maurice Troyer and I wrote as follows:

The cooperative study of teacher education emphasized implementation rather than research and survey. Its purpose was to work with groups on their problems rather than to organize a program or to present formal recommendations. It was to work through and to develop local leadership, and to stimulate thinking and experimentation on basic problems in teacher education. Because there were obstacles to the free play of local initiative, the staff of the cooperative study became increasingly interested in how changes were brought about—in the strategy of effective planning. They and the Commission believe that, in the long run, greater progress in teacher education would be made by each institution striving to improve its own program than by any national organization trying to lay down standard recommendations. Thus, in the cooperative study, responsibility for the improvement of programs and procedures remained within each institution. All this had a direct bearing on the philosophy and services of the evaluation division. It meant that there were no comparative studies, no tests centrally developed for use by all the cooperating centers, no prescribed sequence of steps for each to follow in evaluating itself. There was, in other words, no evaluation of schools and colleges by an external agency; rather, there was a working with schools on evaluative tasks chosen by them.

If one purpose for evaluating a program is to improve it then how the evaluation is conducted becomes important. The people whose actions are necessary to change a program need to be active participants in the evaluation. It is sometimes easy to reject recommendations from an evaluation made by someone else but not by an evaluation they themselves have made. Evaluation needs to be a cooperative, collaborative undertaking. The process may determine what is done with the product. Evaluation should not only contribute to learning; it should itself be a learning activity. The simple fact that evaluation has human consequences means that how people are treated will influence what they learn from evaluation. What does a student learn from the teacher’s evaluation? What does a teacher learn from the evaluation made by a supervisor? Has the evaluation contributed to learning? The final report of the evaluation activities includes extensive examples of what was done in the colleges in their teacher education programs—evaluation related to selection and admission, orientation programs, general education, professional education, practice teaching, follow-up, and evaluation in-service on the job. This sequence is one way to view an institutional evaluation plan, with concern about the consistency and relationships among the stages. Another integrating focus is a cumulative record kept by the student. This is described as self-evaluation under guidance. In today’s language this would be described as a portfolio.

The philosophy and practice of evaluation exemplified in the work of the Commission is very similar to what is being advocated today under the label of assessment. Both are seen as avenues for learning and development, for change and improvement. And how the work is conducted is probably as important for assessment as it was for evaluation. Our own views and the record of what the colleges did is reported in the book Maurice Troyer and I wrote, Evaluation in Teacher Education, published by the American Council on Education in 1944.

OPINION POLLS AND ATTITUDE SURVEYS: THE BUREAU OF NAVAL PERSONNEL

Lieutenant Commander Alvin C. Eurich was the officer in charge of the standards and curriculum section in BuPers. He asked me to join the staff. I did, not as an officer, but as a civilian scientist in the civil service. This activity in the Bureau constructed the Navy's general classification test, and other personnel selection instruments such as reading comprehension, arithmetic, mechanical aptitude, etc.; wrote training manuals and achievement tests for Navy specialties such as Gunner's Mate, Electrician's Mate, etc.; and carried out studies predicting performance from test data, evaluated the instruction and utility of various training programs, and kept in touch with similar programs in the Army and Air Force. I worked in the research unit of this Navy program.

After a year or so most of my work consisted of developing, analyzing, and reporting the results of Information Surveys which I initiated. These were opinions and attitudes of enlisted men regarding their training, and other aspects of their naval experience. The first of these Information Surveys asked men for their opinions about the training they received in their specialty. In other surveys there were items about such topics as job satisfaction, attitude toward officers, pride in outfit, Navy fairness, efficiency, importance. The most extensive survey was an evaluation of the educational services program in the Navy which was under the direction of Lieutenant Commander Earl McGrath. This program included off-duty classes, courses available from the Armed Forces Institute, getting information to ships and bases about progress in the war. The questionnaire was given to samples of officers and men in eight advanced bases in the Pacific, personnel on nearly all types of ships mostly from the Pacific, and the ship's company from one training center in the U.S. A series of nine reports was produced—Getting
RECOLLECTIONS AND REFLECTIONS

The final experience from my Navy Department days relates to the creation of the Office of Naval Research. My Minnesota classmate and close friend, Jack Darley, was involved with a few others in drawing up the plans for the Navy to contract with universities for conducting research of interest to the Navy. Jack discussed the plans with me from time to time. ONR was established and a review committee to read and recommend proposals was named. I was designated as the official representative on the committee from the Bureau of Naval Personnel. I knew many of the psychologists on the committee—Jack Darley, Rensis Likert, Lowell Kelley, Carol Shartle, Donald Marquis—and others on the committee I knew only from reading but not in person—Margaret Mead, Ruth Benedict, Erich Fromm. At our first meeting, after we asked Ren Likert to leave the meeting, we approved his proposal for studies of leadership, supervision, and productivity with a long range grant that made possible the development of the Survey Research Center at the University of Michigan.

As that Center grew and expanded over the years it became, in my view, one of the two or three most productive and influential centers of social science research in the country.

Developing Scales for Higher Education

After the war, during the six months between my departure from the Navy Department and my arrival as a faculty member at Syracuse University, I had an opportunity to design an alumni survey for American University. Remembering the complexity of the Minnesota study and my recent knowledge about Guttman scaling theory and methods, I decided to see if I could construct scales for the alumni survey. These would measure aspects of adult life that could be related to general education—to humanities and arts and sciences. For each of seven areas of adult experience—political affairs, civic affairs, art, music, literature, science, and religion—I wrote check-lists of participation. For each topic there were ten activities and alumni were asked to check each activity they had engaged in during the past year. The activities in each topic ranged from ones that were commonplace to ones that suggested a greater level of interest, and that the uncommon activities subsumed participation in all the more common ones. In other words the content of each set of activities would form a unidimensional and hierarchical universe of content as in a Guttman scale.

Here is a brief example: if a person contributed a regular sum of money to the church, you can be pretty sure that person is a member of the church, and attends church services; or if a person says he voted in the last primary or local election, you can be pretty sure that person also voted in the last national election, has listened to political speeches, reads about current events in the paper, and talks about politics with friends. In analyzing the responses to each of these ten item activity check-lists, I found that usually seven or more of them did form a scale as defined by Guttman. So, I learned that I could develop scales to measure some important outcomes of higher education. This had a big influence on much of my
future research and on my views about measurement. In achievement testing the
length of the test contributes to its reliability. Tests of 100 or more items are typi-
cal. With a good scale, however, one can get good reliability with fewer than a
dozens items. This means that the range of ones inquiry can be greatly extended
and still be feasible administratively. It also means that one must carefully define
the concept being measured and determine empirically whether it is a scalable
universe of content.

TEACHING, RESEARCH, ADMINISTRATION: THE SYRACUSE
UNIVERSITY YEARS

When I went to Syracuse in September of 1947 it was my first employment as a
university faculty member. The trustees and administration of the university had
decided to undertake a comprehensive survey of its programs and services that
would be useful in planning for its future development and needs and that this
should be a self-survey. Maurice Troyer, head of the Evaluation Services Center
at Syracuse, was asked to be director of the self-survey. At the same time, Dr.
Troyer was developing a graduate program in higher education. The need for
research in the self-survey plus the need for teaching courses in higher education
created the job opportunity for me.

The Self-Survey

The self-survey dealt with a broad range of topics: curriculum and instruction,
personnel services, administrative organization, plant and facilities, finances,
research and scholarly productivity, library, public relations, with each topic stud-
ied by a committee. Every committee included faculty members selected by the
academic senate, administrators, and a trustee. In going around to many commit-
tee meetings, I could help them gather the information they wanted, merge over-
lapping requests, devise questionnaires as needed, etc. Altogether at least 100
faculty members and administrators participated in this work, and without any
released time from normal responsibilities. The Evaluation Center staff adminis-
tered the Cooperative General Culture Test and the Current Affairs test to sam-
ple of sophomores and seniors. Faculty and students indicated how they rated
the importance of each of a list of 18 objectives of general education, with stud-
ents also indicating how much they thought they were getting toward these
goals, and faculty indicating how much responsibility their area assumed for
helping students attain the objective. And other systematic data gathering surveys
and ratings were undertaken. Each committee prepared a report of its inquiry and
its recommendations during the summer of 1948. These reports added up to
nearly 1,000 pages. At that point I was asked to prepare a condensed report of
less than 100 pages which would accurately reflect what each committee did and
which could and would be read by all the members of all the committees, and
ultimately by all the faculty members and administrators in the university, and
would clearly indicate all the important recommendations that were made. In the
fall of 1948 representatives of the survey committees met with the University's
top administration for three days to discuss the findings and lay plans for translat-
ing recommendations into action. At this point the public relations office argued
that the reports should be carefully guarded lest the Syracuse newspapers would
write articles about all the things that were wrong at the university. Happily, most
of the people at the meeting rejected that view. Ultimately, a 75-page self-survey
report to faculty was attractively printed and very widely distributed.

What a wonderful way to begin a career as a professor of higher education! In
just over a year I became acquainted with all the university's administrators,
many of its faculty, some of the trustees, and they with me. I learned how the
place operated—not only its curriculum and instruction, but also its financial
structure and needs, its administrative organization, graduate and research pro-
grams, student personnel services, its plant and facilities, and other features.

The Chancellor's Office

Not long after completion of the self-survey the Chancellor, William P. Tolley,
asked if I would be willing to spend part of my time working in his office as a
Special Assistant to the Chancellor. I said OK provided he understood that I
would be a special assistant, not an administrative assistant or an expeditor or a
gopher. For the next three years I had a chance to write some speeches for the
Chancellor, assemble materials for his annual reports to the Board of Trustees,
answer some of his mail, and participate in many discussions about university
policies, personnel, and budgets, and at his request I talked with Deans and
Directors about their aspirations and their budgets.

Before WW2 Syracuse had been mainly an undergraduate institution with an
up-State Methodist clientele. After WW2 the University, under the Chancellor's
leadership, rapidly grew into a comprehensive research university, with a diverse
student body, an aggressive research enterprise, and a big-time athletic program. I
sat in on many discussions where these decisions were made—how to organize a
research institute, how to determine and distribute overhead, etc. There had been
no institutional research office so I instituted various data sets such as informa-
tion about class size, teaching loads, faculty ranks, instructional costs. Once,
when the Chancellor was out of town for more than a month prior to the annual
meeting of the Trustees, I wrote the report from beginning to end, and the first
time he saw it was when his secretary handed him a copy at the Board meeting.
The opportunity to see how the university really worked, from inside a chancel-
lor's office, what problems and issues were discussed, how decisions were made,
and who made them, all added greatly to my knowledge about higher education. I
saw that the knowledge and experiences that characterize life in a chancellor's
office are very different from the day-to-day knowledge and experiences of fac-
ulty members. I have subsequently felt that misunderstandings between adminis-
The Evaluation Center

The Evaluation Services Center had played a significant role in the conduct of the self-survey. The Center had a director and an associate director, part-time, a full time secretary, and three or four graduate student research assistants, mainly funded by the Chancellor’s office. Several research inquiries for the self-study were carried out as doctoral dissertations by research assistants in the Evaluation Center. After I had been at Syracuse two years, Maurice Troyer left to become vice president of a new International Christian University in Japan, so I became director of the Evaluation Center. The research assistants in the Center were always doctoral candidates in the graduate program in higher education. There were four basic courses in the doctoral program—history, administration, curriculum, and evaluation. I always taught the evaluation course, and sometimes the history or curriculum course. Altogether at Syracuse I supervised twelve dissertations in higher education, six of them by research assistants in the Evaluation Center. In addition to the studies of objectives and achievement tests previously noted there was a study of religious beliefs and social values and a study of Syracuse alumni.

The alumni survey questionnaire was a revision and extension of the questionnaire developed at American University. There were Guttman-type activity scales related to Politics, Civic Affairs, Religion, Literature, Music, Art, and Science, with each scale having eleven items. Then there were opinion measures related to Politics, Civic Relations, Government, The World, Philosophy, Literature, Music, Art, and Science, with each measure having six statements. And, there was the same list of eighteen objectives that had also been used in the studies of students and faculty. The survey was mailed to graduates of the class of ’47 and to classes at five year intervals back to ’27. The response rate was 52 percent.

With respect to the eighteen objectives there was substantial agreement among students, faculty, and alumni as to their importance—the rank order correlations ranging from .81 to .91. There was also high agreement between students and alumni regarding their progress toward the objectives; and between the ratings of importance and progress in both groups. The alumni survey scores on the Guttman-type activity scales and the opinion measures were analyzed in relation to their college major field, and in all instances the pattern of results for adult activities and opinions was congruent with the undergraduate major field. For example, adults who had been Fine Arts majors were also the highest participants in activities related to the arts and in having opinions similar to experts in the arts. This same clear-cut relationship was found when students’ scores on the Cooperative General Culture Test were analyzed—high achievement was always related to the students’ major field.

With respect to religious beliefs and social values there was no difference by majors among freshmen, but among seniors the highest scores were made by students in the humanistic and service-oriented curricula and the lowest scores by students in the scientific and technically oriented curricula, so that one would conclude that scores do not seem to be incidental products of just any kind of college program. On all the other studies of students and alumni it seemed clear that college makes a difference and the particular kind of college education also makes a difference, and that the pattern of activity and opinion among alumni is remarkably similar to the pattern among students.

The network of associations that emerges from these studies is still a basically correct picture of college influences on student outcomes. Perhaps one of the reasons for this stability is the reliability of the measurements. My personal interest in measurement has always reflected my belief in the importance of reliability and validity. For the Evaluation Center studies, here are some examples. In the alumni survey we had a sample of respondents answer the activity scales and opinion measures six months later. The test-retest correlations ranged from .83 to .89 for the activity scores and, for seven of the nine opinion measures from .60 to .70 with two low ones. For the activity scales 85 percent of the responses were identical on the two occasions, and for the opinion items 75 percent were identical. On the measures of religious beliefs and social values, test-re-test correlations were .93 and .87. Validity is apparent from the content of the measures, but also from the fact that results fit into previously known and predicted outcomes.

The Psychology Department

In the fall of 1952 my role at Syracuse changed. I was chosen to be the chairman of a newly formed psychology department. Prior to that time psychologists at Syracuse were located in various parts of the university—some in Liberal Arts, some in Education, one or two in the School of Business, and one in the Maxwell School of Citizenship. Now all the psychologists would belong to a single all-university department. Psychology at Syracuse was not accredited in anything—not in clinical or counseling, not by VA or Mental Health—and there were no regular faculty members in experimental or physiological psychology and no laboratories.

So the job was to help construct a balanced department, become fully accredited, and promote research. Fortunately, this was a period in higher education of growth in magnitude and in the support of research. At Syracuse three people would be retiring within a year or two. We were authorized to hire replacements and to seek two more. I got money from the National Science Foundation for laboratory equipment. In less than two years we had some outstanding young scholars on the faculty; in three years we were fully accredited in everything, we had many research contracts and grants; and within five years we were rated as among the top 20 departments in the country.

I served as department chairman for nine years. We were always well represented at the annual meetings of the American Psychological Association, and
the American Educational Research Association. Locally I spent much of my time just talking with every member of the department so as to know fully what research was being carried out and why it was important. We had monthly departmental meetings where procedures as well as policies were openly discussed and voted on. The larger success story of the department was owing to the fact that good people could be hired and money could be obtained for research grants, training programs, and student support, not to any virtues I may have had as an administrator.

When I indicated that I no longer wished to be chairman it was agreed that I would spend my final year working with the administration and the faculty in a national search for a new chairman. This was a very active search, to which everyone in the department devoted a great deal of thought, and which resulted in coming down to three excellent choices that we were ready to send to the Dean with our recommendations in rank order. The Dean, of course, had met all of them when they had come to the campus for interviews. Then the Dean arbitrarily named a person who had not been a candidate or even indicated an interest in the position, and who was known by but not wanted by most members of the faculty. Moreover, the Dean never met with the department to explain his action. As a consequence nearly all the outstanding people who had been hired left Syracuse within the next few years for positions elsewhere and the department's strong reputation rapidly disappeared. I report this event as an illustration of the importance of good faith between faculty and administration, and the consequences of bad faith.

National Connections

During the years I was at Syracuse I had the opportunity to work with several national organizations—the College Entrance Examination Board, the American Council on Education, the Social Science Research Council, the Fund for the Advancement of Education of the Ford Foundation, and the Carnegie Corporation. Frank Bowles, president of the CEEB, invited me to be a member of the Board’s research advisory committee. In the past, all of the research activities of interest to CEEB were performed by the Educational Testing Service. I persuaded the Board that, at least from time to time, it should solicit research proposals from people who had no connection with ETS. One year I submitted a proposal that was funded, and Anne Anastasi also received a grant. I also suggested to the Board that it should have a few people on its research advisory committee who were not from the educational testing establishment and I recommended that Paul Lazarsfeld be invited to bring a sociological and societal perspective to the Board's programs.

For several years I was a member of the American Council on Education’s Committee on Measurement and Evaluation. This committee recommended that the Council approve the creation of the Educational Testing Service. It was this committee that also sponsored the Cooperative Study of Evaluation in General Education, and appointed Paul Dressel to direct the study. A major product of that national study was the development of several tests of critical thinking. These tests represented the thinking of many faculty members about what was meant by critical thinking, and what test items would adequately reflect it—in social studies, communication, sciences, and humanities. People today who think that objective achievement tests can only measure facts are surely not aware of the report of the Cooperative Study by Paul Dressel and Lewis Mayhew.

My notes at home do not indicate when the Social Science Research Council’s Committee on Personality Development in Youth was formed. At any rate I was a member of the committee for the duration of its existence. Ralph J. Tyler was the chairman and the other members included Dana Farnsworth, Chester Harris, Don Marquis, T.R. McConnell, Lloyd Morrisett, Ted Newcomb, Nevitt Sanford, and Robin Williams. A major purpose of the committee was to interest a variety of social scientists in the study of personality development during the college years. The committee provided some financial support to Nevitt Sanford for his book on the American College, and to Chester Harris for a book on the measurement of change. The committee organized a three-day conference at Amherst to which more than 50 researchers were invited to discuss the present status of knowledge and encourage new lines of inquiry. The committee also enabled me to hold a three-day meeting at the Council’s offices in New York to discuss personality measurements for outcomes of higher education, with the following invitees: Ed Bordin, Ben Bloom, Jack Darley, Jane Loesinger, Don Marquis, Lloyd Morrisett, T.R. McConnell, Dave Saunders, George Stern, Harold Webster, Dick Christie, and Robin Williams. Many personality measures are not suitable for measuring outcomes of higher education because they presumably measure stable traits. At the meeting we made a list of important objectives of higher education and for each we listed measures that were available for those objectives. The use of relevant personality measures has an important influence on higher education research because it turns attention to the total college environment—to outcomes that do not come from courses exclusively to ones that look beyond courses and curricula to other features of college life. This wider look at the total college environment can enrich our understanding of higher education. Subsequent events demonstrated that it did.

In 1951 the Ford Foundation established the Fund for the Advancement of Education, with Clarence Faust as president and Alvin C. Eurich as vice president. Over the next several years I had many contacts with Al Eurich—discussing policies and programs with him, and participating in some of the Fund’s projects. My own views often differed from those of the Fund, but that never interfered with the frequent support I got from the Fund. I tried to interest the Fund in setting up some major research centers to focus on bringing together various behavioral scientists in studying major educational problems, but the Fund was more interested in national demonstration projects such as the improvement of teacher education, the use of television in the schools, early admission and advanced
placement in college, promoting equal educational opportunity. Moreover, the possibility of connecting education and the behavioral sciences was ended when the Ford Foundation decided to abolish its Behavioral Sciences Division.

I contributed to several Fund projects. For a number of years money was given to colleges in Arkansas to raise the level of liberal education of prospective school teachers. As the program was ending the Fund offered small planning grants to the colleges to think about how they might maintain some of their activities. I was asked to meet with the presidents, deans, and others to talk about the values that might be gained from these local grants, and then to visit many of the colleges in Arkansas to answer questions, review their plans, and make suggestions. The experience enlarged my knowledge and appreciation of the diversity of higher education in the country.

Another program was the college teaching internship program where several colleges agreed to employ graduate students who were near to completion of their dissertations and give them an introduction to what college teaching was all about, and what it was like to be a faculty member. They visited classes of other professors and discussed their observations, and in most places they also had seminars about higher education, student characteristics, learning, counseling, etc. At the end of the year they were asked to write their observations and thoughts about teaching; and first year instructors who were not interns were also asked to comment on their year's experience. I was asked to read and summarize all these expressions. The task required familiarity with the methodology of content analysis. What impressed me was the difference in attitude between the interns and the regular instructors, with the interns being much more aware and appreciative of student interests, motivations, and personalities. For the colleges the main impact was that college teaching became a normal topic of conversation among the faculty whereas previously faculty members rarely talked about what they did in their classes.

The Fund gave a grant to Sweet Briar college to evaluate the Junior Year in France program which was administered from the college, and which prior to the war had been administered by the University of Delaware. I was asked to prepare a questionnaire to be sent to alumni of this program, the main purpose of which was to increase international understanding. I developed and pre-tested Guttman-type scales on six topics—international activities, language and cultural activities, policies regarding the exchange of people and information, the role of the United Nations, and policies of the U.S., and acceptance of people of other cultures. There was also a short test designed to tap knowledge about other cultures and countries by asking people to think of the names of outstanding contemporary contributors to literature and the arts, science, philosophy, and other fields from the United States, Great Britain, France, and any other countries. So here was another opportunity to indulge my interest in developing new measures of important variables that had not previously been measured.

In the post-war period there were 14 colleges that had sent 10 or more students to the program, and from those colleges we got a control group of similar liberal arts students. The results of the study were clear. Compared with their contemporaries from the same colleges, alumni of the Junior Year in France program were: 1) personally more tolerant in their acceptance of people who differ from themselves; 2) more fully aware of significant intercultural contributions to life in the twentieth century; 3) more frequent and more active participants in internationally-oriented activities both of a political and cultural sort; and 4) more inclined to endorse policies which promote the freer exchange of ideas, goods, and people among countries. These differences could not be attributed to travel by itself, or to language majors, or to gender. My report of this study, The Junior Year in France, was published by the Syracuse University Press in 1959.

Beginning a New Line of Research

The combination of scores on the Scholastic Aptitude Test and high school grades usually predicted college success (freshman grades) by a correlation of about .50. Adding other variables to the prediction equation did not seem to increase the correlation by any great amount. As these studies were discussed in the research advisory committee of the College Board, I wondered whether the relationship might be improved by expanding the criterion end of the equation. Perhaps one could think of college success as involving something more than a grade point average. The Board appointed a subcommittee of Quinn McNemar, Anne Anastasi, and me to think about non-intellectual factors related to college success, including factors in the environment as well as in the individual.

One of the new psychology department faculty members at Syracuse was George Stern, who was familiar with Henry Murray's theory of personality needs and environmental press. A press is an aspect of an environment that is compatible with a personality need. A person with a need for order would like an orderly environment, for example. Performance is best when need and press are similar. Stern had constructed a questionnaire to measure 30 of the personality needs described by Murray. So we proposed to the College Board to construct a set of environmental characteristics parallel to the personal characteristics in Stern's questionnaire. With funds from the CEEB we constructed the College Characteristics Index (CCI) and tried it out on samples of students at five very different colleges. Responses to some of the statements of press were clearly different between some of the institutions, and this suggested the possible value in further exploration of the need-press idea. Much more information was needed about the presumed parallelism between needs and press, for example how well can it be demonstrated empirically. And more thought needed to be given to the content of the CCI items.

After the completion of the report to CEEB I got a discretionary grant from John Gardner at the Carnegie Corporation that enabled work to continue on the revision of the CCI and the collection of more data. Meanwhile I became increasingly uneasy about interpreting the CCI as if college environments had the same
characteristics as personalities. We revised the CCI twice and by the summer of 1959 we had data from about 50 colleges and universities. By then it was also clear that Stern and I had very different ideas about how the CCI should be analyzed and interpreted. Stern analyzed student responses to the CCI undifferentiated as to the institution they were describing. My interest was in differentiating between institutions so that the college, not the student, was the proper unit of analysis. Also, correlations between pairs of need-press responses revealed that in quite a few cases they were not parallel. In one or two cases the correlation between need and press was negative. Empirically the presumed relation between need and press was not confirmed in some of the comparisons, and therefore it was not appropriate to report all the environment characteristics as if they were parallel to the personal needs characteristics.

From September 1959 to September 1960 I was a Fellow at the Center for Advanced Study in the Behavioral Sciences. During that year I analyzed institutional differences in CCI scores and, from a factor analysis of institutions, obtained four main ways in which environments differed as follows: 1) an intellectual, humanistic, esthetic emphasis; 2) a friendly, group welfare emphasis; 3) a scientific, independent emphasis; and 4) a practical, status-oriented emphasis. In my thinking these four dimensions replaced the 30 environmental press dimensions. Then I also made a content analysis of the 300 items in the CCI to see how well the items reflected characteristics commonly considered in accreditation reviews and college self studies. There were very few items referring to administrative or to academic characteristics and too many items referring to student characteristics. So I developed an instrument in which administrative, academic, and student sources of press were in balance. This questionnaire, called the College Characteristics Analysis (CCA), was also diagnostic of academic and student subcultures in the college. Half of the CCA items were also in the CCI, and half were newly written, for a total of 180 items in the CCA.

At the Behavioral Sciences Center one of the most valuable benefits was the opportunity to have many discussions with scholars from other fields and for me especially with sociologists, anthropologists, political scientists, philosophers, as well as other psychologists and educators. Just as the measurement of individual differences had led to many insights about college students, so now the measurement of college environment differences may lead to further insights about higher education. Before leaving the Center I outlined plans for a ten-year research program.

Back at Syracuse we got CCI responses from more institutions. I received a contract from the Office of Education for a study of academic and student subcultures in college, using the College Characteristics Analysis, which I transferred to UCLA when I changed jobs in the fall of 1961. When I left Syracuse I had available CCI data from about 80 colleges and universities. Since Stern and I were co-authors of the CCI we obviously needed to reach some agreement about who could do what in the future. The result was that Stern would distrib-

ute and analyze the CCI as a counterpart to his inventory of personality needs; and I would use half of the original items in an instrument based on environment differences and without regard for the personal needs in Stern's inventory. Subsequently, at UCLA I found that nearly half of the CCI items did not discriminate between environments, so I produced a 150-item questionnaire with the most discriminating items and called it College and University Environment Scales (CUES).

**UCLA: ENVIRONMENT, EVALUATION, EFFORT**

The professor of higher education at UCLA was Malcolm MacLean, who had been my first employer when he was the director of the General College at Minnesota. His retirement led to my appointment—another instance of my many Minnesota connections over the years. UCLA had a new chancellor, and a dean of Education whose plans for the school emphasized the role of research. I had never before been employed full-time in a School of Education but I knew a few people at UCLA in education and in the psychology department and I accepted the job offer with confidence and pleasure.

**Studies of College Environments**

For nearly a decade much of my research was devoted to studies of college environments. A factor analysis of responses to CUES from 50 institutions led to the production of five factors—labeled scholarship, awareness, community, propriety, and practicality. These same five factors were always reconfirmed in subsequent analyses. Also, from the beginning, the scoring of CUES was unique. Students' responses were interpreted as an opinion poll. Students were reporters about the environment. When students agreed by a margin of two to one or greater that something was generally true about the college, the statement was counted toward the score. The number of such statements along a particular dimension—scholarship, awareness, etc.—was the score on that dimension. This is not a mean or average score, it is a single score that characterizes the institution. The two-to-one definition was an arbitrary choice. I just regarded it as sufficient to describe something as characteristic or dominant. The difference in result between this method scoring and a score based on the average of student responses is dramatic. For example, if there are 20 statements about the environment and student responses are divided 50-50 about whether each statement is generally true or generally false, the average of student responses would be 10. In the rationale for scoring CUES the institution's score would be zero because no statement was regarded as characteristic by a margin of 2 to 1 or greater among qualified reporters. Also, since there is no mean or variance, the usual methods of estimating reliability are inappropriate. Test-retest reliability can be noted just by seeing whether the college's score is the same on two occasions. It almost always
was. Moreover if the first score was 10 and the second score was 11, the second score almost always included the same ten items plus one more. The stability of a score depends on the number of items where the percentage is close to the dividing line of being counted or not counted. CUES was published by Educational Testing Service in 1962.

The research contract to study academic and student subcultures was officially transferred from Syracuse to UCLA as of January 1962, and continued until the final report was issued in 1964. The content of the CCA had three parts, each with its own directions. Part 1, The College or University as a Whole, consists of items that refer to regulations, policies, facilities, and other features of campus-wide relevance. Part 2, Your Major Academic Field, consists of statements about professors, classes, teaching, etc., which the student answers with respect to the academic part of the college he knows best—his own academic major field. Part 3, Your Student Colleagues, is answered with respect to the students and student activities one knows best. Thus, the test provides an indication of the source of press (administrative, academic, and student) as well as the direction of press (humanistic, scientific, welfare, practical). Nine institutions participated in the study. Three were small and not likely to have any deviant subcultures; two were somewhat larger; and four were large complex places assumed to have numerous academic and student subcultures. On the back of the CCA answer sheet students indicated their progress toward each of eleven objectives, their grades, and their satisfaction with college. Also at each of the colleges measures of personality were obtained. These differed in the different schools. The measures included the Allport-Vernon-Lindzey values, sections from the Omnibus Personality Inventory, the Heston Personal Adjustment Inventory, Stern's Activity Index, portions of the California Psychological Inventory.

The general purpose of the CCA was diagnostic. In the larger institutions are there many academic subgroups whose environmental press differs significantly from the institution as a whole? Which has the greater impact on attainment—the subgroup or the total group? What happens to similar types of students in different types of subgroups, and to different types of students in similar types of subgroups? In other words, which is more frequently related to attainment—student characteristics or environmental press? From the many analyses made in this research the results indicated that the largest influence on the attainment of general objectives came from the largest stimuli (the total institution), and that subgroups are smaller stimuli having smaller influence. Also differences between institutions were larger than differences between subgroups. The environmental press of a subgroup was more frequently related to attainment than the personal characteristics of students in the subgroup. Viewing each institution as a case study, one could find many interesting differences in the main sources of press—for example, the press toward humanistic objectives might come almost entirely from faculty sources at a particular college, or, at another college the practical, status-oriented press might come mostly from student sources with some added support from administrative sources but no support from faculty sources.

When this study ended, no further research with the CCA was undertaken. I had suggested to ETS that the CCA would be especially well suited for large universities; but ETS decided not to expand its activities in higher education testing and research.

Meanwhile, many more studies of CUES were underway. One series of studies, supported by the College Entrance Examination Board, compared CUES results from different groups of reporters. CUES scores were highly stable over a period of one or two years; almost identical results were obtained from any group of qualified reporters; and results were not influenced or biased by the abilities or the academic success or the personalities of the reporters. Freshmen, however, are not qualified reporters. It may be valuable to have them indicate what they expect will be true about the college and to compare their expectations with what others say is really true.

Over a period of several years, ending with a report in 1967, a grant from the U.S. Office of Education supported many new analyses of the measurement qualities of CUES, a clearer definition of the five factors, the development of a national baseline, the identification of institutional types, and much evidence documenting CUES validity. CUES scores were available from more than 200 colleges and universities. Each school's profile (its five scores) was put into a cluster analysis computer program to locate schools having a similar profile. This was done in more than one way and with more than one sample of schools. The result was always the same—eight types of institutions. The first five types listed below were the clearest; the others were less sharply defined.

SLA Highly selective liberal arts colleges
SU Highly selective universities
ES Colleges emphasizing engineering and science
Den Strongly denominational colleges
TC Teachers colleges
GLA General liberal arts colleges
GU General universities
SC State colleges and other universities

For the national baseline we selected 10 schools of each type except GLA and GU which had 20. Overall the baseline was a mid-way compromise between a sample based on enrollments and a sample based of institutions. The result is a reference group similar in concept to the Dow-Jones index which is not a sample of stocks but rather a selection of stocks in important categories.

For each of the five CUES scores we plotted the distribution for the eight types of schools. This showed clearly that the range of scores within a type was much smaller than the total range, and that in many cases the lowest score of an institution in one type was higher than the highest score in another type. So, there are important differences between institutional types, and that information could be used in college admission decisions if it became available.
With respect to CUES validity, comparisons were made with results from other studies. The National Opinion Research Center had undertaken two large studies of the educational and career plans of college students. More than 300 different schools were involved in those surveys. Responses to CUES had already been obtained at some of the schools. We made an arrangement with NORC to get CUES data from many more of their schools and in return NORC would share with us all of their data. This enabled us to make many connections related to CUES validity. We also had data from the National Merit Scholarship Corporation, from all of Astin’s studies, and from CEEB, altogether totalling several hundred correlations comprising a validity network for CUES. The conclusion was that campus atmosphere, as measured by CUES, is a concept buttressed by substantial evidence of concurrent and construct validity.

CUES’ second edition, together with a technical manual that included all the new psychometric data, the national baseline, and the validity studies, was published by ETS in 1969.

Evaluation: New Concepts and New Research

In the mid 1960s the U.S. Office of Education decided to fund the establishment of national research centers focused on important educational topics. At UCLA we submitted a proposal for a Center for the Study of Evaluation. With respect to higher education I said we needed a broader range of outcome or criterion measures, and a greater realization of the importance of environmental or contextual measures. When the Center was funded I tried to incorporate these views in the studies of higher education evaluation which I directed for a period of about seven years.

Historically, in education, evaluation was concerned with the evaluation of instruction and courses. One needed specifically defined objectives and specific measures of their attainment. But what if you wanted to evaluate not just a course or a curricula, but the institution, or higher education in the U.S. where there are multiple and conflicting objectives and programs. The familiar experimental models of research and evaluation do not fit this larger problem. A concept of evaluation appropriate for the study of large and complex institutions must ask different questions, proceed in a different style, and have a new view of the purpose of evaluation and the role of the evaluator. In a report I wrote for the Center I summarized this larger concept as follows:

1. It begins with the central question “What are the consequences?” rather than with the more limiting question “What are the objectives?”
2. Its style of inquiry is more aptly characterized by the word "exploration" than by the words “control” and “focus”.
3. It sees the role of the evaluator as a social scientist rather than as a teacher, missionary, reformer, or staff officer to the practitioners.
4. Its purpose is to provide more complex bases for informed judgment.

The relevance of these thoughts to some of the current (1996) activities and views about evaluation and assessment will be discussed later in this essay.

The major project undertaken in the Center’s evaluation program was, not surprisingly, another questionnaire survey of alumni and current students. In this type of national appraisal there are three potential sources of distortion: first, an inadequate range of criterion variables; second, an inadequate range of contextual or environmental variables; and third, an inadequately representative population of individuals and institutions.

For the population of institutions we selected examples from each of the eight types identified from CUES research. We tried to get 100, as in our national baseline; we succeeded in getting 88. Within each institution, depending on its size, we got random samples of 300 or 150 alumni from the graduation class of June 1950, and similar samples of current freshmen and upperclassmen. The response rate from these groups to the questionnaire was 80 percent from freshmen, 66 percent from upperclassmen, and 58 percent from alumni.

For the criterion variables we had 12 activity scales dealing with a broad range of involvement in contemporary society and culture, a measure of knowledge and awareness about major changes that are taking place in American society and attitudes toward such changes, and ratings of benefit or influence from college toward various educational objectives.

For the environment or contextual variables we had the institutional type from CUES, information about the major field, academic performance, participation in extracurricular activities, aspects of the college experience that stand out in memory, and some corresponding information about high school. Personal characteristics of the respondents are also part of the college context or environment. Beyond the usual census data we included questions about personal and family background, places of residence, personality measures of theoretical orientation, autonomy, and complexity and a vocabulary test.

Most of the results of these questionnaire surveys were reported in my book, The Denise of Diversity? A Comparative Profile of Eight Types of Institutions, The Carnegie Commission on Higher Education, 1974. The evidence reconfirms and strengthens the relationships between college environment and experience, on the one hand, and subsequent activities and benefits. And it shows again that there are real differences between some types of institutions, not only in what they offer and emphasize but also in the activities and attainments of their students and alumni.

Unfortunately the full potential of this national study was not realized because the Office of Education terminated the funding. It was the Carnegie Commission, not the federal government, that enabled me to write the book. However, during the years that funding existed for the higher education evaluation program, there were other achievements. This included more than 20 doctoral dissertations that drew upon the data base, or produced additional kinds of measures for future research. In 1971, owing to the government’s wishes, we developed a loose-leaf
notebook of nearly 40 brief measures, which any college could select for its own particular interests. This included, for example, measures of learning styles, campus morale, teaching, educational preferences, involvement in campus events, societal and educational priorities, etc. More than 1,000 copies of this notebook, called the Higher Education Measurement and Evaluation KIT, were distributed, and several hundred colleges used some of the measures. The KIT also included the various measures from the undergraduate questionnaire—the activity scales, goals, etc.

The decade of the 1970s was a busy one. Clark Kerr asked me to write a book about Protestant colleges. I did, drawing on the recent national survey, CUES data, and some campus visits—Education and Evangelism, McGraw-Hill, 1972. I was issue editor for one of the New Directions in Higher Education reports. This was on Evaluating Learning and Teaching, Jossey-Bass, 1974. Then there was a FIPSE-sponsored national program called Better Information for Student Choice in which UCLA was one of the participants. For this I wrote a little booklet called UCLA: Who Goes? What’s it Like? which was widely distributed to high school students, counselors, parents, and others in 1976. In 1978 I wrote a report for the university on what happened to entering students seven years later. And at the end of 1979 Jossey-Bass published a book I had been working on for a couple of years, Measuring Outcomes of College.

In 1973 a large national project proposal was suddenly and without explanation not funded. This was to be a study of liberal education in large public universities. For a couple of years, the National Association of State Universities and Land Grant Colleges, together with the Council of Colleges of Arts and Sciences, expressed concern about the status of liberal education in those institutions. They appointed a subcommittee to consider what might be done. The consensus was that in the NASULGC institutions the traditional emphasis on technology and applied science had resulted in both philosophical and financial neglect of the liberal arts, that studies of such trends were needed, and that measures of students progress toward liberal education outcomes were needed. Representatives of the associations initiated discussions with me about my possible interest in developing and carrying out such studies. I was very much interested and said I could devote a substantial portion of my time to it over the next several years. The association then formally invited me to develop a proposal and to solicit funds for carrying it out. There were many discussions with people in the associations and also with Earl McGrath who was serving as a consultant to the Lilly Endowment, and who was especially concerned about liberal education. Together with the association we developed a three-year inquiry for submission to the Lilly Endowment. It was our belief and also Earl McGrath’s understanding that the proposal would be presented and recommended for funding at the next meeting of the board of directors. But the proposal was never presented to the board. So that was the end of it. Too bad, because the findings from such a study would be especially valuable today as the content of undergraduate education is being very widely discussed. We know a lot about the effectiveness of selective liberal arts colleges, but we really don’t know as much as we should about liberal education in the NASULGC institutions that grant more than a third of all baccalaureate degrees.

Earlier in this section I said we needed to rethink the role of the evaluator. A few years after the UCLA evaluation center was established, the director at the time defined evaluation as follows: Evaluation is the process of determining the kinds of decisions that have to be made—selecting, collecting, and analyzing information needed in making these decisions—and then reporting this information to the appropriate decision makers. This was obviously not my concept. In my view the emphasis on decision makers has several potential dangers. I personally think the evaluator needs to be independent of the decision maker. And who is the decision maker? Is it a professor, a dean, a college president, a board of trustees, a legislature, a governor? Perhaps some distinction needs to be made between an Institutional Research office which obviously provides staff services to the administration, and an Evaluation office where independence is desirable.

I also suggested that the role of the evaluator was that of a social scientist rather than a teacher or reformer. This conflicts with the view of evaluation and assessment in schools as instruments of change and improvement. However, I neither reject nor minimize the role of evaluation as a learning experience leading to changes; I am only saying that for a very large and complex topic, such as higher education in the U.S., or a state system of higher education, it may be desirable to stand back far enough from the activities to view them in a social science research perspective.

The most provocative concept that emerged in my thinking during the 1970s was to question the merit of the input-environment-output model for evaluation and, even more radically, to question the merit of removing student characteristics (input) when judging the influence of environment on outcomes. Before the student arrives at the college, his SAT score is “input”; but after he is in college his SAT score, along with the SAT score of all the other students, is now a characteristic of the student body and is therefore an environment characteristic. If you want to study the influence of the environment why would you want to remove one of the important features of the environment? There are also other grounds for reconsidering input defined by student characteristics. The nature of experience itself needs to be considered. From psychological research one can entertain the following thoughts:

- experience consists of events
- events have a quality as a whole
- this quality is the resultant of the experiencer and the world, or physical event
- the meaning of an event therefore consists of the context which the experiencer brings to it, and the context of the physical event

\footnote{See James Jenkins' conceptual explanations of human memory in American Psychologist, November, 1974}
the college experience consists of the events that occur in the college environment.

since the experiencer is an integral part of the meaning and quality of an event, the characteristics of the experiencer that are brought to bear on any given event are part of the event itself.

how then can or should the person and the event be separated?

These thoughts form what I have called a contextual model. Instead of input-environment-output, the concept is better expressed by environment-experience-development. Perhaps we should abandon the whole idea of student input. The environment is the input. Its what’s there before the student arrives. Then the question is what does the student do in the environment, and how do different students use the opportunities for learning that the environment provides. This line of thinking was one step along the way toward the concept of “quality of effort”.

The College Student Experiences Questionnaire

Several past experiences and thoughts contributed to my belief that something called quality of effort could be measured and might add substantially to our understanding of student learning and development. For example, education is both a product and a process. We have typically thought of educational processes in terms of what they contribute to the product; but we know that some processes are qualitatively better than others, just as some products are better than others. So perhaps we should give more thought to measuring the quality of the processes. One motivation for my desire to measure student effort was the recurring rhetoric about accountability that always blamed the institution for outcomes. If students don’t graduate it’s the college’s fault. If they don’t learn, it’s the teacher’s fault. If the graduates don’t get good jobs, the college is to blame. If you don’t benefit from going to college it’s their fault. This assumes that the student is buying a product when actually the student, at a later point in time, is the product. So, the other side of accountability is the quality of effort students invest in using the facilities and opportunities the college provides. Several years earlier in a questionnaire given to UCLA upperclassmen the following statement was included: “If students expect to benefit from what the institution has to offer, they have to take the initiative.” Ninety-five percent of them agreed with that statement. They know that they have to make an effort, that it’s up to them to take advantage of the facilities and opportunities that are before them.

The model of a Guttman scale seemed to me to be particularly well suited to the idea of measuring quality of effort. A few years earlier I had developed such a scale on the topic of course learning. It occurred to me that the taxonomy of educational objectives developed by Benjamin Bloom and others could be viewed as a hierarchical classification of learning activities as well as a classification of objectives. I developed a scale of learning activities more or less parallel to the levels in the taxonomy, tried it out in a few courses, found that the professors liked it, and that the quality of cognitive effort showed significant differences between A students and C students. The question, then, was whether or not similar scales could be developed for other important aspects of college experience.

Thanks to a grant from the Spencer Foundation I was able to spend two years creating the College Student Experiences Questionnaire (CSEQ). It may be of some value for current researchers and especially those who develop measures to note the decisions and data that ultimately defined the CSEQ. Too often, in my view, questionnaire makers just include items because they are interesting or in some sense relevant to a topic. Given my background, I think of a questionnaire as a test or measure. What is the variable you are trying to measure and how is it defined?

How do you decide what aspects of college life to look at? And how do you decide on the underlying quality dimension of each aspect? There should be some theoretical or conceptual backing for whatever ways one answers those questions. Many college events are related to facilities—classrooms, libraries, etc. These can be seen as behavior settings, for each facility has a purpose and there are characteristic activities that occur in the facility. Major facilities on all or nearly all campuses include classrooms, laboratories, libraries, residences, student unions, athletic and recreation facilities, cultural facilities such as auditoriums, galleries, theater. Other facilities are important on some but not all campuses, such as chapels; and for still others the concept of quality of student effort does not apply such as the health service or counseling office. Other important events are not associated with a specific facility such as a great variety of interpersonal relationships, contacts with faculty members, involvement in clubs and organizations, informal student conversations, etc. So, some scales would refer to students’ use of major campus facilities, and other scales would refer to important personal and interpersonal associations on the college campus.

Then I decided that the items in each scale should be simple, direct, clear statements of an activity so that students would know immediately whether they had engaged in the activity and about how often. There would be no statements of opinions or of likes and dislikes or of students’ satisfaction or dissatisfaction. Statements of that sort are not indicative of action or effort. I also decided that no statements would refer to events off campus or to facilities which the college itself did not provide since the college is not responsible or accountable for them. Each of these decisions clarifies the definition of effort and rules out the inclusion of unrelated content.

As we started to write statements for the scales we sought advice from many people. We talked to librarians about students’ use of the library, with science faculty members about science laboratories, with English professors about writing, with counselors about student acquaintances and experiences that might contribute to self understanding, with student groups in dormitories at UCLA about opportunities for growth and development in that particular setting. Altogether we developed quality of effort scales related to 18 different topics. In some cases we
tried out different directions for students’ responses—for example, referring to a particular course or to courses in general, or referring to events during the current school year or to events in college up to now. Each scale was printed on a single sheet of paper and we asked friends in various colleges to try out a few of them on a few students. In this way we got from 200 to 500 responses to each scale, from a total of 28 different colleges in all parts of the country.

Now we could examine the statistical properties of each scale. Student answers to each activity were marked by checking never, occasionally, often, or very often. Admittedly this is not very precise, but the direction of frequency is clear. We had written from 11 to 14 items for each scale with the expectation that the final scales would each have about nine or ten items. We scored each scale by giving four points for very often, three points for often, etc. Then we looked at the range of scores for each scale, the item intercorrelations, the correlation between each item and the total score (adjusted), the reliability of each scale, and whether from a factor analysis there was one dominant factor as there should be in a Guttman type of measure. With this data we picked the content for each scale knowing that it was a good measure. The final questionnaire, ready for use, had seven scales related to the use of facilities and seven related to personal and interpersonal events.

The complete printed questionnaire had various items about the students’ background and their status in college, their rating of college characteristics along dimensions somewhat similar to CUES, and their estimate of how much gain or progress they felt they had made toward each of a list of important goals. In the spring of 1979 thirteen colleges used the questionnaire—three research universities, three state colleges, two community colleges, and four liberal arts colleges. Now we are ready to find out what we can learn from students’ responses to this questionnaire.

What has been learned is a lot. Most of it has been reported in monographs, journals, and in research papers delivered at the national conferences of AERA, AIR, ASHE, NASPA, and AAHE. Following the initial use of the CSEQ in 1979, other colleges heard about it and asked how they could get it. So we set up a distribution office at UCLA. In 1983 I published CSEQ second edition, and in 1990 the third edition. When I left UCLA in 1993 I transferred the CSEQ to Indiana University, and a similar instrument for community colleges (the CCSEQ) to The University of Memphis. There have now been more than 500 colleges and universities that have used the CSEQ. The variable, “student effort”, is now a basic element in most research designs for studying student learning and development in college. A few excerpts from my observations along the way probably should be noted.

In my report to the Spencer Foundation in 1979 I ended as follows:

The most striking findings from this study are the discovery that quality of effort is the most important factor in accounting for students attainment, and that after all other influences have been added together, quality of effort still makes a substantial additional contribution. So it turns out that the most influential variable of all is one that has not previously been included in studies of higher education. Now that it can be included, we shall see whether new research confirms the conclusion I would draw from the data reported here, a conclusion somewhat at variance with prior research. Put simply, the conclusion is this: What counts most is not who you are or where you are but what you do.

I believe that this conclusion has been consistently supported by the research studies conducted by other investigators as well as by me in the 1980s and 1990s. The concept or variable I have called quality of effort is operationally defined by the activities in the CSEQ, but there has not been a scientific, psychological-educational, dictionary definition. Effort is not the same as motivation. It is not the same as persistence. It is not exactly the same as initiative. It has elements of motivation, persistence, and initiative, but it also has a specific educational context, and its strength probably depends on the context. My general view is that quality of effort describes voluntary behavior. It reflects initiative. It describes the strength and scope of personal investment that students are making for their own higher education.

Based on the CSEQ responses of 25,427 students from 74 colleges and universities obtained during the years 1983 to 1986, I wrote a monograph, *The Undergraduates, UCLA Center for the Study of Evaluation, 1990.* It is not a technical monograph. There are no means and sigmas, no significance tests, no correlations, no regressions—just percentages and averages and descriptions of what students reported about their activities and progress. I wrote it that way because I hoped that nonspecialists could read it, especially perhaps some of the people who had read the critical reports that had been recently published. One of those reports criticized the curriculum, another one criticized the professors, another criticized the lack of evaluation and measurement, and another criticized the administration. None of the reports dealt directly with the students’ experience in higher education. I think the writers just assumed the students were not learning much. From the CSEQ data I came to very different conclusions:

We found, contrary to the critics who claim that students don’t learn anything, that all students believe they have made at least some progress toward every one of the important goals listed in the questionnaire. What is even more dramatically contrary to the critics is the finding that a majority of students at all types of institutions believe they have made substantial progress toward ten of the most basic and historically most highly regarded goals of higher education—gains in intellectual skills, in breadth of knowledge, in vocational preparation, and in personal and social development. If we are to believe the students, the critics are wrong and badly misinformed.

We also found that the “average” student spends about 35 to 40 hours a week on academic activities. In some places, especially in the selective liberal arts colleges, many students spend 50 hours a week or more.

The level of effort and the scope or breadth of effort students put into their college experience was found to be a very good indicator of the quality of the undergraduate experience, clearly associated with progress toward all important goals, with better grades, and with greater satisfaction.
One of the important findings in this national survey is the truly distinctive character of the selective liberal arts colleges. They are uniquely powerful environments for student learning and development. Nationally, the Carnegie Foundation identifies about 125 selective liberal arts colleges, and I estimate that the total number of undergraduates enrolled in them is roughly 3 percent of the national total. Are they an endangered species?

**VIEWS, VIRTUES, AND VALUES**

For about a century the trend in the shape of higher education in the U.S. has been away from liberal arts colleges to large comprehensive universities, and from private to public institutions. In the last few decades there has also been a trend toward greater enrollment of nontraditional students—older, part-time, nonwhite. And most recently there is exploration of higher education via computers, WWW, Internet, etc. All this is relevant to understanding the results of nearly all of our surveys because nearly all of the students who answer our questionnaires are traditional full-time students on the campus. We know that nontraditional students are less likely to answer questionnaires; and we need to figure out better ways to reach them and perhaps also better questions to ask them.

None of this invalidates the results of our past surveys; it just says that there are other populations and conditions we need to consider in the future. The college campus, with full-time students living away from home, is the hard core of higher education. Nearly all liberal arts colleges are in small towns. And many of the major universities are not located in population centers and do not have part-time commuting students—Princeton, Dartmouth, Cornell, Duke, Georgia, Michigan State, Notre Dame, Indiana, Purdue, Iowa, Louisiana State, Stanford, and many others. So for millions of students going to college means going to a particular place and living there; and the college environment is still an influential variable. Higher education is both an individual and a social experience. This personal/social combination, in behavior settings designed to facilitate learning and development, is a memorable and influential experience. “Virtual experience,” via the best of modern technology, is not a substitute for real experience. And part-time education is not the same as full-time education. Whether any of this matters very much depends on what we think higher education is, or should be.

At the AIR Forum in the spring of 1996 I heard a speaker who thought of higher education as a commodity with its condition and value determined by the market. This reminded me to look back at a speech I gave in Iowa twenty-five years ago. Here are some excerpts:

Many of the analogies and models we use in thinking about education and evaluation are drawn from fields that have no necessary connection with the nature and quality of education. Higher education is not a factory, receiving raw material, processing it, and turning out products having certain performance characteristics. Nor is higher education a business, distributing goods at so much per unit cost. Nor is it a bureaucracy run by bosses with flow charts, communication networks, decision points, and job descriptions. But these mechanical and administrative analogies have their counterparts in the language of educational research and evaluation—the measurement of input-output differences, specified performance objectives, college effects, test score gains, etc. Most recently the popular terminology includes behavioral objectives, product development, cost effectiveness, performance contract, management information systems, and accountability. These conceptualizations tend to emphasize and reinforce an administrative and efficiency view of the nature and purpose of education. A college should be judged by the quality of life that it fosters, the opportunities for experience and exploration it provides, the concern for growth, for enrichment, and for culture that it exemplifies. College could be conceptualized as an environment for exploration—of self, of knowledge and skills, of ideas and values, of society, conscience, community, and commitment.

The variables and concepts that are relevant and useful in other disciplines are not necessarily relevant or helpful in the study of higher education. In any case, it is important to realize what image about the enterprise is guiding what we see and say about it, for all thought begins with an image or at least reflects one.

To a large extent the value of our research can be judged by what we look for and how adequately we measure it. From the General College at Minnesota I saw the value of knowledge about college graduates and the scope of what could be learned from a questionnaire. Later I constructed a set of reliable and valid scales to measure the quality of life of college alumni—responsible citizenship, cultural participation, breadth of engagement in adult society. Next, from my association with the College Entrance Examination Board, and the Social Science Research Council, I sensed the value of knowledge about the college environment. This led to the construction of CUES, enabling the quality of the institution to be measured by the collective perceptions of its students. The characteristics of an institution could then be seen as an educational/psychological stimulus or context for student learning and development. Then, from an interest in the evaluation of teaching and learning, and a concern about institutional accountability, I created a way to measure the quality of effort students invest in their own education. I think each of these lines of inquiry has contributed to the enlargement of research in higher education—the quality of life among alumni, the quality of college environments, and the quality of effort among students.

From my experiences as a researcher, a teacher, and an administrator, I see higher education as a field of study, an area of inquiry, not as an academic discipline or specialization. The best background for a college president is a knowledge of history and philosophy—an understanding of how the college came to be what it now is, and what values are crucial in guiding its future. In the more limited field of measurement and evaluation I think the basic requirements for good evaluation are judgment and creativity, and the basic require-
ments for writing good test items are knowledge of the subject and English composition. Beyond that, various psychometric methods and research designs come into play. The danger in thinking of higher education as a narrow specialty is the likelihood that one's research will focus on limited and easily defined topics, and will be another example of what I once described as unsuccessful efforts to find large significance in small phenomena. In the physical and biological sciences the exploration of ever smaller units of analysis has resulted in new insights—a particular gene, a DNA trace, atoms, quarks, etc. I'm not at all sure that the domain of the social sciences is similar. New insights in the social sciences may come from seeking to discover larger combinations of evidence, and understanding the way events and conditions fit together to create a major influence. If so, progress will come from expansion, not from reduction.

Also, we need to be reminded from time to time that the value of results from a questionnaire survey or a test does not depend primarily on the quantity of responses; it depends on the quality of the questions. The emphasis on quantity as in collecting large national samples usually produces very many very low correlations that are nevertheless statistically significant. In some reports I've read, correlations of .10 are significant and also thought to be important. If we recall the old "coefficient of alienation" we would realize that a correlation as high as .30 still leaves 91 percent of the variance unaccounted for. Often too, large national surveys do not have adequate replies from single colleges; yet if changes are desirable they have to be made at the local level. Because large scale surveys attract attention, the quality of the questions is especially important. Most large scale surveys do not consist of carefully developed measures of fundamental constructs. What are the basic variables being measured? and with what validity? The big surveys "cover" various topics, but not fundamental concepts. All this gets us back to thinking about our purposes and programs.

My general interests and inclinations have been to look for relationships in their natural setting—between environments and attainments, between effort and outcomes, between the pattern of activities as college students and the pattern of activities as college alumni, between institutional purposes and institutional influences. I believe that making connections is a good way to stimulate new research and new insight. But who knows what connections will be important in the future? The size and shape of higher education as we know it today may be very different tomorrow. In any case, I am sure that higher education will continue to be a fascinating topic of research.

Reflections on the Study of Effective College Teaching and Student Ratings: One Continuing Quest and Two Unresolved Issues

Kenneth A. Feldman
State University of New York at Stony Brook

When I was invited to address the Special Interest Group on Faculty Evaluation and Instruction of the American Educational Research Association (Feldman, 1995), I must admit that, initially, I was not altogether sanguine about preparing a presentation. After years of reading and integrating the research on instructional effectiveness in higher education, particularly the research on college students' assessment of their teachers, I felt that most (if not all) of what I had to say or report I had indeed said or reported—in a series of research integrations (Feldman, 1976a, 1976b, 1977, 1978, 1979, 1983, 1984, 1986, 1987, 1988, 1989a, 1989b, 1990a, 1992, 1993; also see Feldman and Newcomb, 1969, Chap. 8) and in some occasional pieces (Feldman, 1973, 1990b, 1994). So I was concerned about merely repeating past observations.

Yet, as I reflected on the work I had done over the years, I decided that there might be some value in highlighting a continuing research quest in the field—that of establishing the conditions and contexts that determine the strength and patterning of pertinent empirical associations. It also seemed worthwhile to elaborate on two long-standing issues in the field that have never been fully resolved: one dealing with the question of bias in college students' ratings of teachers; and the other concerning the applicability of the traditional model of psychological

---

1This chapter is based on an invited address (in conjunction with receiving the Wilbert J. McKeachie Career Achievement Award) that I presented to the Special Interest Group on Faculty Evaluation and Development at the 1995 Annual Meeting of the American Educational Research Association (Feldman, 1995). Raymond Perry, an Associate Editor of this Handbook, has for some time been encouraging me to bring together into one piece some of the different parts of my work on teaching effectiveness and student ratings; I wish to thank him for this support and for his most helpful suggestions on drafts of this chapter. William Cashin and Herbert Marsh—two "official" reviewers of this chapter—made thoughtful, penetrating comments on an earlier version. I also want to thank the following persons for their valuable suggestions: Philip Abramie; Anthony Greenwald; Wilbert McKeachie; Harry Murray; and Harry Tagonomi.