

CAREER CHOICE: HOW CAN HIGH SCHOOL GUIDANCE HELP?

Marion F. Shaycoft
 AMERICAN INSTITUTES FOR RESEARCH
 Palo Alto, California

When Albert Einstein was a young man, long before he became world-famous, he was a postal clerk in Switzerland. Obviously he was overqualified for the job. But at least to some extent the job suited him. The duties weren't arduous and he apparently could fulfill them without devoting his full efforts to this purpose. This allowed him to spend long hours thinking about problems of theoretical physics, scribbling formulas, and laying the groundwork for his theory of relativity.

My reason for mentioning this phase of Einstein's career isn't to suggest that the Project TALENT sample is full of Einsteins. An Einstein comes along only once in a century, if that often, and this century has already had its Einstein -- (to use a bit of spurious statistical logic).

The only reason I mention Einstein's postal clerk experience is to point out that the problem of overqualification for a job is not a simple one. It has lots of ramifications and no simple answer. A person who has all the abilities required to perform a highly demanding job successfully may have interests or personality traits which cause him to ^{satisfied with} be ~~in~~ a lower-level job that will be comparatively undemanding. In other ^{words} interests and personality traits may be functioning as moderator variables. (This hypothesis isn't explored in the present paper. It's just mentioned as something that might be explored in the future.)

The problem of overqualification for a job is an important one, and one which must be coped with by anyone involved with problems of educational and

career guidance. And it is a problem that multiple aptitude batteries, at least in the way they have traditionally been used, have not always dealt with successfully.

A multiple aptitude battery is a useful tool in vocational and educational guidance but there has been no clear agreement on the best way to use it. Assuming that the problem is how best to help the individual decide which of several possible occupational categories or educational programs is the one which will be best for him, three statistical techniques for developing a predictive equation for each category have been available: (1) multiple correlation other than multiple point biserial, (2) multiple discriminant analysis, and (3) multiple point biserial correlation. In the interests of brevity, the first named of those three techniques will be referred to in this paper simply as "multiple correlation."

The term "multiple correlation technique", as used in this context, means that a separate within-group criterion of effectiveness of performance is available and that for prediction of membership in each category a separate multiple correlation coefficient and multiple regression equation are computed.

The second technique, multiple discriminant analysis, is applicable when the sole criterion available is membership in a group; in other words when there is no within-group criterion to distinguish among members of the group. The technique produces a set of discriminant functions, each of which is a linear composite of the original predictor variables. The discriminant functions have to be considered in combination, since they do not correspond on a one-to-one basis to the criterion categories.

The third commonly used approach, multiple point biserial correlation is really a cross between the first two. It is multiple correlation against a dichotomous criterion, and therefore, like the multiple discriminant analysis technique, it is useable even when membership-versus-nonmembership in the group is the only criterion information available. The resulting multiple regression function is exactly equivalent (except for a scaling factor) to a discriminant function that discriminates optimally between the group in question and all other groups combined.

But even the simultaneous use of all three of these techniques --- multiple correlation (against a continuous criterion), multiple discriminant analysis, and multiple point biserial R --- may not yield all the relevant information available. None of the three really comes to grips adequately with the problem of overqualification --- in other words the situation in which overqualification for group membership is just about as undesirable as underqualification is. The problem is still further aggravated when, as often happens, group membership criteria are the only kind available. Other difficulties, particularly applicable to ^{the} discriminant function ⁺ approach, lie in the fact that because most of the discriminant functions are bipolar, they are hard to interpret, hard to explain, and generally obscure in meaning. This makes them peculiarly unsuitable for use in guidance. ~~And~~ discriminant functions have the further disadvantage that the results for one criterion category depend to an inordinate degree on what other categories happen to be included in the analysis.

In an effort to avoid all these problems and to supplement information provided by the more usual approaches, I've worked out a new approach, called propinquity analysis, which I think may prove useful. This approach requires no

criteria other than group membership, and although it superficially has some of the features of each of the other approaches, it is actually quite different from any of them.

THE NATURE AND CHARACTERISTICS OF PROPINQUITY INDEXES

Let's assume we have a scatterplot in which each individual's set of scores is represented by a point in n-dimensional space, where the standard score scale on each test constitutes a separate dimension. Let's now suppose that for purposes of developing propinquity indexes for a given occupational group the scatterplot is rescaled, weighting each dimension by a value representing, at least approximately, the relevance of the corresponding variable in identifying group members. An individual's propinquity index with respect to the groups may then be defined as his geometric distance from the group centroid, in this rescaled space. A minus sign is attached to the distance, so that 0 is the maximum value of the index. A zero index --- in other words the maximum score --- indicates that the individual's scores on relevant variables coincide exactly with the group centroid. Thus the higher the algebraic value of the index, the closer the individual is to the centroid. (Hence the term "propinquity!")

Formula 14 (or 15) represents the squared distance in regard to a single variable. Formula 16 gives the propinquity index, δ , with w as the weight representing the relevance of a particular variable. Formulas 19 and 20 show how these weights were computed. Formulas 21-26 show several alternative formulas that were originally under consideration but have now been fairly definitely rejected because they seem to work out less well than Formula 19;

In a small-scale tryout of formulas 21-26 and 19 about a year ago, for eight career groups, results suggested formula 19 was slightly better than any of the others; for six of the eight groups, it produced δ^2 values that had at least slightly higher point biserial correlations with the group membership criterion than did any of the other formulas.

Formula 27, which calls for regression weights in the formula for computing propinquity indexes, produces indexes that ~~XXXX~~ tend to have higher ~~XXXXXXXXXXXX~~ point biserial correlations with the group membership criterion than formula 19 produces; but the formula 27 indexes overlap more with the information yielded by conventional regressed scores than the formula 19 indexes do, and thus are less satisfactory than the formula 19 indexes. These findings on weights produced by formulas other than formula 19 are just mentioned in passing. No data will be presented on them in today's paper, which focuses on propinquity indexes ~~XXXXXXXXXX~~ using formula 19 weights.

Formulas 19 (like formulas 21- 26) gives a weight of 0 for irrelevant variables and a positive weight for relevant variables. There are no negative weights.

EMPIRICAL DATA

Tables 1a and 1b are based on about 14000 twelfth-grade boys tested in Project TALENT in 1960 for whom scores on all 109 predictor variables used in the study are available and for whom follow-up data obtained five years after the class graduated from high school are also available. These follow-up data provided the information about the long-range career plans of the TALENT sample. On the basis of a hierarchical analysis the career plans of these 14000 boys were organized into 20 clusters, one of them consisting of the "undecided" group; and a 21st "pseudo-cluster" consisting of everybody else --- a miscellaneous and very heterogeneous group. Table 1b shows empirical data for these 21 clusters. Table 1a shows data for 16 individual career plan groups selected out of the total of 173 for whom data were available.*

The total group has been divided into two parallel samples, A and B, so that propinquity indexes could be calculated for the individuals in one sample on the basis of statistics (means and standard deviations) determined from the other, in order to avoid capitalizing on chance.

* Ten of the 16 career-plan groups were chosen for inclusion in TABLE 1a primarily on the basis of their size (they are all large enough to provide comparatively stable data) and the variety of fields and levels they represent. The remaining six are the single career-plan "clusters", and they therefore also appear in Table 1b.

All 109 variables were normalized with a mean of 0 and standard deviation of 1, on the basis of Sample A distributions. The same normalization conversion table was then applied to the Sample B cases, to make the Sample B distributions approximately normal. Proximity indexes were calculated for every Sample B case, using weights, means and standard deviations calculated from Sample A data. (Formula 19 was used for the weights.)

Columns 1 and 2 show the number and proportion of cases in each group. The remaining columns contain biserial and point biserial correlations of one kind or another. Columns 3, 4, 7, and 9 contain point biserial or multiple point biserial correlations of the conventional kind, for various variables. Before I go into further detail about these correlations let me say a few words about the interpretation of conventional point biserial correlations. The best advice regarding point biserial correlation coefficients, of the type presented in columns 3, 4, 6 and 9, is not to try to interpret them directly. Basically they are uninterpretable until they have been converted to some other kind of coefficient. Raw unconverted coefficients are shown in Tables 1a and 1b chiefly because a test of significance must be applied to them, to determine whether they (and any coefficients to be derived from them) differ significantly from zero. Once it has been established that they are significantly different from zero they should be converted to other kinds of coefficients and not looked at again! Table ³ summarizes the data that was used to determine levels of significance of the various coefficients in Tables 1a and 1b, based on different numbers of cases and different numbers of independent variables.

The chief drawback to the use of point biserial correlations is that their magnitude is so very much a function of the nature of the split. The more extreme the split --- in other words the further from 50-50 --- the lower the maximum possible value of the point biserial correlation (at least when the continuous variable is normally distributed). The upper limit on point biserial r for various splits is shown in Table 2 (in the last column). It is seen from this table, for instance, that with a 50-50 split on the dichotomous variable, and a normally distributed continuous variable, point biserial r can go as high as about .79. With a 99-to-1 split, in contrast, it can go no higher than .27. And when the split between the two groups is even more uneven the maximum point biserial correlation is even lower.

This flaw in point biserial correlations turns out to be a major one when the total group one is studying is split into a large number of subgroups of different sizes. If we want to compare various career categories in regard to the validity of the Project TALENT test battery in forecasting whether an individual will choose that particular career, point biserial correlations are inappropriate. Not only are the groups representing the different career choices of unequal size, rendering the point biserial correlations not directly comparable, but some of the groups contain an extremely small percentage of the cases in the total group, resulting in misleadingly low point biserial correlations. I might mention, parenthetically, that in Project TALENT we are interested in the validity of prediction of career choice even for the groups containing an

extremely small percentage of the cases, because the total group is large enough that even an extremely small percentage can be a respectably large number of cases and can therefore yield reasonably stable statistical data.

In some circumstances use of biserial correlations rather than point biserials would be a reasonable solution to these problems, since for data for which it is appropriate the value of a biserial correlation is independent of the point of the split. But biserial correlations would be quite inappropriate as indicators of

the correlation between a propinquity index and the dichotomous criterion corresponding to choice of a particular career. Propinquity indexes, unlike the composite scores derived from ordinary multiple regression equations, are not linear composites, and therefore are not subject to the law of statistics that says the distribution of sums (or linear composites) of a large number of variates approaches normality even if the variates themselves are not normally distributed.

The solution is to convert each point biserial correlation to what it would be if ~~the nature of~~ ^{assumed to be} the two groups were unchanged ⁱⁿ character ~~but were~~ ~~assumed to be~~ present in equal proportions. This new kind of point biserial correlation is represented by r' in the handout. Formula 28 is the formula for r' .

Let me emphasize that for r' as defined, formula 28 yields the exact values. No approximations are built into it.

But this formula requires knowledge of the ratio of variance on the continuous variable within the appropriate category to total variance on the continuous variable; and if the ratio is unknown it must be estimated from formula 29, which yields a mere approximation, not an exact value.

Therefore if formula 29 has to be used in conjunction with formula 28, the resulting value of r' is merely an approximation --- and how good an approximation it is depends largely on how close to normal the distribution of the continuous variable is. Consequently I have used formula 29 with caution, and only where absolutely necessary.

In Table 1a ~~2/24/11~~, if columns 6 and 8 are compared, the results show that the propinquity index is about as effective a predictor of group membership as the conventional regression equation, for the following career plan categories:

250.	Architect
841.	Airplane Pilot
422.	High school science teacher
899.	Unskilled labor

This represents four of the sixteen categories shown in Table 1a --- not a large proportion but certainly far more than would be expected to occur on the basis of chance alone, if the propinquity index didn't have any real effectiveness as a predictor.

Further evidence of the effectiveness of propinquity indexes in distinguishing between members and nonmembers of a group is provided in Tables 5 and 6, which show, for four career categories, the distributions for members and nonmembers. Table 5 shows the percentage distribution and Table 7 shows the cumulative percentage distributipn. The differences between the members' and nonmembers' distributions are conspicuous.

When one is using an arbitrarily defined index, like the propinquity index, and is getting promising but not spectacular results, one can't help wondering whether a slightly different index, differently defined, might give better results. For instance would some function of the distance be better than the distance itself? Should we, perhaps, remove the square root sign from formula 16, thus defining the propinquity index as the square of the distance? Or might some other function, such as square root, be better? To get an answer to these questions, I did a little experimenting with algebraic transformations of the propinquity index.

The results, which are summarized in Table 4, certainly provided no compelling reason for switching from a distance measure to a nonlinear transformation of distance --- but perhaps further experimentation would yield other results. The functions I tried were square of the distance, square root of the distance, and some other roots. The results for the four ^{career} categories I did this experimenting on are shown in Table 4. Distance seemed to be a definitely better predictor than square of distance. And while the square root, cube root, fourth root, etc., did seem to give successively higher correlations, the differences were ^{microscopic,} ~~very small~~ and the correlations appeared to be converging rapidly to a value only trivially higher than the correlation of the criterion with the distance. Therefore, for the present at least, distance seems to be a better measure to use than any exponential transformation of it. Of course some other kind of transformation, such as normalizing the distribution of the propinquity index, could conceivably prove more effective --- but that hasn't been tried yet.

CONCLUSIONS

On the basis of data now available concerning the results of the propinquity index approach and the other approach explored in this paper (the use of regression equations corresponding to multiple point biserial correlations of the conventional type) neither of the two approaches is clearly and universally better than the other. Rather, for certain career categories, they supplement each other, since each provides some information of a kind that the other doesn't. Therefore it seems desirable to use the two techniques in conjunction with each other, at least in the case of those career categories where each provides unique information not yielded by the other.

In summary, then, the results are all significant, and the validity coefficients, though not large, are substantial.

Membership in career category is correlated with

the 109 predictors used in the conventional way to determine a conventional multiple regression equation. It is also correlated with the propinquity index derived from these same predictors. But why aren't the correlations higher than they turned out to be? The methods used, multiple regression equation, and propinquity indexes, seem to work --- but why don't they work better? I suggest that there are at least two major reasons, the first of which is

that when jobs, or even career fields, are the basis of categorization, there is a lot more heterogeneity among jobs in the same category than one might suppose. If you doubt this, just consider the wide variety of jobs represented by the people at this convention, most of whom are fall in the category of "psychologist". The

second reason the validity coefficients didn't turn out higher may be in part a consequence of the first; a lot of people are going into jobs for which they are misfits and too few are going into jobs for which they are optimally suited.