Conversation Hour presentation, APA 1983

MAXIMUM UTILIZATION OF HUMAN RESOURCES

Paul Horst - August 1983

The Undergraduate Years

By the time I entered college I had already decided upon my future career. I had been class poet in high school and had already sold two short stories to Sunday School publications (one for five dollars, the other for ten). I was sure I could write. I was going to be a novelist.

With this as my goal, I enrolled at the University of California, Southern Branch, in Los Angeles, where I planned to major in English. I had been told that a good novelist must know a lot of psychology, so I started immediately to take courses in psychology. When these first courses did not seem to improve my fiction writing ability, I took more psychology classes. Even these did not contribute greatly to my writing ability, but I was accumulating a lot of credits in psychology. When it became necessary to declare my major, psychology was the only subject in which I had enough course credits to complete a major in the time I had left.

In those days many students at U.C. Southern Branch went to Berkeley for their last year. This was some years before UCSB became UCLA and moved from the North Vermont campus to Westwood. Since an AB from Berkeley was thought to carry more status than one from the Southern Branch, I went to Berkeley for my last year.

It so happened that Edward Tolman had been on sabbatical from Berkeley the previous year and had just returned from studying Gestalt psychology with Köhler and Kofka in Germany. He appeared to be very much intrigued by the new school. He tried to explain Gestalt psychology to us but admitted that he didn't understand it very well. Neither did I understand his explanations very well so I tried to read Kofka's GESTALT PSYCHOLOGY and Köhler's MENTALITY OF APES. But I still didn't understand Gestalt psychology very well and I could not see that it had contributed anything to my fiction writing ability.

At this same time there was a young instructor at Berkeley who later achieved distinction at Princeton for his work in audition. His name was Ernest Glenn Weaver and he was offering a course in psychophysics. Perhaps this would inspire and facilitate my fiction writing skills. Of course it didn't, but the idea that sensation increases as the logarithm of the stimulus somehow had more meaning for me than "the whole is not equal to the sum of the parts." I had studied no mathematics in college and the Gestalt psychologists did not point out that the whole might be a weighted sum or a multivariate polynomial function of the parts.

Glenn Weaver's psychophysics course required a laboratory project. Mine had to do with the discrimination of thermal stimuli. I contrived a rather complicated gadget for the purpose. Whatever its scientific and technical merits, the contraption did seem to arouse Dr. Weaver's interest in my own career plans. I told him that I was planning to write fiction. His response was to the effect that, if I was interested in fiction, there was plenty of that in the current state of psychology. But he also assured me that there were no jobs for A.B.'s in psychology and that, if I wanted to get a job as a psychologist, I would have to get a Ph.D. in psychology. I knew I could never be a Gestalt psychologist but perhaps I could be a psychophysicist.

During the time that I was considering the possibility of changing my future career from novelist to psychologist, a graduate student at Berkeley was doing some research work with something called the Stanford-Binet and enlisted me as a subject. She asked me a series of questions about a ball in a field and a lot of other things, and made notes on a pad. This didn't seem like a very precise way of finding out much about me and it vaguely occurred to me that some of the things I had learned in psychophysics might suggest better ways of doing what she seemed to be trying to do.

About this same time a young man on leave from some eastern university was giving a course in which he demonstrated what was called the Stenquist Assembly Test, which struck me as an ingenious way of detecting special abilities in people. Also, at Berkeley one could not avoid hearing such words as "mean", "median", "mode", and "standard deviation." I was told these were important and that, if I wanted to learn why, I should take such and such a course that was to be given in summer school. I did take the course and, with the help of what high school algebra I could recall, it wasn't too difficult.

As the summer drew to a close, I felt I was still far from having the skills necessary for becoming a successful fiction writer. My A.B. degree would be in psychology but I couldn't get a job as a psychologist without a Ph.D. degree. I could never be a Gestalt psychologist but perhaps I could be a kind of psychologist who, by means more precise than the use of ambiguous words, could help people find out what they were good for, if anything.

Where could I find graduate work in this sort of thing? This was before the term "quantitative psychology" was invented. I had a hunch that somehow the word "statistics" would be involved in such a graduate department. Perhaps some psychologist might even have written a book on the subject. Why not look under "statistics" in the card catalogue at the University library? Perhaps one of the books might have been written by a psychologist. Sure enough, I found FUNDAMENTALS OF STATISTICS by L. L. Thurstone, Department of Psychology, University of Chicago. Who else would be at Chicago and what had they done? A University of Chicago catalogue of courses, together with a check in the card catalogue, showed that Arthur Kornhauser and Forest Kingsbury were on the Chicago faculty and had written a book on industrial psychology. Perhaps one could make a living as an industrial psychologist.

Graduate School

So much for vocational guidance circa 1927. I went to Chicago that fall. In those days there was little formality about applying and being accepted. I found the psychology department which was located in an old apartment building on Drexel Avenue. Thurstone's office was on the second floor. He greeted me warmly and I told him I wanted to be a graduate student. So I was.

The first quarter I took Thurstone's course in psychophysics. Whether I was doing so well that he thought I should have some enrichment or so poorly that I needed supplementary nourishment, I don't know. In any case, he suggested that I enroll in Henry Schulz's course in advanced economic statistics, which I did. It turned out to be quite a struggle. I got so discouraged that toward the end of the course I went to Schulz's office for whatever advice he might offer. And this is what he said, "Horst, you're clever enough in spots but you just need to know a lot more mathematics."

This confirmed some of my earlier suspicions. I bought some second-hand trigonometry, solid analytics, and college algebra books. Also a brand new Granville DIFFERENTIAL AND INTEGRAL CALCULUS. That summer on my own I did learn a little mathematics. In the fall, without any formal college mathematics background, I enrolled in a course in Theory of Equations and another in Differential Equations. I did not breeze through either of them but I did learn a little more mathematics. Also that summer, Clark Hull was a visiting professor at Chicago and I took his course in Aptitude Testing. He had not yet got involved in Learning Theory.

For my second year at Chicago, I was awarded a fellowship that paid $500 a year.

At this point I should indicate how I managed to support myself, both in college and in graduate school, for it may throw some light on my impatience with verbal expositions which have persisted in characterizing much of the psychological and other literature. During my high school years, I learned the automobile trimming and upholstering trade from my father. I had already decided to be some kind of a writer. Early on, I discovered several trade journals in the area of my own trade. The readers of these journals were craftsmen who were less than critical of imperfect diction and syntax, and very few professional writers knew anything about automobile trimming and upholstering. It was therefore without too much difficulty that I established a regular contributing connection with several of the journals. My articles were of the "how to do it" type and I would furnish my own illustrations. Working on an average of about two evenings a month, I could produce enough copy to bring in from sixty to one hundred dollars for the month. But I would compose on the typewriter and I got paid either by the word or by the column inch. There was therefore the temptation, when a sentence did not come out quite right, to write another sentence clarifying what I had meant to say. I rarely revised my manuscripts. This was all before the days of word processors. It was not a procedure to improve one's writing skills but it had definite economic spin-offs. My publishers never complained. However, I learned to be conscious of what I was saying and what it meant, if anything. But further, I think, it also taught me to inquire more critically, in reading what other persons have said or written, what, if anything, it meant. This early experience probably led me to the formulation of Horst's last law of communication, viz., "Most things that most people say most of the time don't mean much of anything unless proven otherwise."

But by my second year of graduate school, I had been so far away from my early trade, and styles in automobiles had changed so much, that I was running out of things to write about. So the stipend of $500 a year was most welcome.

Fortunately, in those early days there wasn't nearly as much for graduate students to read as in later years, even though some of it may have been just as dull. Then there was more time to think of research, which

Thurstone encouraged me to do. He had just completed his study on the mental growth curve, which suggested that the absolute zero of intelligence occurred about nine months before birth. His method of curve fitting was before the days of maximum likelihood. I had learned about the method of least squares from Henry Schulz and by this time knew a little about fitting polynomial functions to data. I thought it might be worthwhile to try some of the more sophisticated methods on Thurstone's mental growth data. He was somewhat non-committal but told me that Truman Kelley from Stanford would be in town shortly and was being entertained at the Thurstones' home for dinner. How would I like to join them so that we could discuss my proposal? Naturally, I would like it very much. So I did and we discussed the proposal. Dr. Kelley was even more non-committal than Thurstone about the value and feasibility of my proposal. And so we were back to square one, as is sometimes the case even today.

At this time, Thurstone was very much involved with his attitude measurement and scaling techniques. He remarked repeatedly how sterile the work in psychophysics was and how boring it was to lift weights. But he was excited about the possibility of modifying and adapting the techniques to the psychological scaling of a variety of stimulus elements. He had formulated his Law of Comparative Judgment, with its special cases. The five scale types had not yet been established and declared constitutional but common sense told us that one such scale should be a ratio scale with an absolute origin. Thurstone's scaling models did not provide for this. Perhaps the development of a method for doing so would be worth a Ph.D thesis. Somehow in those days, things just happened sort of naturally without formal proposals and committee meetings and there was a general understanding that my thesis proposal was accepted.

By this time, I had had experience with some of the more verbal courses in the department. I still had no way of deciding how to tell whether what was being said meant much of anything but I began to suspect that a high precision communication system might help. There was one course of a highly verbal nature that did intrigue me. It had something to do with philosophical systems. I read some of the standard philosophers with limited enthusiasm. But Pearson's GRAMMAR OF SCIENCE did keep me awake and I still defend Percy Bridgeman's LOGIC OF MODERN PHYSICS, despite his detractors. How one can use words without underlying operational definitions I, to this day, cannot understand.

I was approaching the end of the first quarter of my second year. I had been struggling to solve for the roots of quartic equations by Ferrari's method and to solve differential equations, without having had formal courses in college mathematics. I had worked all of the previous year and all through the summer. To put it bluntly, I was fed up with school. Just about that time, a call came to the department from Kansas State Teacher's College at Emporia, Kansas, for someone to teach 3 hours a week of introductory psychology and 2 hours of child psychology. The appointment was for the spring semester and summer term. My $500 stipend wasn't going very far and the pay at Emporia seemed very good. I had never taken a course in child psychology but this did not daunt me, since I had not yet had children of my own to give me preconceived notions. Besides, here was a marvelous opportunity to get data for my thesis from the sons and daughters of the Kansas farmers. So I took the job. It was my first teaching experience. So far as I know, my teaching was adequate and, if anyone resented my using these students as guinea pigs, it was never brought to my attention. I used all of my classes to collect a large number of statements about situations which they regarded as varying from very unpleasant to very pleasant. Eventually, these were culled and presented in two formats: (1) I would rather have S_i than S_j, and (2) I would be willing to take S_i if I could have S_j.

As I look back over more than half a century, there are two things that impress me about my thesis. (1) The experimental design did recognize implicitly the fact that Thurstone's scaling procedure yielded a non-basic matrix, even though no one at the time could verbalize this. And (2) we accepted uncritically the hypothesis of a unidimensional continuum for the affective value of stimulus elements, just as even today many assume unidimensionality in item response theory and adaptive testing technology.

During my stay at Chicago, Spearman's THE ABILITIES OF MAN (1927) had just appeared. We began to hear and read quite a bit about Spearman's tetrad equations. Before I left Chicago, Thurstone would remark occasionally, "We really must look into this work of Spearman," but it was not until several years later that he seriously got into the generalization of Spearman's model. No one recognized at the time that this model was simply the sum of a diagonal matrix and the major product moment of a vector, and that it was by virtue of this simple model that theoretically the tetrads should vanish.

But Spearman early on found it necessary to include group factors along with the general factor and Truman Kelley (1928), about the same time, in his CROSSROADS IN THE MIND OF MAN attempted some preliminary algebraic modeling of these group factors. It remained for Thurstone to generalize the expanded model as the sum of a diagonal matrix and the major product moment of a vertical basic matrix.

Another book appeared about the same time as that of Spearman and caused considerable stir among the few graduate students who were struggling for a scientific footing more stable than the slithery verbalizations of the traditional courses. This was STATISTICAL METHODS FOR RESEARCH WORKERS, by R. A. Fisher. The book did not seem to fit into the patterns of contemporary quantitative approaches of other social and biological sciences at Chicago, which at the time were somewhat limited. A few of us therefore arranged for a graduate student in mathematics to tutor us in Fisher during evening sessions. How we located this student or who, if anyone, paid him and how much, I do not recall. We didn't learn very much, however, and soon the tutoring sessions petered out. The success of this enterprise contrasts sharply with the vast influence Fisher's early work has exerted in the social and biological sciences in later years.

My fellowship of $500 for the year was renewed for my third year at Chicago and, with the money I had saved from my teaching at Emporia, I was fairly well set for most of the year. But, having lost two quarters while teaching in Kansas, it became clear that I could not finish my thesis by the end of the year. Fortunately for me, about this time Dr. L. J. O'Rourke, a good friend of Dr. Thurstone and Chief of the Research Division of the U.S. Civil Service Commission, wanted to add a Research Associate to his staff. Thurstone recommended me, with the understanding that I would finish my thesis in absentia and come back for my prelims as soon as possible.

The Civil Service Commission

Thurstone had suggested from time to time that I should be able to get a job at $5000 a year. This seemed an incredible salary to me. And since the country was reeling from the stock market crash in the early fall, $5000 a year was not only incredible, it was impossible. However, I did get a P-3 rating at $3200, which in the fall of 1930, when many persons were already being non-utilized, did seem quite generous.

The Civil Service Commission was in the old Patent Office Building which, I believe, has long since been demolished. A major responsibility of the Research Division was to teach the Examining Division how to make objective examinations. As it turned out, our efforts were at times only moderately well received. G. M. Ruch (1929) had just published his THE OBJECTIVE OR NEW TYPE EXAMINATION, which was probably the most, if not the only, authoritative work on the subject at the time. One of my first assignments was to master this text.

At the same time, the Division was attempting to develop a model test for the Examining Division. The procedure was to have the stems of the proposed items typed on separate cards. There was space beneath for a proposed answer and four proposed distractors. By a series of negotiations, discussions, and disputes, a half dozen of the Research Staff would more or less agree on a correct answer and the distractors. There were no guidelines as to level or range of difficulty of the items but rather general consensus that the item should be neither too easy nor too difficult. This was before the days when it was recognized that the difficulty of a multiple-choice test item is primarily a function of the difference between the degree of plausibility of the most plausible choice and the next most plausible. A major concern at the time was that the incorrect choices should all be as nearly equal in plausibility as possible. However, no objective guidelines were available for achieving this equality.

Most of the testing at that time was of girls for clerical, typing, and stenographic positions. And there was constantly an overriding consideration in the construction of all test items. Is there anything about the items that might lead a girl who failed it to complain to her senator or congressman about the fairness of the item. This sort of judgment implied a great deal of knowledge about each of the girls who might be taking the item and about all of the senators and congressmen who might be called upon to protest. I could never convince myself that we had this expertise.

One of the techniques we used to select items was the determination of its selective value. The selective value of an item, it turns out, is closely related to the item-test covariance, which is also closely related to what item response theorists and adaptive testing technologists call the discrimination parameter of the item.

This concept of selective value, and the emphasis on it, bothered me rather vaguely from time to time. In the first place, I was not clear why each item should have a relationship with all the other items, for it seemed highly probable even then that ability to answer the items correctly should have something to do with what people do in government. I was beginning to develop the vague suspicion that some people can do some things better than they can do other things and that they can do some things better than other people can do them. If this was indeed the case, why should a test item be considered good just because it has a high correlation with all other items or bad because it does not?

I did not quite see how to resolve this issue of the desirable unidimensionality of test items and the obvious multidimensionality of human abilities but it occurred to me that one might start as Spearman did with the correlation matrix of measures of ability. He had postulated one general factor and a lot of specific factors. Kelley (1928) had suggested a sort of piecemeal cut-and-try method of solving for more than one common factor. By this time, I had had considerable experience with the Doolittle method of solving for multiple regression weights. In the process, it became clear to me that one could solve for what I later learned was the Gram-Schmidt method for the triangular factoring of a matrix. It seemed rather neat to me, so I wrote up the procedure and sent it to Truman Kelley for his comments. So far as I knew, he was the ultimate authority on such matters at the time. He promptly wrote back and pointed out gently and kindly that my method did not give a unique solution but that the result depended entirely on the order in which the variables were arranged and that, if there were $\underline{n}$ variables, there would be factorial $\underline{n}$ ways of factoring the matrix. By that time, I knew what a factorial was and it didn't take long to see that, with only six variables, there were 720 different solutions.

For the time being, I had reached a dead end in that direction. But I was now beginning to feel strongly that mathematics had much to offer in keeping our fuzzy verbal solutions and formulations on track and there were other practical problems to be dealt with. I wrote a number of papers on the chance element and difficulty of multiple choice test items and multiple choice test item alternatives. The rationales in these papers were all based on the assumption of unidimensionality but I had a nagging suspicion that somehow it should be possible to generalize to multidimensionality. However, I didn't know how or where to start and, so far as I knew, neither did anyone else.

Another practical problem arose in our studies of observed distributions of test scores. I had accepted the prevailing belief that, other things being equal, one would much prefer to have normal distributions. Accordingly, I wrote several articles showing how one could readily transform any unimodal distribution into one approximating a normal distribution. These exercises early made me question the absoluteness of measurement and scaling systems in general, for it was quite obvious that most quantification systems were arbitrary. I also discovered that current computational systems for calculating multiple regression constants could be much improved and expedited and I wrote several articles showing how.

In general, most of these papers were motivated by practical problems of measurement encountered in the Civil Service Commission. It was not that I had mastered a body of mathematical methods and tools, and then struck out for problems to solve with them, but rather that I was encountering practical problems for which I needed methods of resolution more precise than mere ambiguous verbal formulations. The distinction between these two approaches was brought out more clearly some years later, after I remarked to the University of Chicago mathematical statistician, Walter Bartke, that in the applied social science field we continually encounter practical problems that call for mathematical solutions which seem to tax the resources of those mathematicians to whom we have access. He replied, "Yes, that is true. But we have an easy way around the difficulty. When people come to us with problems we can't solve, we try to sell them problems that we can solve." He of course was too hard on himself and his colleagues but perhaps the point he made was not wholly without merit.

One thing became increasingly clear during my work at the Civil Service Commission. In spite of all we did to improve our psychological measurement techniques, we had very little idea of how useful our instruments were in helping to select and place people so as to maximize the utilization of the human resources in government positions. Actually, we didn't even know much about the human ability requirements of the government departments and bureaus.

There was, in addition to the Civil Service Commission, a Personnel Classification Board, one of whose functions was to formulate standards of personnel requirements for these government agencies. Some of us in the Research Division developed social relationships with some of the people in the Classification Board. But there was an unwritten understanding that these contacts should not extend to official and technical matters that might logically exist between the measurement of human abilities and the performance of human beings in their productive activities. It was tacitly recognized that, to adequately protect the respective fiefs of the two agencies, it was best to maintain a bureaucratic curtain between them.

But it was inevitable that the recognition of the need for some sort of validation of our tests should persist and we cast about to see how at least a small beginning in this direction could be made. Any efforts to collect criterion data would have to elicit the voluntary cooperation of whatever agency might be involved. Presumably, some agency with a strong quantitative and measurement orientation would be a good starting place. And what more logical agency than the Bureau of Standards? After considerable investigation and exploration, we finally located several people in the Bureau of Standards who agreed to discuss a validation project with us. Several of us went to the Bureau and did discuss the project with more or less interested personnel. To make a short story shorter, our efforts much resembled an attack on a mountain of feathers and the project was soon abandoned.

Nobody was to blame for its failure. I merely mention the experience because it was my first of many to learn what colossal administrative, technical, political, and educational problems are involved in the establishment and operation of a continuing validation program in a large social system.

During this time, I was making desultory efforts to finish my thesis and thinking about going back to Chicago to take my prelims. I wasn't making very rapid progress. Then one day Thurstone came to Washington and stopped in to see my boss, Dr. O'Rourke. Shortly thereafter, O'Rourke called me into his office and told me that Thurstone insisted I should come back and take my prelims. He told me that he too would insist that I do so, for my own good and for that of the Research Division. Also, I was to prosecute the completion of my thesis more vigorously.

I did go back and take my prelims and, not long afterwards, returned again to Chicago for my thesis defense. My thesis had been done, of course, under Thurstone but, ironically enough, on the day set for the defense Thurstone had been called to Washington in some consulting capacity so could not attend the meeting. The other members of the department did not pretend to understand the mathematics of the dissertation so we passed the time rather informally and pleasantly in a sort of general discussion. So that was that. The next day at commencement the new, young, and somewhat controversial president, Robert Maynard Hutchins, placed the hood around my neck and, by the authority vested in him by the Board of Regents of the University of Chicago, conferred upon me the degree of Doctor of Philosophy.

Years later at the University of Washington, some of my students would leave for jobs before completing their Ph.D. degrees. I would nag them and prod them until they finally completed their work. When subsequently they wondered, gratefully, why I went to all that trouble, I would tell them about Thurstone's insistence on my behalf, without which I would never have completed my degree work.

One development that came to my attention while at the Civil Service Commission turned out to be very significant. Even at Chicago, it had become evident to me that mathematics could at best be only a precise communication system and that, to make it come alive, you needed numerical data, and numerical data called for computational devices. Already we had become slaves to the hand-driven Marchants and Monroes, for even in those days you could collect or generate data much more rapidly than you could grind out multiple regression weights, especially for some of Henry Schulz's assignments in economic statistics. One never had quite enough computing power. This also proved to be the case at the Civil Service Commission. Then we got rumors of something exciting that was going on over at the Census Bureau where they had, we were told, literally mountains of data. It seems that a machine had been invented by a man named Hollerith that could work miracles with huge batches of data in very little time. We never did see this machine but we did see a card with some round holes punched in it and were told that such cards were the real secret of success of the new machine. We were even given some hope that, before too many years, the Civil Service Commission might have access to such a machine.

I have mentioned some of the papers I wrote during my work at the U.S. Civil Service Commission. The publication of these papers was another matter. At that time, a mathematical equation in an article was a red flag for editors of the limited number of psychological journals in the early thirties. I was convinced by this time that more rigor and precision of expression was needed in the psychological literature. I wrote two articles without a single equation in either. One was entitled "Psychology and the Scientific Method" (1931) and the other, "Measurement Relationship and Correlation" (1932). Both of these were published in the Journal of Philosophy, of all places. I still stand by most of what I said in them. I quote from the second article. "In a verbal analysis the ambiguity of words is at once the refuge of the slovenly and the despair of the precise thinker."

But to find publishers for the other papers was not easy. About the only regular outlet among the psychological journals for articles with mathematical equations was the Journal of Educational Psychology. I published several articles in it and several in the Journal of the American Statistical Association. The new Annals of Mathematical Statistics also accepted several papers. Among the psychological journals that accepted at least one of my papers were the Journal of General Psychology, the Journal of Experimental Psychology, and the Journal of Applied Psychology.

Incidentally, it was quite a switch when I discontinued writing trade journal articles and began writing for professional journals. For the former, I had been earning at the rate of as much as ten dollars an hour, while for the latter I was paid nothing and, at that, it was always touch and go whether I could get a paper published if it had so much as a little simple algebra in it.

I began to consider the possibility of starting or buying a journal that would not only tolerate but welcome mathematically-oriented articles. The most likely prospect, it struck me, might be the Journal of Educational Psychology. I was making the lavish salary of thirty-two hundred dollars a year and, knowing nothing about the cost of publishing or the ownership of publications, it did not seem unreasonable that I could make a deal with the owner of Ed Psych. It was published by a Mr. Bucholz in Baltimore, only several hours drive from Washington. In those days we worked Saturday mornings, but on a Saturday afternoon I drove up to Baltimore to find Mr. Bucholz. I located him in a sort of shed in the back of his home where, in an inkstained apron, he was apparently working on getting out an issue of the journal. I told him of my interest and asked whether he might consider selling the journal. He answered with an unequivocal "No" and turned back to his work, just as though I were no longer there--immediately after which I was not.

Some years later, I got to know Jack Dunlap who was then editor of Ed Psych. He told me about a story related to him by Bucholz, the publisher. It seemed that some brash young squirt had wandered into his shop one time, pretending that he wanted to buy the journal. Naturally, it didn't take him long to get rid of the chap.

The Procter and Gamble Years

Sometime afterward, in the spring of '32, I received a letter from Marion W. Richardson, later of Kuder-Richardson fame. We had been classmates at Chicago and he was now in the Personnel Research Department of the Procter and Gamble Company in Cincinnati. He wanted to return to Chicago to finish his Ph.D. work and serve on the newly created Board of Examinations under Thurstone's auspices. He was recommending me for his job as Supervisor of Selection Research. Would I be interested in the job at a salary of about four thousand dollars a year?

I wasn't quite sure how to respond or what to tell my boss. However, very shortly Dr. O'Rourke decided the issue himself. This was the summer of 1932 and the great depression was well under way. One morning a staff meeting in the Chief's office was announced. We filed in and he received us with ominous gravity. He told us that very hard times had fallen upon the government and that all appropriations were being greatly reduced. He wasn't sure just how drastically the Civil Service Commission and the Research Division's budgets would be cut but there was no doubt the reduction would be severe. He urged one and all that, if we had any opportunity for employment elsewhere, we should by all means accept it. So I went to Procter and Gamble.

Marion Richardson had been in the process of developing and validating tests for the selection of salesmen for the Procter and Gamble Distributing Company. Frederic Kuder of Kuder-Richardson fame had been assistant to Richardson while this work was going on, but was a casualty of the Company's depression-inspired 10% reduction in force. He returned to Ohio State to complete his Ph.D. and develop his highly successful Kuder Preference Record. Later, of course, he launched <u>Educational and Psychological Measurement</u>.

I inherited Richardson's test development project which was already quite well along. An interesting feature of the project was the emphasis on criterion development as a basis for test validation. Richardson had been exposed to Thurstone's work on attitude measurement at Chicago. It occurred to him that one might develop a proficiency evaluation model by adapting Thurstone's technique to the measurement of supervisors' attitudes toward their subordinates. He collected from supervisors a large number of statements that they might make about good, indifferent, and poor salesmen. These he culled and had scaled by one of the currently standard methods. A rating form was then made up from these and supervisors would check the appropriate items for a given salesman. The score then was the average scale value of the items checked.

Now during my first year or so at Procter and Gamble, Richardson was given a consulting connection with the Company to monitor my work. He would come down from Chicago every few months, at the rate of twenty-five dollars a day. (This was in the depths of the depression.) One of these times he suggested that perhaps it should not be necessary to scale the descriptive rating statements before administering the form. Why not start out by assuming that (1) the average scale value of a statement should be the average of the scale values of the persons for whom the items were checked, and (2) the scale value of a person should be the average of the scale values of the items checked for him. He did not know quite how one would start or even proceed. I didn't either. But after thinking about it a while, some time later I worked out an iterative method with punch cards and large sheets of cross-ruled paper. I soon discovered that the rationale required further constraints on scale and origin for either the person or statement dimensions. Much later, after I had learned something about matrices and characteristic roots and vectors, I found that the model could be naturally extended from a unidimensional to a multidimensional system. Of course, this was long before the days of electronic computers and the optimal multidimensional solution was not feasible with the available computational equipment.

It was in the early thirties that Thurstone became seriously involved in the development of his factor analytic models. Hotelling (1933), about the same time, published his principal components method in the Journal of Educational Psychology. Thurstone (1933), of course, had provided his centroid method which, if the number of variables was not too large, was computationally feasible. But Hotelling's early analysis based on solving for the characteristic roots of a determinant was not. I derived a method for solving for these roots, using Newton's identities and a theorem about the elemental powers of determinants. I still think it was a pretty clever solution. It had only three flaws. (1) It involved determinants rather than matrices, and probably anything that can be done with determinants can be done better with matrices. (2) The method was computationally infeasible because the dominant root obliterated most of the others. And (3), as I learned about a quarter of a century later, the method was not original but had been discovered by Leverrier (1840) almost a century earlier.

However, the editors of the Annals of Mathematical Statistics apparently either did not know or did not care that the method was not novel, for they published it. But not before some preliminary difficulties at Procter and Gamble. It was understood that, before an article was published by a P & G employee, it must be cleared with higher authority. So I would ask my boss who would ask his boss. His boss would ask, "Is there anything in it that anyone in Lever Bros. or Colgate could understand?" If the answer was "No", it was O.K. to publish. But this policy had posed some problems. If no one at Lever or Colgate could understand the article, perhaps very few of the readers of psychological journals could understand it either, and the editors would be reluctant to publish such an article. What was needed was a psychologically-oriented journal that was not afraid to publish mathematically-oriented articles. Such a journal, together with a society to publish it, was established in the middle thirties. Much of the story of how this came about was chronicled by Jack Dunlap (1942) in his Psychometric Society presidential address, and in his address (1961) to the 25th Anniversary meeting of the Psychometric Society. I might add a few notes not included in his reports.

During the early stages of the development of the journal and the society, I happened to know a member of the Procter and Gamble advertising department whose younger sister was studying art at the University of Cincinnati at the time. I drew up a brief prospectus of my own ideas for the journal and offered a prize of $30 to the budding young artists for the best cover design for the proposed journal. I got back a number of varied and imaginative responses which I found very useful in stimulating support for the project. I don't recall who won the prize or just how the winner was chosen. But it isn't particularly relevant, since none of these designs was ultimately adopted.

By this time, several of us had elicited Thurstone's enthusiastic support of the project. Knowing that his clout in getting the journal under way was greater than that of all the rest of us put together, we were more than willing to give him a free hand in anything he was willing to do. One thing he offered to do was to design the cover for the journal. This he did, and his design has persisted to this day, in spite of some interest in recent years in changing it. In the early years we did receive a little good-natured ribbing because of the first line immediately below the heading "Psychometrika." This line has been quoted as reading "a journal devoted to the devil."

We were fortunate in the beginning to find a good and inexpensive printer. He had been just recently discovered by the newly-founded journal Econometrica. His company was the Dentan Printing Company in Colorado Springs which contracted to print our journal for $3.25 per page, including tables, equations, and figures. Those days are, of course, long gone.

One very strong supporter and advisor in our early efforts to get the society and the journal going has never, I think, been given the credit due him. This was Donald Patterson who at the time was secretary of the American Psychological Association. If it were not for his expertise in the organization and operation of social agencies and his interest and assistance in our early efforts, I am not at all sure that Psychometrika and the Psychometric Society would have been launched.

Getting papers together for the first issue of Psychometrika took some doing. Thurstone was, of course, the first president of the Society and he had prepared and given his presidential address. It would seem logical that this address should be published in the first issue of Psychometrika, and he was more than willing that it should be. Some of us felt rather strongly, however, that it would be lost in the first issue of an obscure new journal and that it would be best for the prestige and promotion of the fledgling enterprise if it were published in a widely read and respected journal like Science, the official publication of the American Association for the Advancement of Science. Thurstone did accede to our wishes but I think quite reluctantly. I have often wished since that we had published his address in the first issue of Psychometrika. It would have pleased him and been a great tribute to the first, if not all subsequent, issues of the journal.

At Procter and Gamble we had been working with the applications of the new factor analytic methods to personnel data. I had learned of a new iterative procedure that Hotelling at Columbia had developed for finding the principal axis factors of a correlation matrix, and we were using it at P&G.

About this time, during the fall and winter of 1935, several major events occurred. The first of these was my marriage in September to a childhood acquaintance whom I had only recently rediscovered after the lapse of many years. The second was the purchase by Procter and Gamble of a British subsidiary in Newcastle, England. The company decided that a colleague and I should go to England to develop selection tests after the pattern we had recently followed in the States. My chief, Mr. R. F. Lovett, had most generously arranged that the project could be, for me, a combined business trip and delayed honeymoon, with our departure immediately after the first of the year. I was very eager that the first issue of Psychometrika, due for March publication, get off to a good start. It would be a fine thing if we could get a manuscript from Hotelling on his new principal axis technique. My new bride and I would have a layover of a day or so in New York before we sailed for England. Perhaps I could see Hotelling personally at Columbia during this time and persuade him to contribute his manuscript for the maiden issue of Psychometrika. I did meet him in his office and asked him whether he could give us a manuscript on his new method. He at first was markedly cool to the idea and I suspected that he was not eager to conceal his production under the cover of a dubious new journal. I then told him that I very much wanted this method published in this first issue and that, if he did not feel he could do it, I would reluctantly publish the method myself and of course give him full credit. With this, he decided to provide the manuscript himself (1936), and we remained good friends as long as he lived.

From March 1940 through June of 1942, I assumed the managing editorship of Psychometrika. My wife and I conducted the operation in a basement room of our Cincinnati home, with the aid of an old second-hand ditto machine. Of course, the nitty-gritty, real work was done in those early years by the able assistant managing editor, Dorothy Adkins.

During my tenure as managing editor, I don't recall ever rejecting a manuscript. Sometimes, however, my suggestions for revisions were so extensive that I could be reasonably sure the manuscript would not be resubmitted. But one manuscript I am proud to have published was by George Ferguson (1941). In it he showed that a matrix of perfectly reliable and homogeneous binary item scores could not be of rank one unless the items were all of equal difficulty. The rank of such a matrix was obviously equal to the number of distinct difficulty values for the set of items.

It remained for Carroll (1945, 1961) to generalize this phenomenon to n-ary measures. This has come to be known as the disparate distribution problem. It is only in recent years that the importance of this phenomenon for psychological measurement is beginning to be recognized. Even yet, its importance is becoming generally recognized only by unidimensional binary models. Some slight recognition of its importance is reflected in recent work with multidimensional binary models but little, if any, work is being done in the very important area of multidimensional n-ary models.

My work at Procter and Gamble was most challenging and rewarding. I became more and more convinced that one of the most important objectives of a social structure was the maximum utilization of the human resources with in the structure. Along with this conviction grew the realization that human beings and their effective interactions with one another and their environments involved vastly more complex mechanisms than those of the physical sciences. If those sciences required the use of efficient and precise symbolic systems, how much greater was the need of the behavioral sciences for more efficient symbolic systems. To me, the ever-growing and proliferating jungle of verbal and semantic underbrush being propagated by the behavioral disciplines seemed to be getting nowhere. The soap-selling problems of the real world in real world settings suggested the need for more than armchair approaches, however eloquent and persuasive. For example, in our search for better methods of evaluating employee proficiency, we found that, for a specified group of employees and a specified group of evaluators, not all evaluators could evaluate all employees. Consequently, a matrix of ratings of the employees by the raters would have missing elements. This was a special case of the missing data matrix and led us to seek an optimal mathematical solution. We could not find much help from statisticians and mathematicians. In later years, we identified many special cases of the missing data matrix. In one such special case, we were told by an eminent statistician that this problem had no solution and he tried to sell us a problem that did. Unfortunately, we did not have the problem he tried to sell us.

We did work out a practical solution for the incomplete rating problem. As time has passed, I have come to the conclusion that all multivariate analysis models should assume incomplete data matrices, of which the complete data matrix is but a special case. In fact, one of my tests of a model for handling an incomplete data matrix is to see whether it makes sense in the limiting case of a complete data matrix.

Speaking of matrices, the more I worked with data at Procter and Gamble the more I became convinced that the scalar mathematics I and most mathematicians were using was not very efficient, even with the aid of determinants, about which I had learned something from Dickson's (1922) THEORY OF EQUATIONS and Bôcher's (1927) LINEAR ALGEBRA. In my earlier years at P&G, I began to feel that we needed someone in the department who knew more mathematics than I. We got a young man, Carl Frantz, from Yale who had satisfied all of his requirements for the Ph.D. in mathematics except a thesis. He knew more about matrix algebra than I but did not pretend to be an authority. We found a book by Wedderburn (1934) on matrix theory and decided to hold some evening study sessions. I do not recommend the book for beginners. But with the help of Frantz, I did learn some matrix algebra which I found extremely useful in working with the multivariate analysis models that emerged from our work with personnel data.

Soon after this, the third issue of Psychometrika came out with an exciting article by Eckart and Young (1936), entitled "The Approximation of One Matrix by Another of Lower Rank." I shamelessly plagiarized the concepts of this paper, lock, stock, and barrel, and from it developed the obvious concepts of the basic structure and general inverse of a real matrix. Since that time, as Hamilton and Caley are my witnesses, I have not knowingly stooped to the use of determinants in public. From that time to the time I wrote my MATRIX ALGEBRA FOR SOCIAL SCIENTISTS, I have tried to shuck away from the traditional treatments of matrix algebra those vestigial notational, semantic, and conceptual excrescences of pre-matrian times that I have not found useful in the analysis of real world data.

"The Prediction of Personal Adjustment"

By the winter of '41, World War II was in full fury in Europe and there was much concern in this country about National Defense. The Social Science Research Council had a Committee on the Prediction of Personal Adjustment, of which Samuel Stouffer was chairman. This committee believed that a monograph setting forth the current state of the art in this area could contribute to national defense efforts. Stouffer, who was then in the Sociology Department at Chicago, asked Thurstone to recommend someone to come to Chicago for several months and head up the writing of a monograph on the subject. Thurstone recommended me for the assignment. P&G agreed to donate my time for this project and out of this came the Social Science Research Council Monograph (1941), "Prediction of Personal Adjustment." The monograph was a collaborative enterprise. An important contributor was one of Stouffer's young graduate students who later achieved fame by his many brilliant contributions to quantitative methodology in the behavioral sciences. His name is Louis Guttman.

I was responsible for writing the body of the text of the monograph and also contributed a series of five technical appendices which grew out of my work at Procter and Gamble. Briefly, these are as follows:

I. "A Multiple Rating Problem and its Mathematical Solution." This was a sort of extension and culmination of my work with forced choice techniques. I have been credited, or accused as the case may be, by Robert Wherry and others, of having originated the idea for forced-choice item formats. I did give a talk on this technique in the early thirties at a meeting at Ohio State which Bob Wherry attended. However, the relationship to Thurstone's paired-comparisons technique is obvious.

II. "The Problem of the Matrix of Incomplete Data." This is a generalization of the case of multiple raters where not all raters rate all subjects.

III. "The Role of Prediction Variables Which Are Independent of the Criterion." This is a preliminary approach to the suppressor variable model.

IV. "Approximating a Multiple Correlation System by One of Lower Rank as a Basis for Deriving More Stable Prediction Weights." This model arose as a result of the fact that at P&G, like many other places, the number of observations compared to the number of variables was too small to give stable multiple regression weights. In effect, we developed a reduced rank regression solution which increased the number of degrees of freedom of the model for a given data set.

V. "An Analytical Formulation of the Multiple Cutting Score Technique." This model was an attempt to replace ad hoc multiple cutting score techniques using multiple definite integrals.

In the body of the text, we repeatedly emphasized the need for more powerful computing equipment. Electronic computers had not yet been developed. Even today, however, it appears that no matter how powerful the computing hardware, we continue to develop models and data generation procedures that require still more powerful computational hardware.

Another point we emphasized in the monograph has had relevancy in the computer age. We pointed out that job analyses and instructional manuals should not be produced by experts in the field of activity but rather by persons sufficiently in-expert to recognize problem areas in the activity complex. This principle has been consistently and persistently violated in the preparation of documentation and user manuals for computers since the early days of computer programming. A perusal of such manuals by novices shows that, rather uniformly, computer instructional manuals are written by experts for experts who do not need them.

In the U.S. Army Air Forces

The fall after my return to Procter and Gamble from Chicago came Pearl Harbor and our country was soon in the midst of World War II. Considerable pressure soon began to be exerted on psychologists with experience in personnel work to join the war effort. Finally, in the summer of '42 I was commissioned in the Psychological Division of the Air Surgeon's Office. This work was headed up by John Flanagan.

From the point of view of the technology of human resource utilization, a number of interesting events transpired with our work in aviation psychology.

One of the first direct manifestations of the suppressor variable which came to my attention was a multiple regression study we did with three pencil-and-paper tests against a criterion of pass-fail in primary pilot training. The three tests were verbal, spatial, and mechanical. The verbal test came out with a negative regression weight. Obviously, the verbal test was suppressing non-valid verbal variance in the mechanical and spatial tests. As it turned out, this result was not utilized in actual selection. We concluded that it would be too difficult to explain to the generals why we would penalize an applicant for having a high score in a selection test and reward another for having a low score. I am still not sure that suppressor variables are sufficiently exploited in personnel selection and classification programs.

A standard procedure in the aviation psychology program was to administer several experimental tests along with the operational classification tests to batches of incoming cadets. These cadets were then followed up some months later, after pass-fail criterion data became available. We could then estimate by well known methods whether the experimental tests could increase the predicted efficiency of the battery.

It occurred to me that it would be useful to know at the time the experimental and operational tests were administered how probable it was that the experimental tests would increase the predictive efficiency of the battery. Knowing the validity of the operational battery and the correlation of the experimental test with the battery, it was possible to estimate what the validity of the experimental test would have to be in order to increase the validity of the battery by a specified amount. If this required validity was obviously higher than could reasonably be expected, then we could forget it and go on to design and/or introduce other experimental tests. I do not recall of any test that we ever tried out which had a much higher validity than we thought it might.

One of the problems we encountered early in the program was that of sample selection and its direct and indirect effects on correlation. Solutions to the various forms and special cases of such problems are now well known, but in those days they were not. I did succeed in working out what I regarded as a rather elegant set of solutions in supermatrix notation and showed it to Walter Deemer who was head of our statistical unit at the Headquarters of the Flying Training Command in Fort Worth. Several days later, he showed me he had located another solution of the same problem in Philosophical Transactions. Karl Pearson (1903) had solved the problem about forty years earlier.

However, Pearson's solution covered a number of pages and was done with clumsy determinants and mine was neatly done in several lines of matrix equations, so I considered myself still at least one up on Pearson. But my joy in achievement lasted only several months. I happened to run across a paper in the Proceedings of the Edinburgh Mathematical Society by A. C. Aitken (1934). He gave a solution in matrix notation identical to mine and it had been published almost a decade earlier.

One advantage that the work in the Air Force had over that in Procter and Gamble was the large numbers of cases we had to work with. It also had the advantage of follow-up criterion data which we did not have in the Civil Service Commission. However, not all of our criterion data were of the highest quality. We were working primarily with pilots, bombardiers, and navigators. For pilots, we worked for the most part with wash-out or pass-fail criteria in primary training schools. For navigators, we could get school grades. For the former, we could get multiple R's around the low .50's and for navigators in the low .60's.

But for bombardiers, we were not so successful. The standard criterion measure was average circular error. We were lucky to get multiple R's in the low .30's. The measure was obtained from photographs of bomb drops on ground targets in training missions. We decided to make a closer study of the bombardier criterion by going to some of the bombardier schools and actually going on some of the training missions. One observation we made during these missions was that the pilots not infrequently were kindly and considerate to the student bombardiers. If, on a bomb run, the pilot saw that the released bomb was going to miss the target badly, he would sometimes not click the camera but would make another bomb run, with the hope that the student would do better. What such consideration for students did to our criterion is perhaps obvious.

But this departure from objective evaluation was not all. At the end of the day, the reels of film would be read and the circular error recorded by hand for each student. This was a tedious business, with miles of film, and it was reported that frequently the film readers, after several hours, would become tired and hungry and, tossing the remaining unread films in the bin, would say, "To heck with it. Let's put down some numbers and call it a day."

This example is in no sense a criticism of the bombardier schools in the U.S. Army Air Forces during World War II. It is but another example of the great administrative, operational, and political problems involved in the acquisition of adequate criterion data in psychological measurement and prediction programs within large and complex social systems. These problems persist even today in the three great testing organizations of our time, viz., the Educational Testing Service, the U.S. Office of Personnel Management, and the Department of Defense. There is no want of activity, expertise, and technical achievement which these three organizations exhibit in the development of psychological measurement models, instruments, and technology. But all three are seriously handicapped for want of adequate validation opportunity. The sporadic and piecemeal validation studies that these organizations are able to conduct are far from adequate to counter the onslaught of political and special interest groups that have plagued psychological measurement programs for years, but with increasing violence in recent times.

Assuredly, there is no shortage of expertise and off-the-shelf models and techniques to provide adequate comprehensive proficiency evaluation programs within our major social institutions, whether they be industrial, educational, governmental, or military. Such programs could greatly augment the utilization of the human resources within the organization, as well as provide a sound basis for the validation of psychological measurement and prediction instruments.

The problem is not one for scientists and technologists, however, but rather for the power structure of our society.

During World War II in the Flying Training Command of the U.S. Army Air Forces, our problem was much simpler. We had only three job categories to deal with during most of the war: pilots, bombardiers, and navigators. Our criterion measures were reasonably adequate and continuously available. We could make up bar charts to show simply the effectiveness of the classification tests, which the generals could understand.

It should be remembered that the aviation psychologists functioned under the Air Surgeon at this time. In the classification of air crew candidates, the Flight Surgeons had the last word. We contrived a little experiment to test the relative contributions of the surgeons and psychologists in predicting success in pilot training. The procedure was as follows:

A specified group of 1000 candidates for pilot training would be tested on the classification tests, and the flight surgeons would indicate their evaluation as to whether each one would be successful in primary pilot training. However, all of the candidates were to be admitted to pilot training, irrespective of their classification test scores and the flight surgeons' evaluation. The surgeons had access to the classification scores of the psychological tests before making their evaluations. John Flanagan succeeded in selling the proposal to the cognizant authorities. Just how he did it, I never knew.

The validation follow-up of this group was most interesting. As I recall, the correlation of the tests with the criterion was around .60. This was higher than the low .50's we usually got because there was no restriction of range on the tests. The correlation of the flight surgeons' categorical evaluations was somewhere in the low .30's. A test of how much the addition of the categorical evaluation to the test score would increase the multiple correlation was made. As I recall, the increase in predictive efficiency to three decimal places was .000. The results of this little study were never given great currency during the War, nor since, so far as I know.

One problem that began to nag me during the war years was that of the optimal classification of air crew personnel. Gradually, I realized that, theoretically, if you assigned each candidate to the air crew category for which his prediction was highest, you might not just happen to come out with the right number of bodies to satisfy operationally- or administratively-determined quotas for the three air crew categories.

This, however, was not a very serious practical problem. In general, those with the highest prediction for pilot were assigned to pilot training until pilot quotas were met. Of the remaining, the same procedure was used for navigators. Bombardier trainees were taken from what was left.

Logically, this sequence seemed a little odd. One could argue that the most important objective in a bombing mission was for a bombardier to hit the target. The next most important was for the navigator to get you to the target. And last, but of course not least, that the pilot should be able to fly the plane. Later, of course, I learned that optimal utilization of human resources wasn't all that simple.

Back to Procter and Gamble

At the end of the War, I returned to Procter and Gamble. In my earlier capacity as Supervisor of Selection Research, I had been given a reasonably free hand in my research. It seemed pretty much that what was good for my research was good for P&G and what was good for P&G was good for my research, just like in General Motors. I had said repeatedly, before the War, that I would never wish to have my boss's job, which was that of Manager of the Personnel Research Department. I had been back only a few days when I was informed that my boss was resigning from the Company to take another position and that I had been promoted to his job. In those days, one did not summarily turn down a promotion at P&G. For two years I struggled with problems of personnel and budgets, and with interfacing my staff department with the hard-hitting line organization of the Distributing Company.

During this time, I told my wife that, if I could get a professorship at a much lower salary at Boulder, Seattle, or the Bay Area, I would take it. In the summer of '47, I got a letter from the Department of Psychology at the University of Washington, asking whether I could recommend someone for a professorship in Industrial Psychology. I responded that I did, and that I was it. So I went to Seattle.

Toward the end of my second year at the University of Washington, I was offered the position of Director of Research at the recently created Educational Testing Service in Princeton. I accepted it on the basis of a one-year leave of absence from the University of Washington.

The Educational Testing Service

Ledyard Tucker, Frederic Lord, Harold Gulliksen, and Bill Mollenkopf were all at ETS at the time, together with some very bright young Psychometric Fellows by the names of Bert Green and Warren Torgerson. John Tükey and Sam Wilks were at Princeton University and worked closely with the

ETS group. It was a most stimulating environment. I started my work on optimal test length. Also, I was president-elect of Division 5 of APA that year and I was working on my presidental address, "Most Men Are Created Unequal."

One of the tenets of the paper was that a person's rewards from society should be a monotonic function of his contributions. By the time the paper was finished, it pretty well stated my social philosophy. Eventually, I submitted the paper to the American Psychologist for publication. It was rejected because the editor did not consider it suitable for publication in that journal. It was later accepted and published in Science. But it then seemed that one editorial writer, for the Boston Globe, didn't think it suitable for appearance in any publication. He very much deplored these psychologists who presumed to play God. However, I continued to believe that efforts to further the maximum utilization of creation's crowning achievement could be justified.

One problem I sensed soon after coming to ETS was the same as one we had had at the Civil Service Commission. There was no ongoing comprehensive and continuing follow-up procedure for accumulating criterion data and maintaining a continuous validation program. It is true that fragmentary and sporadic arrangements could be made with a few colleges and universities for limited and occasional validation studies for the various testing programs. In addition to the SAT verbal and math tests, there were, of course, a number of other specialized testing programs, such as the Graduate Record Exam, as well as law, business, teacher, Coast Guard, and other testing programs.

Toward the end of the year, the Psychology Department at Washington needed to know my plans for the following year. By this time I had developed a very strong conviction that any testing program that was seriously concerned with the optimal utilization of human resources must be closely articulated with a multidimensional validation program. I recognized only too well the herculean administrative, operational, and control problems in test validation for organizations like ETS and the Civil Service Commission in a relatively unstructured and fluid society.

Back to the University of Washington

However, it occurred to me that, within a state the size of Washington, it might be possible to develop a multiple differential prediction program as a cooperative enterprise between the high schools and the institutions of higher learning in the state. I outlined my tentative ideas to the appropriate administrators at UW and was offered, in addition to my professorship, the Executive Directorship of the Division of Counseling and Testing Services. I accepted this arrangement, which involved also a Research Program.

At the same time I was still actively involved with the Psychometric Society and Psychometrika as a member of the Editorial Council and the Psychometric Monograph Committee.

During these early years, something revolutionary had been developing in the American Psychological Association that had definite implications for the Psychometric Society. The Divisional Structure was inaugurated and one of the divisions was No. 5, Evaluation and Measurement. The inclusion of this division was, I suspect, a belated recognition that the interests of an emerging group of psychologists had not been adequately represented by the old guard. Back in the middle '30's, the very efficient and popular secretary of APA, Donald Patterson, had encouraged the organization of the Psychometric Society to represent the interests of this inadequately recognized group.

For some years after World War II, Division 5 and the Psychometric Society held joint sessions at the annual meetings of APA. In the early years after the War, I began to have second thoughts about my earlier role in the organization of the Psychometric Society. Was the separation of the Society from APA draining off the kinds of rigorous quantitative approaches to scientific development that all branches of psychology needed, particularly the applied areas?

Some wags were commenting, somewhat unfairly, that the most important discovery of clinical psychology during and after the War was money. In my own activities at the University, I was working closely with the clinically-oriented group. Some of my own students were also closely associated with the clinical interests in the department. I would be less than candid if I did not admit that I thought this close association of quantitative and clinical interests was important for both. Surely, Division 12 and Divison 5 could profit from a higher degree of association than presently characterizes these relationships. It should not be surprising to learn that I believe all areas of psychology could profit considerably from the use of a more rigorous and precise symbolic system than is provided by ambiguous and redundant verbal systems.

It used to be said by some that, if you can't put it onto an IBM card, it's literature. I have at last decided to come out of the closet and admit that, for me, if you and I can't explain it to a computer, we should not try to explain it to one another.

This is all by way of saying that the influence of Division 5 in the APA is far less than it should be, and perhaps the Psychometric Society has been a drain on its vitality. I suggested some years ago, at a meeting of the Psychometric Corporation, that perhaps the Society had served its purpose and could now have its assets and functions absorbed by Division 5. It was at once pointed out that many of the very scholarly and productive members of the Society were from other disciplines and not even eligible for membership in the APA. Their expertise was sorely needed by psychology. I had no good solution for this problem. But in the ensuing years, I have sensed that the relative influence of Division 5 on the APA has continued to decline.

In the meantime, other quantitative behavioral societies and journals have sprung up. Notable among these is the Classification Society. This society is taking up seriously the most fundamental problems of taxonomic research, which the factor analytic models have only begun to scratch. Hopefully, it can ultimately help to protect us against the almost certain collapse of the rhetorical towers of Babylon that social scientists have been so assiduously constructing over the years.

We obviously can't stuff the genie back into the bottle. Nor at this stage of its growth and expansion should we want to. But perhaps Division 5 can fashion a conduit through which other divisions of APA can draw much needed methodological and scientific nourishment from these other organizations.

At the University of Washington, one of my major activities was the development and implementation of the Washington State differential grade prediction program. This proved to be a rather long and arduous enterprise. In the first place, there were many technical and methodological problems because we were dealing in a real world with real-world problems. High school students came from many different schools throughout the state to the various institutions of higher learning within the state. The problem was how to set up a multiple prediction model with a very complicated incomplete data matrix so as to use optimally all available information. I must confess that we knowingly cut some statistical Gordian knots, the cutting of which horrified one or more mathematical statisticians.

Some years later, Ledyard Tucker (1963) published his <u>Psychometric</u>

Monograph, entitled "Formal Models for a Central Prediction System." This he had formulated as a basis for the kind of continuous validation program which would have been highly desirable for ETS to develop. It is probable that the budgetary, administrative, operational, technical, and public relations problems incident to the development of such a program were insurmountable at ETS. In any case, it is interesting to speculate how much of the public relations and political hassles of the past few years could have been avoided if the program could have been successfully implemented and its results widely disseminated.

In Washington, we could probably have utilized a model similar to one of Tucker's models, but that it would have been cost effective is not at all clear.

The technical and operational problems encountered in the development and launching of the Washington State differential grade prediction program were not greater than some of the public relations problems that emerged. The major institutions of higher education in the state each had its own Board of Regents and each struggled for its fair share of college-bound high school students in the state, as well as for its fair share of legislative appropriations. The University of Washington was by far the largest of these institutions and the smaller institutions would brook no actions from it which might lead to a hint of subservience to the giant octopus on Puget Sound. Considerable time, effort, and patience was required to develop a program which would be viewed as entirely cooperative among them all. In addition, the high schools insisted on complete confidentiality of test results, so that one school could not be compared with another and possibly be shown to be inferior.

Two entirely different kinds of reactions were noted from high school counselors throughout the state. One was a fear that the new guidance tool would be too complicated and would make their jobs more difficult. The other was that these tools might be so effective as to make their jobs unnecessary and thus eliminate them. A series of conferences and orientation sessions were conducted throughout the state which eventually succeeded in allaying the anxieties and objections of at least the most articulate of the counselors.

But there were also objections from department heads at the University. The new differential prediction techniques were designed to give estimates of expected grade-point averages for each entering student in each of the departments of the University. Some department heads feared that the new techniques would draw students from their departments and channel them into other courses. This would be bad, for departmental budget allocations were determined in part on the basis of student enrollments within departments.

There were also more sophisticated objections from the mathematics department. Who were these psychologists who, by mathematical and statistical devices, presumed to tell students what courses they should take? By this time, we had produced a number of mathematical and technical reports on predictor selection techniques, optimal test length, and differential prediction models. The head of the mathematics department was one of the most vocal skeptics, so I persuaded him to accept a set of reprints and reports on our work for his own evaluation and for that of any members of his staff that he might designate. Having heard nothing from him after some weeks, I asked him to have lunch with me. He brought with him the documents I had given him and we had a pleasant lunch together. He then returned the documents to me, expressing regret that, after glancing through them, he decided it would take too much of his time to acquire sufficient background to give me a competent evaluation. After that, we received no more objections from the mathematics department.

It took a year or two for the high schools and the colleges and universities of the state to get used to the new system and to utilize it routinely. But even before the development and implementation of the program, I and others had recognized that there was more to the optimal allocation of human resources than the differential prediction of success and that the development of underlying models would be much more involved for optimal allocation than for differential prediction.

Soon after World War II, some of us began to realize that, even if we could predict reasonably well how successful persons would be in each of a large number of jobs, this might not be an adequate basis for achieving an optimal utilization of human resources. To direct each person to the job for which he had the highest prediction of success could result in assigning more persons than are needed for some jobs and not enough persons for others. This situation stimulated the development of optimal assignment models, as distinguished from differential prediction models. Among the early contributors to this development were Brogden (1946), Dwyer (1954), Lord (1952), Thorndike (1950), and Votaw (1952). Subsequently, as computers became increasingly powerful, linear programming techniques based on the transportation model began to be adopted for military personnel classification.

My objections to these procedures were (1) that they did not adequately address the optimal criterion function and (2) that they were essentially tied to a batching procedure which was not appropriate for sequential assignment, rather than to a sequential assignment model that could also be used in batching assignment operations.

Later, I developed a preliminary sequential optimal assignment model (Horst, 1960). Subsequently, Sorenson (1965) elaborated the model and applied it to experimental data. Still later, Horst and Sorenson (1977) published an amplification of the theory and technique of the sequential assignment model.

In the early '60's at the University of Washington, I resigned my administrative position and devoted the rest of my time to teaching, research, and writing, until my retirement in 1969.

One of the things I discovered during my years at Washington was that most things published are already obsolete by the time they are in print. This is true of much that I have published. I have also noted that frequently critiques or reviews of my published work have been trivial or even incorrect. But what is even worse, the really serious defects, ones which I myself have recognized after publication, are not even discovered by reviewers. My own experience in reviewing manuscripts or publications, and in having my own work reviewed, has led me to a tentative conclusion: If one is competent to review in several hours, or at most a few days, what it has taken another even moderately competent person months or years to write, perhaps the reviewer should have written it in the first place.

Retirement

Since my retirement, I have been doing pretty much the same sorts of things as I did before. The chief difference is that I haven't been paid nearly as much for it. I have written quite a lot, just before and since my retirement, but have published very little. Some of this unpublished material seems to me to be more important than my pre-retirement publications. Examples are my work in generalized scale-free factor analysis with variable loss function, optimal assignment modeling, commonality analysis, and expectancy theory modeling. I find that since I am no longer eligible for merit increases, academic promotion, or other academic recognition, my motivation to publish is not strong, especially since I have already perished. It must be that I just like the work.

But my conviction is stronger than ever that the optimal utilization of human resources should enlist our greatest efforts, even though a major thrust in the world these days is the utilization of our natural resources for the production of instruments to destroy these human resources. And this effort can only become more effective as psychology becomes more scientific. But verbal formulations can only contribute to this objective as they have scientific meaning. After considering the matter for more than sixty years, I have finally established a series of rules by which I can test whether a verbal formulation means anything scientifically. (1) Try to translate the words into mathematical notation. (2) See if you can generate numerical data to test the mathematical formulation. (3) Try to explain the mathematics and the data to a computer. (4) See if the computer's reply means anything. If it doesn't, perhaps what was said in the first place doesn't have scientific meaning.

REFERENCES

Aitken, A. C. Note on selection from a multivariate normal population. Proc. Edinburgh Math. Soc., Ser. 2, 4, 106-110, 1934.

Bôcher, M. Introduction to higher algebra. New York: The Macmillan Co., 1927.

Brogden, H. E. An approach to the problem of differential prediction. Psychometrika, 11, 139-154, 1946.

Dickson, L. E. First course in the theory of equations. New York: John Wiley & Sons, 1922.

Dunlap, J. W. The Psychometric Society--roots and powers. Psychometrika, 7, 1-8, 1942.

Dunlap, J. W. Psychometrics--a special case of the Brahman theory. Psychometrika, 26, 65-72, 1961.

Dwyer, P. S. Solution of the personnel classification problem with the method of optimal regions. Psychometrika, 19, 11-26, 1954.

Eckart, C. & Young, G. The approximation of one matrix by another of lower rank. Psychometrika, 1, 211-218, 1936.

Ferguson, G. A. The factorial interpretation of test difficulty. Psychometrika, 6, 323-329, 1941.

Horst, P., et al. Prediction of personal adjustment. Social Science Research Council Bulletin 48, 1941.

Hotelling, H. Analysis of a complex of statistical variables into principal components. Journal of Educational Psychology, 24, 417-441, 498-520, 1933.

REFERENCES (continued)

Kelley, T. L. Crossroads in the mind of man. Stanford, Calif.: Stanford University Press, 1928.

Lord, F. M. Notes on the problem of multiple classification. Psychometrika, 17, 297-304, 1952.

Pearson, K. On the influence of natural selection on the variability and correlation of organs. Philos. Trans., A 200, 1-66, 1903.

Ruch, G. M. The objective or new type examination. Chicago: Scott, Foresman and Co., 1929.

Sorenson, R. C. Development and evaluation of a matrix transformation useful in personnel classification. Unpublished doctoral dissertation. University of Washington, 1965.

Sorenson, R. C. & Horst, P. Matrix transformation for optimal personnel assignment. Technical Note, Navy Personnel Research and Development Center, May 1977.

Spearman, C. The abilities of man. New York: Macmillan Co., 1927.

Thorndike, R. L. The problem of classification of personnel. Psychometrika, 15, 215-235, 1950.

Thurstone, L. L. A simplified multiple factor method and an outline of the computations. Chicago: University of Chicago Bookstore, 1933.

Tucker, L.R. Formal models for a central prediction system. Psychometric Monographs, No. 10, 1963.

Tucker, L. R. Some mathematical notes on three-mode factor analysis. Psychometrika, 31, 279-311, 1966.

Votaw, D. F., Jr. Methods of solving personnel-classification problems. Psychometrika, 17, 255-266, 1952.

Wedderburn, J. H. M. Lectures on matrices. New York: American Mathematical Society, 1934.