DOCUMENT RESUME

ED 061 524 CG 007 039

AUTHOR Lunneborg, Clifford E.
TITLE Psychometrics Discovers the Computer: Discovering Test Items and People.
INSTITUTION Washington Univ., Seattle.
PUB DATE Feb 72
NOTE 37p.

EDRS PRICE MF-$0.65 HC-$3.29
DESCRIPTORS *Career Choice; *Career Planning; Careers; Guidance Counseling; Guidance Functions; *Measurement; Measurement Goals; Measurement Techniques; Models; Prediction; *Predictive Measurement; *Psychometrics; Psychometrists

ABSTRACT

This is the text of a colloquium of recent computer technology on psychometrics, particularly as practiced in the University of Washington's Bureau of Testing. Seven stages of career guidance are enumerated: (1) deciding what to measure; (2) developing the measures; (3) administering the measures; (4) scoring the measures; (5) weighting the scored measures; (6) reporting predictions; and (7) evaluating effectiveness. The impact of the computer on each stage is elaborated. The author argues against what he considers to be the fallacious assumptions underlying guidance in the past. He cites as the major shift as looking at career decision making from the point of view of the decision maker rather than from the point of view of society or one of its institutions. The paper concludes with a description of a decision value oriented career guidance system. (TL)

Bureau of Testing
University of Washington—Seattle

Bureau of Testing

University of Washington

February 1972

Psychometrics Discovers the Computer

Discovering Test Items and People

Clifford E. Lunneborg

This is the text of a colloquium prepared for presentation to the Department of Psychology in January 1972. After reviewing the impact of recent computer technology on psychometrics, particularly as practiced in the Bureau of Testing, a decision value oriented career guidance system is described.

Bureau of Testing Project: 223

Psychometrics Discovers the Computer Discovering

Test Items and People

A new or advanced technology can impact on any particular field of study in at least four ways:

1) Technology can facilitate the application of known problem formulations
2) Technology can reveal that a pre-existing problem formulation carried to its technological limits is unsatisfactory
3) As a response to 2 or independently the technology may suggest answers to problems for which there was not a pre-existing satisfactory formulation
4) After the technology has been developed new problems may surface which may be solved by the technology

These might all be seen as some results of problem-development and technology each having a separate time course, sometimes independent, sometimes interactive, sometimes with one in the lead, and sometimes with the roles reversed. The technology I will be concerned with is the computer and the problem area that of the psychometric role in what I will call "Career Guidance." I will cite a number of examples of computer impact and then explore in greater detail one particular example of the fourth kind. Though my psychometric chauvinism is high I will not quite suggest that psychometry invented the computer.

First a word of orientation about psychometrics and career guidance. I use the terms guidance and career as a convenience and out of some ignorance. Guidance is an old word. It passed out of vogue, largely I gather because of some unacceptable connotation of authoritarianism, telling people what to do. It has had a rebirth recently, surprisingly so because in its rebirth it has been intimately linked, as we shall see, with letting people do, in some sense, their own thing. By career I mean most generally what one does with one's life or, if the moralistic overtones are too heavy in that formulation, at least what one does in one's life. More specifically, career guidance has to do with educational/vocational histories or, to glimpse where I'm headed, educational vocational decision-making. The psychometric role--as it has operated in my professional life--is easily summarized by saying that we want to find out what characteristics of people are associated with educational/vocational success or failure. If we can then measure those characteristics in people making educational/vocational choices, that will help them make better choices. I don't want to insult your intelligence, to use a disputatious word, by offering an oversimplified explanation. It is a very simple, if not simple-minded orientation.

Now let me own-up to some more of my short-comings and throw off what I shall cavalierly regard as excess baggage. I don't want to talk about careers at all but merely educational decision-making. Indeed, let me be so pridefully insular as to say I will be concerned only with choosing an undergraduate major field of study at the University. I could argue, but won't, that what is involved in that decision is no different than what is involved in any other career decision. I generally believe that. Perhaps after I've discussed it you will agree. Whether it is a generalizable example or not that is the small dark corner of career guidance I know best. (Any psychometrist in the audience will note that I made an ipsative and not a normative statement. I may know that area best but whether I know it well is another matter.)

I've left the computer idling now too long. What has its impact been on the career guidance I've practiced? Our approach has been, by and large, a predictive one and I can enumerate the several stages and then take them up one by one.

1) Deciding what to measure
2) Developing the measures
3) Administering the measures
4) Scoring the measures
5) Weighting the scored measures
6) Reporting predictions
7) Evaluating effectiveness

Number 1 has two aspects--a theoretical one and an empirical one. The academic prediction game grew to its current popularity within a factors of intelligence theoretical framework. The promise was that once we had correctly identified those few factors which together made up intelligence and had then written tests to measure each of these pure factors that the matter of academic prediction was well in hand--at least as far as academic performance was under cognitive control. The results have been mixed. On the one hand the extremely widespread use of the twin measures Verbal and Quantitative Aptitude attests to an important factorial contribution. For that, however, the computer was only minimally helpful. When the computer permitted us to explore the cognitive domain beyond, however, the results were painfully close to Quinn McNemar's cynical pre-computer predictions. We have proliferated cognitive factors beyond belief. To assess for an individual each of the cognitive factors enumerated by any one of the current factorial theorists is beyond man's endurance. Besides, it turns out we can't really have those pure measures of the factor--in order to understand the directions for taking this test you must have a certain amount of Factor A even though we intend this to be a measure of Factor K. And finally, beyond V and Q and possibly spatial ability or perceptual speed additional cognitive factors have not materially improved academic prediction when they have been evaluated.

I don't hold the computer responsible for this state of affairs nor can we blame the technique of factor analysis. It is our notion of the factorial nature of intelligence, that the technology permitted us to evaluate, that may indeed be found to be wanting.

I said there was an empirical side to 1 which again has been facilitated by the computer. Motivation, interests, dimensions of personality, birth order, socioeconomic status are all examples of "non-cognitive" characteristics suggested as predictors of academic success. The computer has permitted us to develop and apply prediction equations with ever increasing numbers and variety of predictors but with minimal, if any, increase in the accuracy of these predictions. The expectation that the computer, by its computational speed, would allow us to make vast strides forward by assessing great masses of potential predictors has not been realized. The shotgun, no matter what its sophistication, would not appear to be the instrument of choice.

Lest my remarks seem overly critical let me quickly enter some waivers. I am not saying that cognitive factors are valueless sources of educational decision data. I doubt there is much gain where we see our guidance goal as providing the same set of academic predictions to all members of a somewhat heterogeneous class of clients. If we can pick and choose in some way which information to use with which clients there is greater potential merit. I will have more to say on this later.

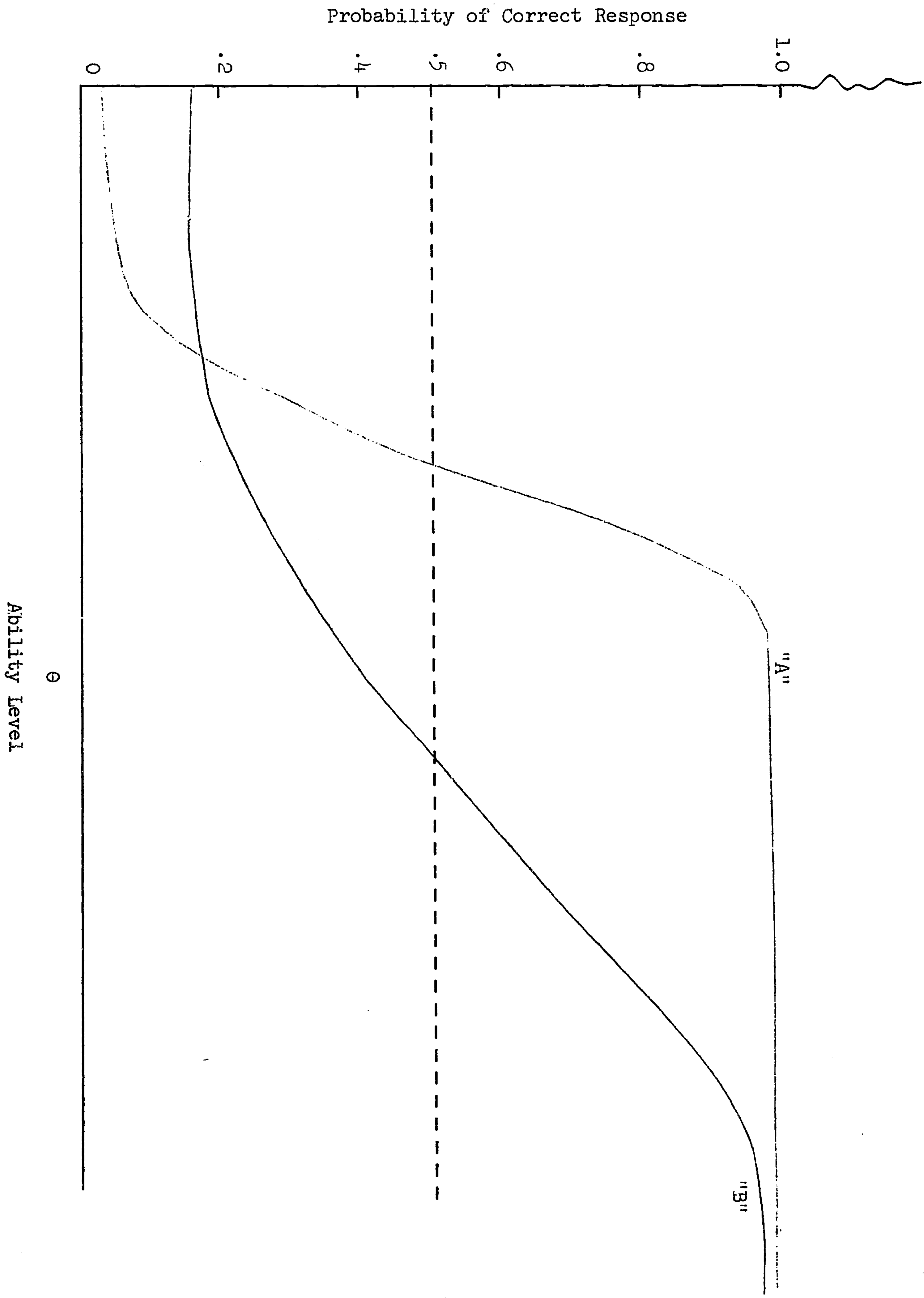
I am also not saying the computer cannot help us discover new measures. Quite the opposite. Any of you in attendance at this colloquium last quarter when Buz Hunt spoke know that we have embarked on what is to us a very exciting investigation of individual differences in cognitive functioning that is very much computer oriented. Our theoretical approach derives very much from psychology's reasoning from how computing machinery is organized and made to operate and the measures we are obtaining are ones which the computer technology makes possible but which would be precluded in either mass paper and pencil testing or even in one-on-one clinical interviews. I don't intend to discuss that research except to note that its payoff lies not in using the computer to solve old measurement problems by doing what we already knew how to do only faster, on more subjects, or over more dimensions but in capitalizing on the computer to provide new problems and new dimensions.

What I have said about the first point is largely true for each of the others. Where we apply the computer to, in essence, add more rapidly, we quickly reach limits. For example, point 2, the computer has taught us relatively little about putting together standardized tests. Why? Because, with rare exception, we employ the same strategy to select good and bad items that we did without the computer. Throw out items that have low inter-item correlations, retain items that have high inter-item correlations. Following the same rules the computer cannot give us tests that are any more homogeneous, any more internally consistent, that have higher reliabilities than ones we would assemble by hand. It only does it faster.

Fastness can make the difference, however. Let me cite one example of developing a standardized test and then go on to the more interesting, to me, non-standard test construction. We had an undergraduate honors student recently, Sue Gibson, who taught a computer to write spelling tests. It's basically a simple problem. You can tell a computer what kinds of spelling errors people make and the computer can gin up words with and without errors to form whatever kinds of distributions you like. People can do that too, of course. It just takes them a long time thumbing through dictionaries, word frequency lists, etc. Sue and the CDC cranked out strictly parallel, 50 item test forms at the rate of five tests per minute. Those tests, as nearly as we could determine in trial administrations, were every bit as good as commercially available tests. Better potentially as they were designed to permit diagnosis of particular classes of errors. A problem in which there is a thundering disinterest.

The computer mother-lode as far as test construction is concerned is the design of the non-standard test. If you want to use a standard test of, say, quantitative ability with a subject population that has considerable variation on that trait and if you want that test to provide useful--i.e., reliable--measurement all up and down that continuum then the test has to have items that function or discriminate at different levels. As a result the test is relatively long and the typical student is asked to respond to a number of items that are for him very easy and to a number which are perhaps impossibly difficult. For a variety of reasons, not the least of which is that we would prefer the student not be bored or frustrated, it would be desirable to test using only a few items of appropriate difficulty. In the in-group jargon we would like to tailor the test to the individual. The computer did not invent the goal, but it does make it realizable. We simply permit the computer to choose the next item in the sequence on the basis of responsesto previous items. The problem is essentially identical to the one Dick Rose and Davida Teller have been working with in psycho-physical scaling--there they want to choose stimuli of appropriate magnitude so as to zero in on some particular point of a psychophysical curve. The problem in psychometrics is only a bit more complicated in that our stimuli--test items--may differ on more than one dimension.

Here (Figure 1) are what are called item characteristic curves for two hypothetical items. The X axis spans the ability we are interested in measuring from low to high. The Y axis gives the probability of answering the item correctly. By our current test item theory these characteristic curves display three characteristics. Let me illustrate by reference to the two items. Item B is more difficult than item A--the ability level at which the probability of a correct answer is .50. is greater for B than A. Item A is a more discriminating item--there is a sharper transition from low to high probabilities of correct response. Finally, item B is more likely to be "guessed" than item A--at very low levels of θ , p is higher. All three characteristics are important in determining an optimal--shortest--stream of items for an individual. This three parameter tailoring model has been only recently worked out and evaluated by my colleague in the Bureau of Testing, Vern Urry. Vern joined us last year from Purdue and we will be fortunate in having him offering some graduate specialty work in the department.


Probability of Correct Response
0
.2
.4
.5
.6
.8
1.0
"A"
"B"
θ
Ability Level

A funny thing has happened on the way to the market as it were with tailored testing. It has been suggested by the dean of test theory, ETS's Fred Lord, that tailored testing works no better than a good standardized test. He may be correct, Vern argues that he is not, but if Lord is correct it is only because he views the problem within a constricting classical test-theoretic goal: namely, to move from no information to an accurate estimate of an individual's "true" score on one unobservable trait or factor. That is, I think, not the best psychometric goal as far as career guidance is concerned. There is often prior information, we are seldom concerned with evaluating a single trait or, indeed, with several pure traits. We hope to show that when more realistic goals are adopted tailored testing does have a positive payoff.

To this end we have several projects under-way. Carl Jensema, a graduate student, is evaluating the contribution of prior information. We often know something in advance of testing that ought to be useful in tailoring. For example, to use the empirical setting in which Carl is working out his approach, in the WPC testing program for high school juniors there is a need to measure mathematical ability. There is no need, however, to assume that each student tested is the hypothetical average junior. You know in advance how much school mathematics he has studied and how he has been graded in these courses. This ought to permit you to zero in on his aptitude somewhat more rapidly. What Carl will find out is how much faster for what kinds of prior information and, more practically, how should we go about constructing an item bank--from which the computer will draw its tailoring items--so that bank will take optimum advantage of a particular kind of prior information. If the approach is Bayesian it may only be with a very small b.

Dr. Urry is extending his tailoring model in two important directions. As I mentioned earlier one of the practical disappointments of factorial studies as they relate to test development was the difficulty in writing "pure" items on tests. With tailored testing this disadvantage may actually be an advantage. The typical item has a non-zero regression on more than one trait. If, as is commonly true, you are interested in measuring more than one trait then the item may reduce uncertainty simultaneously about the individual's position on more than one trait. We ought to be able to build into our tailoring strategy, then, some way of taking advantage of this to zero in on more than one measure. It would certainly be foolish to go on to assess a second trait ignoring what we had learned in assessing the first.

Dr. Urry's second model extension has even greater appeal to me. Recalling to your attention that I am presuming to talk about career guidance it seems clear to me that the student we are testing is much less interested in his mathematical and spatial ability scores than in what he can find out that will help him decide whether he should major in chemistry or biology. Vern is proposing to take criteria of that kind--differential success--rather than a trait score or scores and study the advantage of tailoring over administering a subset of standardized tests. Some very rough work Buz Hunt, Tom Love and I did a year ago dealing with some problems like this--am I more like an engineer than a Business major--from a pattern recognition orientation suggests that rather inaccurate estimates--a "high" score on vocabulary--may give you as much information as you can use.

To facilitate the development of this type of criterion oriented tailoring we--Dr. Urry, Dr. Hunt and I--have recently submitted a proposal to the U. S. Office of Education. A second--though not secondary--aspect of that proposed line of research leads me to the third and fourth parts in my list. The format of current guidance-related psychometric instruments is largely fixed by the need to have self-administering paper-and-pencil tests that can be administered to large groups and scored mechanically. Surely it ought to redound to our everlasting shame to constrict our conversational or interactive testing--to stake out some names for our computer aided tailoring--to having the computer type out multiple choice items for examinees. The computer will certainly do that less well than the printing plant. The computer can draw displays, measure reaction time, recognize "free" responses, etc. In short we can create items and evaluate responses that are quite unlike what has been feasible up to now.

We don't, however, want our technological capability to run too far ahead of what precious little theory we have and while the exploration of what is possible in the way of unconstrained response recognition, for instance, will provide some important pathways for development, we also want to flesh out the reasonably good models that have been developed for multiple choice items. One example of this is the attempt we will be making, where we can utilize an interactive approach to testing for, say, placement in introductory University courses to evaluate so-called confidence testing.

Earlier, in introducing tailored testing, I suggested that one of the characteristics of a test item was the extent to which it was susceptible to or promoted guessing. Although items do differ on this when we sum across respondents it is the differences between individuals as guessers rather than between items as occasions for guessing that interest us and provide the background for confidence testing. Every multiple-choice test has built into its scoring formula either an explicit or implicit way of controlling for guessing. You are doubtless familiar with such formula-scoring. What may not be obvious is that any fixed scoring strategy must differentially reward or, by comparison, punish the guessing strategies of individual students. Without presenting any formal analysis let me simply state that any given fixed scoring strategy (way of counting correct and incorrect answers) will yield higher scores for a particular guessing philosophy than for any other. Respondents more or less cautious than this optimum will be penalized.

The idea behind confidence testing is to get people to indicate not only which answer they think is correct but, as well, how confident they are that answer is the correct one. Oversimplifying, the test is then scored by weighting the responses made by the individual by the expressed confidence in the response--being very confident of a correct response produces more "points" than giving that same correct answer but with some ambivalence; similarly, to express great confidence in an incorrect answer brings greater penalties than offering that wrong answer with a stated uncertainty as to whether correct or not. In such a scoring setting it can be shown--to fall back on a favorite psychometric expression--it can be shown that the individual will earn his highest score by honestly reporting his confidence in his answers. Understating confidence reduces the individual's credit for correctly answered items. Overstating confidence brings greater penalties for wrong answers.

The success of confidence testing depends upon providing respondents with a way to express their confidence and, more importantly, upon training them to match expressed confidence with what they feel that to be. There has been only limited success with confidence testing in standard group paper and pencil test administration. Where testing is interactive with a computer one has an ideal way of setting up the training and considerably greater flexibility in setting up techniques for subjects to report their confidence.

Let me quickly move through points 5 and 6 on my list of elements of the predictive model and get on, finally, to what I really want to discuss. On five, weighting the scored measures, we have an example of a veritable technological Lorelei. The predictive model most frequently used in guidance psychometrics is the familiar linear regression one. Optimal but fixed weights are obtained for each predictor variable. Although it is a compensatory model--a good score on one predictor can offset bombing out on another--it has been critized non-stop for the past twenty-five years because it is insensitive to the predictive information potential in patterns, profiles, or interactions among the predictor measures, because the weight or importance attached to one predictor is fixed and does not vary dependent upon the score obtained on some second predictor.

Well, that criticism has a lot of appeal--particularly if you are not particularly pleased with the degree of predictability a straightforward linear model provides--and computers have given us the muscle to investigate more complicated models. So, we've sought after patterns and profiles, moderator variables, differentially predictable sub-groups, individual regression weights, interactions, departures from the simple linear model by a hundred different names. It has to be regarded as a bust. I'm in my thirteenth year looking for predictive patterns. My failures way outnumber my successes. We are attracted, of course, because there is no reason why, <u>in principle</u>, patterns should not contain predictive information. With any particular data base, however, we are most likely to observe that there isn't any pattern contribution. The problem is never laid simply because we can never seem to get beyond saying "these aren't quite the right measures" and searching for some more promising set or adopting this year's popular name for patterns.

Some years ago when I was really down on patterns I gave a paper at W.P.A. Buz Hunt was chairman of the session and, as I recall, he did not share my pessimism. Pattern recognition was having a fruitful life in computer science and it seemed it ought to have implications for the psychometric version of the problem. It really is the same problem--one of the few instances in science I would guess where the independent use of the same household word to name two looks at the same phenomenon has not completely disguised the fact that they are the same phenomena. Pattern recognition has succeeded in computer science because machines can indeed be taught to recognize patterns when these patterns do in fact occur in nature. It has failed in psychometrics, not because we don't know how to look for patterns, I have often argued, but simply because the blasted things aren't there.

Though I do believe that, I must report my one recent success with patterns came with the same data I mentioned earlier--data on choice of University major--that Buz and I had looked at from a pattern recognition point of view. When I reported these results--which I did in a bit more "classic" psychometric framework--at an APA symposium last fall mine was the sole pattern success. I don't want to make the success out as all that impressive, it may only be a chance phenomenon. But, in my optimistic moments it suggests that if we can learn, perhaps from the computer, some new way of looking at patterns we may find a bonanza. Of course that is absurd but why is it that in experimental work the analysis of variance makes so much room for the detection of interactions between predictors that are dichotomous--treatment conditions--when naturally-occurring, continuous predictors just don't seem to interact?

The computer has taught us little about point 6--reporting predictive data--and point 7--evaluating the effectiveness of educational prediction--and I would like to use the relative dependence of these two aspects on a previously stated (or, implied) theory or philosophy of career guidance to question first whether that philosophy should not now be abandoned and second whether our computer technology does not now provide us ways to develop a more defensible career guidance. Now, let me dispose of points 6 and 7. From the psychometric point of view our career guidance reports to clients have largely been of the form of "how well you are likely to perform" and we have evaluated our success by correlating "how well clients performed" with "how well we expected clients to perform."

I think those two statements say a lot about where psychometric career guidance has been. Let me crystallize it just a bit more, however. I have been associated for a number of years with a career guidance program--the Washington Pre-College Testing Program--which, as many of you know, tests high school juniors and then on the basis of these test results and the student's high school record provides the student and his high school counselor with a prediction of how well he might expect to do if he were to take college level course work in each of a number of different areas of study. In some respects it is a classic example of a predictive guidance program, in other respects it has some uniqueness. My point today is that despite its uniqueness it is still largely fixated by an unstated philosophy of guidance.

Elsewhere I have extolled the virtues of the WPC program. It is guidance rather than admissions oriented. The predictors were selected for their contribution to individual rather than institutional decision-making. The program doesn't focus on the absolute level of performance but upon differences in performances, not with how well a student will do but rather with whether he will do better in area A than in area B. All of these presumably nice characteristics stem from the adoption of what is called a differential prediction model developed by Paul Horst in the mid 1950's. In going that route what the WPC program does in hunting for predictors is to select measures which best account for or predict observed differences in an individual's performance (earned GPA) in, say, mathematics and biology. Predictions based on these selected measures should be maximally informative about expected differential performance.

I don't want to labor further the details of the program. I simply wanted to set the stage, to give you enough information that you can empathize with me in the following situation. Presume I have explained to someone, whose intellect I respect, how the program operates and she asks what is perhaps the most natural question in the world: "Does the program work?" My immediate response in the past has been to say that we obtain correlations up to .60 between the predicted GPA's in an area and the grade averages actually earned by students taking course work in that area. But, my questioner persists: "I understand," she says, "that your predictions are as accurate as those of any educational testing program. What I want to know, however, is whether the program works in the sense that students use these data to make wise decisions about education." In candor I would have had to reply as recently as a year ago that we really don't know, that we have surveyed high school and college students and their counselors and that they report they find the information useful. In short, however, no study has been made of whether students actually use the data to make better decisions.

Probably in the recent past such a conversation would have stimulated me to explain to my questioner, or at least to myself, how I would go about investigating that question scientifically. Before I tell you how I would have done it permit me to digress and tell a story which I hope will establish the guidance testing philosophy of the fifties and sixties. The story is one of two fables--I call them--written by Paul Horst. (The first was titled "All men are created unequal" and I don't know whether Prof. Horst considers himself fortunate to have published it when he did but it would certainly have been a different kettle of fish to have cast it up in the Jensen-Herrnstein era. That shift in mood, however, is not the one with which I am concerned today.) It is the second fable that interests me now. The title is unimportant as is the moral, which relates to sharing information with clients. The story line is what is important.

A passenger liner goes down at sea. Two groups of four men each are washed ashore, one group on each of two equally uninhabited but potentially habitable islets. One member of each group is a vocational counselor who has managed to salvage the aptitude tests basic to his trade. We leave our castaways for some months or, perhaps, years until a rescue mission visits the two sites. On the one island the rescuers find a happy, industrious society. One man is busily engaged in trapping, penning, raising, breeding, and slaughtering for food the wily wild beasts of the island. A second has domesticated and planted acres of heretofore wild turnips and rutabagas. A third has been erecting an array of huts, pens, and storehouses utilizing native stone and vegetation. The fourth, our guidance counselor, serves as a storekeeper and, in my version, leads nightly T-groups by the communal fireside. This happy state of affairs, we are told, came about because our counselor tested his three fellow islanders and told the three in turn that they had the aptitudes of a trapper, farmer, and carpenter respectively. (He, we must suppose, had a tested aptitude for testing the aptitudes of others.)

On the second island, however, there is no sign of life. Our four castaways have all perished. One we find buried under the debris, rocks and branches of a rude and poorly constructed shelter. A second has inadvertently hanged himself with a clumsy noose of vegetable fibre we must presume was intended to snare some lesser species. A third met his demise, cause unknown, while attempting to scratch a pathetic furrow in rocky soil with a fire-hardened stick. The counselor, features impassive, is found tenaciously pressing to his bosom, if that is permissible post mortem, a locked briefcase. In the briefcase are discovered full aptitude profiles for his fellow unwilling settlers and a half-finished scholarly manuscript. The aptitude profiles reveal, as you must have guessed, the three clients, had they been so advised, would have made an excellent carpenter, a resourceful hunter, and a clever agriculturist. The manuscript is a closely reasoned argument for allowing the vocationally undecided to work out their own destinies with minimal direction.

The story bears on what has been accepted guidance practice from a psychometric point of view in two ways. First, there seems to be a major premise which is moral: One _should_ do that which one is best at. That is how societies survive and prosper. One should accept as his or her goal that which will most nearly insure that the society achieves its goal. To this moral major premise has been attached a behavioral minor premise: Rational man _will choose_ to do that which he is best at. If someone fails to behave in this way it must be because he is uninformed, uninformable, or irrational.

That may seem stark to you but it is not far wide of the mark. Testing for guidance has been almost exclusively limited to providing a data base for telling the client "Your aptitudes are more nearly those of a sociologist than a physicist" or "Your interest pattern matches that of

Y.M.C.A secretaries considerably more closely than it does the pattern for successful C.P.A.'s." Although the decision is the client's, the "hidden" message here is clear enough--if in the light of this information you persist in the study of mathematics you are either an ingrate or I, the counselor, have failed to impress upon you how it really is.

Returning now to my WPC inquisitor I would earlier have proposed that the way to answer the question "Do the data help students make wise decisions" would be to find out whether students confronted with these data tend to abandon educational plans which would take them into areas of study where they might expect to do relatively poorly in favor of alternatives which promise them greater social approbation, higher grade point averages. Fortunately, I now believe, we never quite got around to doing that study. I say fortunately because it is no longer at all obvious to me that what is a wise decision for an individual to make should be the same decision as the one society would have him make in order to maximize the gain to society.

My desertion of the tried and true path is largely due to Lee Beach and his work on decisions and utility. Although he started me thinking along these lines and I am arrogant enough to believe that what I'm about to say bears somewhat on what Lee treats in decision processes I want to take him off the hook a bit. My naive use of the technical terms and operations of his trade should in no way reflect on him. Lee has afforded me a way of looking at career guidance which, when I adopt it, both seems to make a good deal more sense than my earlier orientation and allows me to say loudly to myself as I read the literature-- "Hell, I knew that last month."

The major shift is one of giving up looking at career decision making from the point of view of society or a societal institution like the university of Washington and to look at it instead from the point of view of the decision maker. Lest I be accused of deserting science to coddle students or otherwise turn soft let me offer my rationale. Students are going to make their own educational decisions. I see no mechanism looming over the horizon which would assign students to fields of study. If the purpose of career guidance is to somehow improve or facilitate this decision-making, then we ought to take advantage of what we know about decision-making.

The starting point for me has been to accept that decision-making will be regulated by the decision-maker's perception of what the relative pay-offs to him will be among the several alternatives. We have prided ourselves in the WPC program that we were concerned with the individual decision-maker and not with the institution. The program, to now, has not sought to provide institutions with data to be used in determining whom to admit or reject but to provide the student decision-maker with relevant information. What, implicitly, has been our operative notion of how that student sees the pay-offs? A simplistic one. We have behaved as though the only thing associated with choice of major at University that has utility to the student is the expected GPA. To the extent he or she perceives other aspects as bearing on the decision to be made we have been non-helpful.

The WPC program has not been quite this consciously antedeluvian. In fact it has been the recent addition of Pat Lunneborg's Vocational Interest Inventory to the test battery--the first inclusion of a measure not directly oriented to predicting GPA--that has provided a second impetus to examine how psychometric data should interface with the decision maker's activities. We ought to be able to do something more effective than to provide him with bits and pieces of information about himself and leave him to evaluate these as best he can.

Let me finally sketch out one way a guidance testing program with this decision-maker orientation might function. I am by no means alone in thinking along these lines but to date there has been, on the one hand, little integration of this orientation with the earlier tradition and, to tie back into my computer technology text, the proposed use of computer resources in guidance, where it hasn't been gimmicky, has been largely a patching up operation. It is by no means easy to move tradition. The recent report of a blue-ribbon Commission on Tests appointed by the College Board to offer suggestions on the future of that career guidance enterprise is testimony to that. They coined a memorable phrase--Symmetry of Choice--but the chapter and verse is largely more of the same. Symmetry of choice is meant to contrast with traditional, asymmetric college admissions in which colleges have required that applicants come clean about themselves; reveal their academic records, take tests, etc.; and then the college decides whom to accept. The new look, symmetry, would require that colleges would also have to describe their attributes to potential students. Students and colleges then would make parallel (but independent) choices. Where the symmetry breaks down is the point where it must be recognized that while there is no shortage of expertise to guide the colleges in converting data about students into useful decision information, the same can hardly be said for the student. How is he or she to evaluate the data provided? The experts are largely mute.

I have the same problem with regard to the healthy number of vocational or educational guidance information systems that have been developed or at least announced in recent years. These systems, based upon conversational computing capabilities, permit the student decision maker to acquire large amounts of data about potential choices from stored banks. Again little attention has been paid to how the decision maker is to integrate these data into his or her value structure. It was this situation I had in mind earlier when I remarked that the term guidance had had a rebirth but in a totally non-authoritarian context--here is "information," do with it as you will.

My problem, which can be badly stated if I am not quite careful, is that I suspect a good many potential decision makers do not trust their impressions of what is important to them. Or, to put it another way, they are not ready to decide. If that is true, data about choices may be largely irrelevant to moving them in a career development sense. A word about movement. I've had a second bitter pill to swallow. Not only have I had to accept that the decision maker will operate within his own frame of reference--meaning I can't expect him to decide in the way I would decide for him--but I am now close to deciding that the test of the effectiveness of guidance information is not even in whether it leads the client to make a decision, whatever its basis. For a long while I thought in terms of getting better decisions made. Now it seems perfectly reasonable to me that a client may exit saying something like "I know that there is nothing more that I can find out about myself or about the world at this point in time that will make a decision clear to me and I'm going to stop worrying about it. Perhaps next month I'll feel differently." This points up a second, though quite important aspect to the computer technology that could be utilized, ready access to the guidance testing program. Users should be able to access the system when they are ready to do so. It is clearly quite limiting to say to a class of 400 high school seniors on 15 September "OK, here are your test results. They tell you what you are like. Learn all you can about areas of interest to you by the first of October and we'll schedule counseling sessions between then and 15 October." If they aren't in 400 different places on 15 September they're certainly not all at the same place.

What I am suggesting by way of a career guidance testing system is in four parts as illustrated in Figure 2. If all went well one would move from (1) assessing client's values to (2) reporting utilities of options to client to (3) measuring client's skills and capabilities to (4) predicting success in options. The system would hopefully provide cumulative information relating to decisions with the client controlling the rate of progress through the system--moving on to the next stage only when satisfied with the information accumulated so far. As we shall see in a moment, the typical client may also acquire information at one stage that will prompt him to cycle back to an earlier stage to revise information generated there.

Now let's hypothesize a client with a problem, someone who feels like making a decision or at least that he would like to find out whether he can find a basis for decision. For concreteness assume a University student who might say "I think it's time I found a major and right now

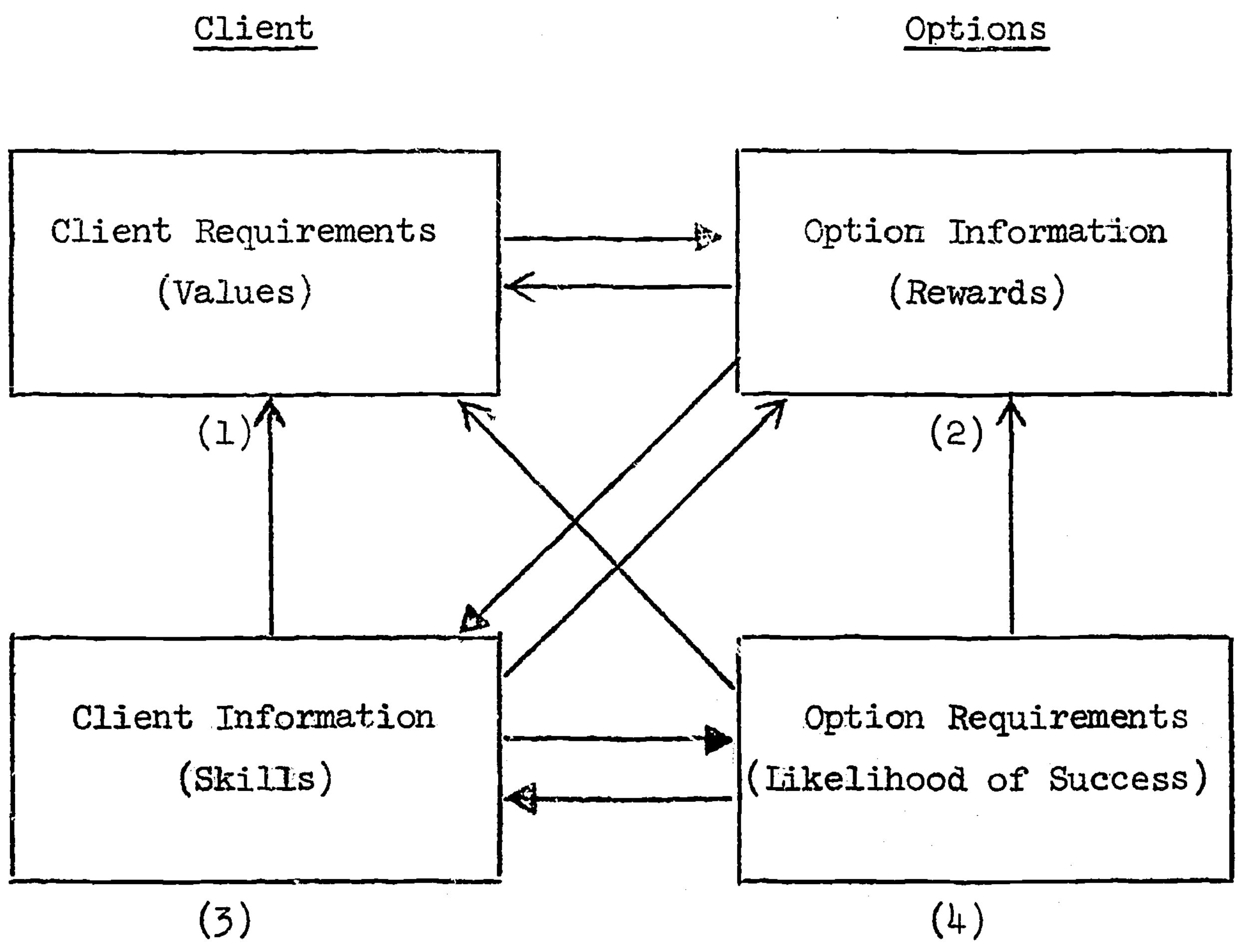


Figure 2. Proposed Flow Through a Guidance System.

→ acquiring new information

→ Revising earlier information

there are aspects of zoology, mathematics, psychology and sociology that all appeal to me."

Stage One. The focus here is on letting the client find out whether he can articulate satisfactorily, for himself, what is important to him. What are the personally relevant aspects to the options? At once we have a problem. We can opt for a measurement scheme that will produce the most accurate, reliable, replicable results; that is well standardized and probably subtle--designed to frustrate lying, malingering, yeasaying, or trying to look good. In short, sophisticated measurement oriented to reveal the true values of the client. The alternative, to draw the other extreme, is simply to ask the client "What is important to you?"

The measurement specialist should probably be expected to load his choice towards the first alternative. I do not. To me one of the most frustrating comments I hear from the consumers of test data is that they don't believe the results. In the case of the institutional user who disbelieves the normative data I still reserve the right to cretinous appelations but in the case of the individual confronted with his scores disbelief should not be dismissed. The client will find it difficult to do anything himself with such tainted data and he will almost certainly reject any fancy manipulations the specialist or his proxy computer performs upon these data. Credibility is a great deal more important than decimal accuracy.

But, technology imposes compromises. Completely unconstrained responses are too troublesome and I foresee getting our client underway by confronting him with some structure in this way: I (my computer assisted system) would say: "Here are a number of things other students have felt were important in making choices. Which ones do you think are important and which would you choose to ignore?" To this I would append an opportunity for the client to describe any aspects to the choice that he felt important but were not on my list. Hopefully, the list would be inclusive and this problem would, with some experience, be overcome. Again, however, it is important that we not try to "process" the client while he has reservations about whether what he and the computer are talking about is important.

Let us take it, however, that our client responds by saying, "These four things are what have importance to me:

1) Doing well academically, earning good grades;
2) Preparing for a job that will pay reasonably well to start;
3) Having a good chance to continue study beyond the B.A.; and
4) Doing work, after graduation, that will help solve the problems of man's impact on his environment."

We would follow up on this, since we shrewdly know something about the metrics behind these dimensions, by asking the client to operationalize these values. This is illustrated in Figure 3. Here our client translates the first into earning an average grade of B or better, the second into earning at least $9,000 the first year out of school; the third into having odds of 4 to 1 in his favor of getting into graduate school; and the last into having at least a 50-50 chance of finding a job with a man-environment impact.

These are important to the client but how important? With our client able to manipulate computer directed graphic displays it would be easy to

Selecting Values	Operationalizing Values	Scaling Values
Earning good grades	B or better	40
Earning good pay	$9,000 or more	20
Continuing education	3/1 chance of graduate school	30
Impact on environment	1/1 chance of job in area	10

Figure 3. The Value Definition Stage

have the client scale for himself the relative importance to him of the four operationalized values. Here, again in Figure 3, we have taken our client to be something of a self-professed egghead: earning good grades and getting into graduate school are relatively dominating. (We have taken the scaling task here to be one in which the client distributes 100 points over the values.) This concludes the first stage but, it should be remembered that it may be only a first run-through. As he progresses the client may well want to come back and revise this scaling (or, even the set of values that are relevant).

Stage Two. Although there was only minimal normative data involved in the first stage, the initial value list, this second stage is dominated by masses of normative data. This stage is the one that corresponds to a guidance information system. The task here is to respond to what the client has said is important in terms of the track records of the several options. It depends upon considerable data collection for each option over all of the value aspects. The relevant data base for the example here discussed would be that provided by all students completing baccalaureate programs.

Zeroing in on the particular options and value dimensions selected by our client the data to be accessed here would include, for each of the four areas:

1) GPA distributions of graduating seniors;
2) Salary distributions for graduates on their first jobs;
3) Proportion of graduates undertaking advanced training;
4) Proportion of graduates who have found employment meeting the environment criterion.

These data would be reported back to the client, as illustrated in Figure 4, relative to the particular quantitative categories he had earlier selected. Thus, he would learn the proportion of graduates in each area with GPA's of 3.0 or higher or the proportion accepting jobs that command salaries above $9,000. (The percentage values in Figure 4 are biased random numbers. They have no basis in reality and are included merely to illustrate the kinds of data to be reported.)

We now introduce a second scaling task. Although it will be a normative scaling it seems crucial, because of that, to take particular pains to explain and rationalize it to the client; to involve him actively in seeing how the percentage data are scaled. At base we want to replace the arbitrary percentage values with some index numbers which will reflect, credibly to the client we hope, the relative standing of these particular prospects of value fulfillment against the prospects across all options open to him--not just the options he has selected to study. Thus the parenthesized integers in Figure 4 would be induced by involving the client in a study across all University graduates of, say, first year post B. A. income. The particular numbers I've chosen to use here come from a "standard fives" scaling in which the complete distribution--across all options--would be chunked up into five equal width intervals. Low is 1, high is 5.

With this scaling completed the client would be led into the third phase of this stage: summing up to produce what is labelled on Figure 4 as the value return for each of the options. If I read Lee Beach correctly, clients should have little difficulty in doing this. We are simply

Options

Value Scaling	Zoology	Mathematics	Psychology	Sociology
	70% above B	50% above B	85% above B	95% above B
40	(4)	(2)	(4)	(5)
	30% above $9000	80% above $9000	40% above $9000	35% above $9000
20	(3)	(5)	(3)	(3)
	35% to grad school	65% to grad school	50% to grad school	30% to grad school
30	(2)	(4)	(3)	(2)
	75% environmental	5% environmental	30% environmental	20% environmental
10	(5)	(1)	(2)	(2)
	160+60+60+50	80+100+120+10	160+60+60+20	200+60+60+10
Value Return	330	310	300	330

Figure 4. Stage Two Information about Options

proposing that the relative reward offered by each option be weighted by the importance to the client of that reward and then sum these up over the aspects he has nominated as relevant. I've illustrated this numerically in the figure but with computer displays this would probably be done more effectively pictorially--with line lengths and areas showing how the scaled values and scaled rewards permit the options to be stacked up against each other. Unfortunately, my random numbers don't immediately suggest important differences among the options. Or, it may be a magnitude problem. If the summed returns are all big numbers differences may not be apparent. Some experimentation with scaling constants is clearly indicated.

With a couple of exceptions this is what happens in Stage Two. I expect for some clients--perhaps most--the information generated in Stage Two to lead to recycling to Stage One, re-examining and recalibrating values. I see no problem with that. After all, I have no illusions about finding out about true values. At best I hope the client can come to a picture of himself and of his future with which he can live. Indeed, I propose stimulating reexamination as a part of this second stage. We've assumed here the client nominated certain options. Our data base has information, presumably, about a great many other options. It is more than trivially probable that our computer can identify an unnominated option that appears to provide a better return than one or more of the client's selections. I would have the client confronted with this--not so much to influence any potential choice he might make as to be able to say "Look, economics, in which you are not interested, provides a better match to the way you have stated your values than does sociology. Does that suggest that you might like to revise what you now feel you value most?" Once the client has worked, to his satisfaction, back and forth between the first two stages he can elect to go on.

Stages Three and Four. The value returns for the options developed in the second stage are the returns to the client for satisfactorily completing those programs. We now need to deal with establishing the likelihood that our client can achieve one of those states of grace that is identified with a B.S. in zoology or a B.A. in sociology. The third stage, then, is the closest to a traditional guidance testing program. The fourth, running in harness with the third, presents our predictions of success based on that testing.

My reason for coupling the two--rather than doing the actuarially optimal testing first and then presenting the results--is again credibility. I want to avoid as much as possible any under the table or behind the console number shuffling. The client should be able to see each card as it is played. To be sure, we will select the cards but, to labor this ill-chosen analogy, it is probably more convincing to the client to see his possible inside straight go aglimmering than to simply have announced to him at the end of play: "You lose!"

Here, incidentally, is where I expect that both Horst's differential prediction philosophy and the tailored testing techniques I discussed earlier can be effectively used. Differential prediction to date has suffered under the load of having to predict all possible differences for all possible clients. Tailored testing has concentrated on establishing reliable estimates of true ability. These severe constraints can be lifted here. We can select those tests for the individual client that are differentially relevant to the choices he has in mind and we need administer only so much of a particular "test" as will contribute to that differentiation.

What the client might see is summarized in Figure 5. I've skipped over the initial stages, pooling the prior information about the client into a single step. In practice the client would be led through this a piece at a time--natural science course work, social science courses, high school test results, etc. What should be communicated by Figure 5 is that the client sees the decrease in uncertainty about the relative likelihood of success for the options as a consequence of the testing sequence. Note also now the likelihoods, at each step, are referred back, as multipliers, to the Stage Two value returns to provide the final expected values of the options to the client. These expected values are the final, informational product of the system; a synthesis of (1) what the client values, (2) what satisfactions the options can provide, (3) what the options require for success, and (4) what skills, competencies, etc., the client can bring to the option.

I'll stop short of going on to tell you how the "rational" client should now behave and thank you for your patient attention.

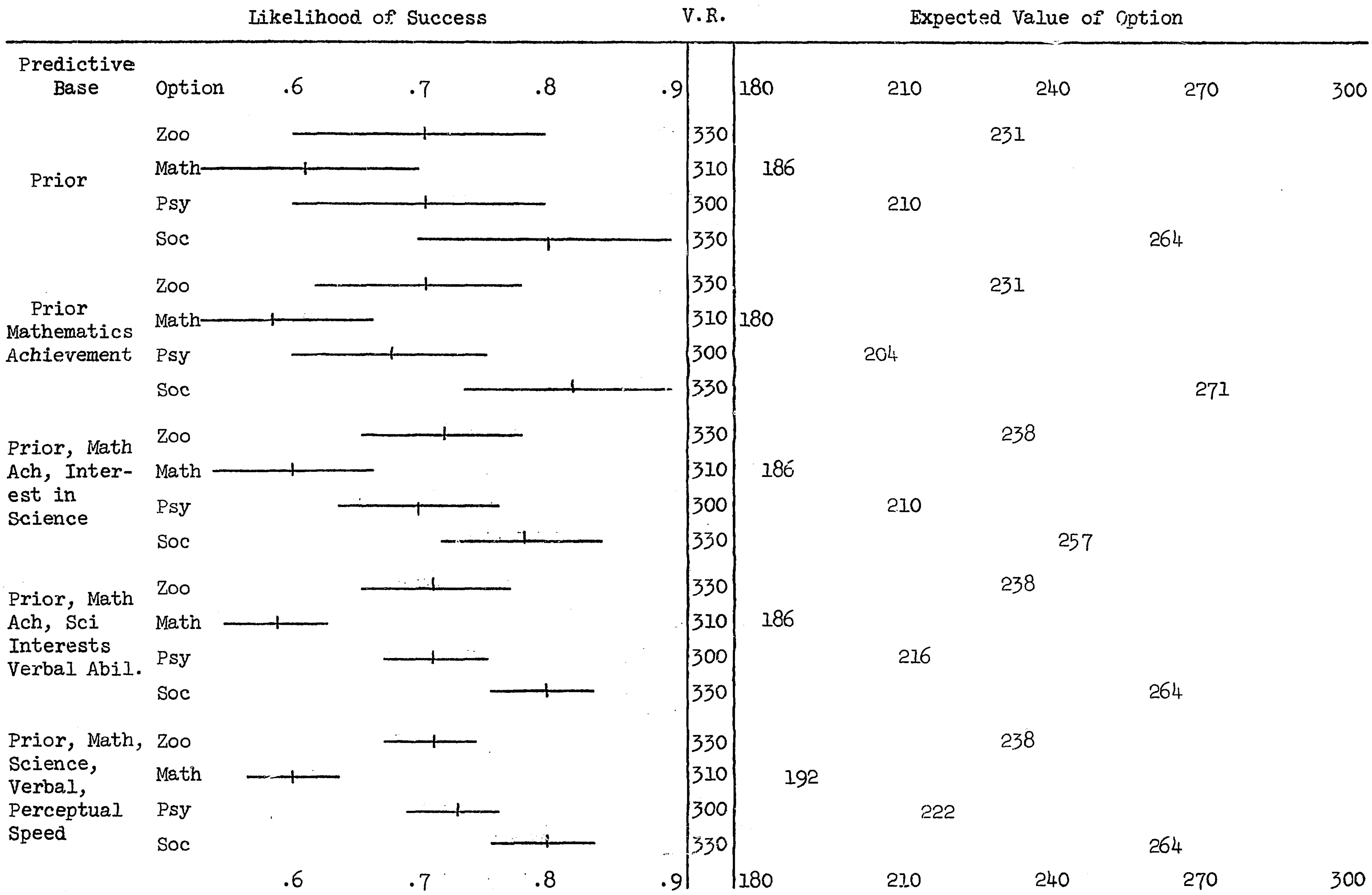


Figure 5. Likelihood of Success and Expected Values of Options