AN EMPIRICAL GROUPING OF JOB FAMILIES

Marion F. Shaycoft
American Institutes for Research
Palo Alto, California

ABSTRACT

Long-range career plans have been reported five years after high school by over 14000 of the young men who participated in Project TALENT in 1960 when they were in grade 12. These TALENT participants were classified in about 250 categories on the basis of their career plans. The patterns of TALENT scores characteristic of each of these groups were determined on 64 cognitive variables (aptitudes, abilities, and achievement levels) and 45 noncognitive variables (interests, personality traits, activities, and background variables). After reducing the number of career categories down to 93, and the number of cases down to 10494, by eliminating almost all of the groups with fewer than 50 cases, inter-group matrices containing algebraic measures of the difference between group profiles were analyzed, using a combination of hierarchical and inspectional procedures, in an effort to see what career categories cluster together on this basis.

Twelve job families, accounting for 86 of the 93 career categories, have been tentatively established on this basis. The remaining 7 of the 93 career categories are tentatively considered single-category "job families" (pending analysis of additional data, which may result in their being grouped with some of the categories not included in the 93 of the present analysis). The twelve job families containing more than one category each are:

1. engineering and applied physical sciences
2. math and physical science (other)
3. biological sciences (theoretical and applied)
4. "people-oriented" professions in the sciences
5. professions in the social sciences
6. teaching and other "people-oriented" professions (non-science)
7. businessman, salesman, and related occupations
8. farming
9. technician
10. miscellaneous skilled occupations
11. miscellaneous blue-collar jobs (trades, etc.)
12. protective.

The single-category groups are:

1. airplane pilot
2. architect
3. clergy
4. college English professor
5. high school math teacher
6. high school science teacher
7. high school physical education teacher.

It appears that at least in terms of the aptitudes, abilities, personality traits, background factors, and other relevant characteristics typical of the members of a career category when they are still in high school, there is an enormous diversity among career categories (even for traits that are quite homogeneous within category). Careers do not all fall neatly in a small number of categories. This is interpreted to mean that it should be easier for a person to find a career field suited to his pattern of abilities, interests, and other characteristics than if his choice were restricted to job families corresponding to a small number of profile patterns, none of which might resemble his.
AN EMPIRICAL GROUPING OF JOB FAMILIES

Marion P. Shycott, American Institutes for Research

Project TALENT is a long-term educational research project that started about ten years ago and is expected to continue about 25 years altogether. The project has now reached the point where questions of whether jobs can be grouped into families, and if so how, are important.

Project TALENT: The source of the data

In 1960, a comprehensive battery of tests and questionnaires lasting two full days was administered to about 400,000 students in over 1000 secondary schools. This very large sample, consisting of the students in grades 9, 10, 11, and 12 in a stratified random sample of all secondary schools -- public, parochial, and private -- in the United States, has been followed up by questionnaire one year and five years after high school graduation. The plans of the project include further follow-ups ten and twenty years after high school. Though the follow-up questionnaires cover a fairly wide range of areas, they focus most sharply on post-high school education and on jobs and long-range career plans.

One of the major purposes of the study is to provide a basis for improving vocational and educational guidance in the high schools by finding out what kinds of aptitudes, interests, achievement levels, personality traits, and other characteristics manifest at the high school stage of development are predictive of success in specific occupations. In the questionnaires the basic questions to elicit information about career plans were:

(a) What occupation do you plan to make your life work? Be as specific as possible. For instance, if military service specify type of work.

(b) What steps have you taken in this direction? (Mark as many as apply.)
   a. I now have or have had a regular job in this field.
   b. I now have or have had a job as a trainee in this field.
   c. My present job may lead to work in this field.
   d. I am doing or have done volunteer work in this field.
   e. I have had special training or education in this field.
   f. None of the above.

(c) If you have had special training or education in this field, how or where did you get it? (Mark as many as apply.)
   a. In high school.
   b. In college as an undergraduate.
   c. In graduate school or professional school after college.
   d. In some other kind of school, since high school.
   e. An apprenticeship program.
   f. On-the-job training (informal or formal).
   g. An informal program: reading or other independent study.
   h. Some other way.
   i. I have had no special training or education in this field.

Project TALENT's procedures and instruments are documented elsewhere [1,2,3,4,6,7].

Purpose of grouping

The career fields indicated by Project TALENT participants when they were followed up five years after high school were initially coded into nearly a thousand categories in order to retain as much information as possible, and to permit subsequent collapsing of categories in a multiplicity of ways. It was recognized, of course, that some collapsing would be necessary, since 1000 categories would be far more than could be handled conveniently in any data analysis. Although we could have bypassed the thousand-category stage altogether, this would have been undesirable because having more detail initially than would be needed in any one analysis would allow maximum flexibility in combining categories later on, and thus would permit different kinds and degrees of condensation of categories for different purposes.

As the first step in reducing the number of categories for potential use in educational and career guidance of high school students, the original categories were collapsed on a judgmental basis to about 250. It was felt that this was the most that could be done safely on the basis of subjective judgment; that any further combining of groups should be based on empirical data. More specifically it was hoped that on the basis of empirical data the 250 categories could be condensed into a much smaller number of groups, such that the categories combined in a single group would be relatively homogeneous in terms of the patterns of aptitudes, abilities, achievement levels, interests, personality traits, and background factors characteristic of their members. What was sought was relative homogeneity within groups and heterogeneity among groups with respect to scores on 64 cognitive variables and 45 noncognitive variables from the TALENT battery. This might simplify educational and vocational guidance to some extent, by making it possible for guidance counselors in high schools to advise the student in terms of families of jobs for which he is suited, rather than in terms of a small or large number of specific career fields from which to choose.

Besides convenience, there were other reasons for
wanting to determine some "job families."
Despite the very large number of cases Project
TALENT had started with initially, some of the
less usual career fields had very few cases in
them, and therefore might not provide stable
data unless they were combined with other close-
ly related groups.

**Methodological considerations in using hierarchi-
cal analysis**

Hierarchical analysis seemed like a promising
way of establishing job families. But the term
hierarchical analysis doesn't cover just one
specific procedure. Rather it represents a
whole family of procedures -- so that once one
has decided to use hierarchical analysis his
decisions have just begun.

In a hierarchical grouping procedure, one
generally starts with a matrix showing the
degree of difference (or similarity) between
each individual and each other individual. The
hierarchical analysis procedure operates in a
stepwise fashion, to combine individuals (or
groups) for whom the index of difference is as
low as possible, or the index of similarity as
high as possible. Difference may be expressed
in a number of ways: as distance between indi-
viduals (or groups), as dispersion of the com-
bined group, or perhaps as a ratio of distance
to dispersion. Likewise similarity can be ex-
pressed in numerous ways -- as amount of overlap
between groups, for instance, or perhaps as
correlation between group means where the means
are expressed as ipsative scores.

Some of the methodological considerations in
hierarchical analysis are discussed below.

1. **Deciding on the order of merging**

To get an idea of some of the problems, and
the point of view underlying the methodologi-
cal decisions that were made, let's look
at Figure 1. (In the interest of simplicity
this diagram and all others in this paper
are limited to the two-dimensional or two-
variable situation, but the problems and
conclusions are readily generalizable to any
number of variables.) Each circle (or
ellipse) represents a group -- let's say a
group of people in the same career field.
The radius of the circle represents the
dispersion of the group, and the dimensions
of the ellipses have an analogous interpr-
etation, so that each of the circles and
elipses in the diagram encloses the same
proportion of cases in the group it repres-
ents. Let's suppose, for instance, that
each circle or ellipse encloses 95 percent
of its group. Which pair of the groups in
Figure 1 should be combined first?

a. Should groups A2 and B2 be combined
before groups A1 and B1 are? The cen-
troids of A2 and B2 are closer together
than those of A1 and B1, and it also
appears that the A2-B2 combination
would be more compact than the A1-B1
combination. But in terms of degree
of overlap the A1-B1 combination seems
about as good as A2-B2.

b. Now let's look at A3 and B3, the two
very compact distributions represented by
the tiny circles toward the bottom
of the chart. Their centroids are
about the same distance apart as the
centroids of A1 and B1. But probably
in view of the differential overlap
rates, A1 and B1 are better candidates
for merging than are A3 and B3.

c. Should Groups A4 and B4 be merged before
A2 and B2 are, or should the A2-B2
merging take precedence? The A4-B4
pair of circles looks about the same (in
size and overlap) as the A2-B2 pair but
the A4 and B4 groups contain 200 and 400
cases respectively while the A2 and B2
groups are much smaller, containing only
50 cases each. Does this affect their
mergeability? Remember that in the
diagram the radius of a circle represents
the dispersion of the entire group rather
than sampling error. Because the A4 and
B4 centroids have smaller sampling errors
than the A2 and B2 centroids, the dis-
tance between centroids A4 and B4 is
statistically significant to a greater
degree than the distance between A2 and
B2. But this wouldn't be any reason for
merging the smaller A2 and B2 ahead of
A4 and B4. We aren't trying to limit
the merging to groups that do not differ
significantly. We are quite willing to
admit that probably no two of the popu-
lations represented by the various groups
are identical in their statistical charac-
teristics. In other words it is quite
likely that all the groups -- even the
thes we merge -- differ significantly.
The important questions is not whether
they differ, but how much they differ,
since we would like the merging confined
to groups whose differences are relatively
small. The A2-B2 pair and the A4-B4
pair are equally good candidates for merging.

d. What about the mergeability of the groups
represented by the A5-B5 pair of ellipses
in comparison with the mergeability of
the A6 and B6 groups, represented by
ellipses of similar size, shape, and
orientation? Note that the distance
between centroids A5 and B5 is the same
as that between centroids A6 and B6.
But despite all these similarities
between the A5 and B5 pair and the A6-B6
pair there is vast difference in their
mergeability. Groups A6 and B6 overlap
substantially while groups A5 and B5
hardly overlap at all. This corresponds
to the fact that in the case of the A6-B6
pair the dimension in which the distance
between centroids lies is the dimension in
which within group dispersion is largest,
while in the case of the A5-B5 groups the
opposite is true. There might be some
justification, then, for merging A6 and B6, but there is probably none for merging A5 and B5.

e. How about A7 and B7? The distance between centroids is about twice as large for this pair as the A6 and B6, but the dispersions are also about twice as large, and the A7-B7 configuration is entirely proportional to the A6-B6 configuration. The 90° rotation, reversing the relationship to the horizontal and vertical dimensions, of course doesn't alter this. Thus the two pairs are equally mergeable.

f. As for A8 and B8, this pair is about the same general configuration, except for the 45° difference in angular orientation and the greater overlap, as the A5-B5 pair. Actually the A8-B8 pair has the same amount of overlap as A6-B6 or A7-B7, and is equally mergeable.

To recapitulate our conclusions in regard to Figure 1:

1. A1-B1, A2-B2, and A4-B4 are all equally mergeable.

2. A6-B6, A7-B7, and A8-B8 are all equally mergeable, and each of these pairs is far more mergeable than A5-B5.

3. A3 and B3 should not be merged.

If these are the decisions we want our hierarchical procedure to result in, what kind of formula should be used as the basis on which merge decisions are made? Mere distance between centroids, merging the two groups whose centroids are closest together geometrically, won't give the desired result. Nor will minimizing any kind of variance measure such as the mean square distance of points in the new combined group from the centroid of the new group.

Formula 8 in the Appendix gives the geometric distance between two centroids, plotted in n-dimensional space. This is the generalized Pythagorean Theorem. Formula 9 gives the distance between centroids when each dimension is appropriately scaled in terms of the standard deviation of the distance between centroids along the dimension. The formula 9 value (or any monotonic transformation of it such as its square, given by formula 7) will give the desired results. Formula 7, therefore, is the one we would have liked to use in our research on grouping of jobs. Because of practical considerations, however, we actually had to do this research as a two-stage operation, using formula 5, which gives the square of the distance between centroids, in the preliminary stage and formula 7 in the final stage. The preliminary stage consisted in hierarchical analysis; the final stage consisted in using formula 7 to check on the tentative groupings from the hierarchical analysis and making modifications where appropriate.

Practical considerations, such as limitations on computer capacity, precluded using formula 7 in the hierarchical analysis itself, since we wanted to analyze up to 173 job groups in terms of as many as 64 variables. To do this with any kind of reasonable efficiency would have required about 2½ times as big a computer as formula 5, and about 2½ times as big a computer as we had available. Although formula 5 was known in advance not to be the ideal formula for our analysis, it did turn out to be a very useful one and to work well. Several other formulas (formulas 11-14) were tried out as alternatives to formula 5. All these alternative formulas represented efforts to achieve partially the advantages of formula 7 over formula 5 without requiring a computer with any more core capacity than formula 5 requires. However most of these alternative formulas, when applied to our data, turned out to give roughly the same results as formula 5, and none gave any better results. In the interests of simplicity, therefore, formula 5, which was by far the simplest of the formulas tried in the hierarchical analysis program, was the one used operationally.

Therefore the input to the hierarchical analysis was a matrix of $d^2$ values. At each stage of the hierarchical analysis the two jobs or job groups were combined whose centroids were closest together. Formula 6 was used to compute the square of the distance between the new group thus formed and each of the other groups.

2. Kind of scores scales to be used

Having decided what formulas to use (formulas 5 and 7) to express difference between groups, the next question is what kind of scores to apply the formula to. In other words should we use the initial test scores in their raw form? Or should they be converted to some kind of factor score, or discriminant function, or some other kind of derived scores? And should the number of variables be reduced through some such procedure as converting to factor scores and then using only the first few factors? It was decided that orthogonal scores would be quite necessary but that dimension reduction would be extremely undesirable. The advantage of orthogonal variables was that they would result in a meaningful $d^2$ matrix uncontaminated by the effects of correlation.

But what kind of orthogonal variables? It was decided that our needs in this direction would be served best by principal components, scaled in the usual way — with zero means and unit standard deviations. The possibility of using discriminant functions instead
of principal components in order to get orthogonal variables was given careful consideration and rejected. Discriminant functions, unlike principal components, are normally scaled in such a way that their variances are proportional to their overall effectiveness in discriminating among groups. Principal components with uniform standard deviations of 1 obviously lack this feature, as far as individuals are concerned. But for group centroids this deficiency is self-correcting, since the dispersion among groups means is of course far greater for the principal components that discriminate effectively among groups than for those that don't.

3. Should a scale-free method be used?

All the problems concerning choice of a measure of geometric distance and/or dispersion suggested an entirely different possibility -- the possibility that perhaps we should bypass all these considerations of a strictly metric nature by using a method that is both simple and computationally invariant under monotonic transformation of the data -- Johnson's ultrametric maximum method, for instance, or his ultrametric minimum method [5]. After careful consideration it was decided that these methods were not suitable for the kind of data we had. Let's look at some strictly hypothetical data illustrating one of the disadvantages. Figure 2 shows an example of the ultrametric maximum method and Figure 3 the ultrametric minimum method. Both are artificial data, but the results are bizarre enough to give some idea of the sorts of peculiarities that may result when useful parametric data are ignored. According to the maximum method, the pair of groups to be merged is the pair for which the maximum distance between a point in one group and a point in the other group is smallest. This procedure is intended to yield maximally compact groups -- but that isn't always the actual result. Figure 2a presents a 14-point grouping problem. Three of the 14 points of Figure 2a form a very compact cluster at the left, while the other 11 form a somewhat more diffuse cluster at the right. Two solutions are presented -- one in Figure 2c and the other in Figure 2d. Figure 2b shows two intermediate stages of grouping which would occur if the \( d^2 \) criterion (square of geometric distance) were used. (The first of the two would also occur with the ultrametric maximum method.) Figure 2c shows the final grouping resulting from the maximum method. Three of the points that seem rightfully to belong in the right-hand cluster are joined with the left-hand cluster. This strange result seems to be a blatant instance of "empire-building" by group J. Figure 2d shows the more normal results obtained through the use of \( d^2 \). (Table 1 is the distance matrix corresponding to Figure 2. It shows the distances themselves, not their squares, but this makes no difference in the results.)

The ultrametric minimum method merges those two groups closest to each other, when the closeness of two groups is defined as the distance between the two points that are closest to each other. Figure 3 shows a dumbbell configuration of points with a small hexagonal arrangement near one end of the dumbbell. As shown at the bottom of Figure 3, when the minimum method is used the entire dumbbell turns out to be one cluster, even though the two obvious clusters of the dumbbell are actually connected by only the most tenuous chain of points. If any single point on the chain joining the two ends of the dumbbell were dropped from the configuration the dumbbell would collapse into two parts immediately.

4. When to stop merging

Getting back to our old-fashioned metric data of formulas 5 and 7, how does one decide when these values have become too large to warrant further combining of groups? For formula 7 the answer lies in the fact that there is a way of interpreting the numerical values in geometric terms. This is shown in Figure 4. As in Figure 1, the centers of the Figure 4 circles represent the centroids and the radii are assumed to indicate the dispersion.

Since the circles include almost everyone in the group, the last pair corresponds to two groups that have almost no overlap. The \( D^2 \) is 8 (where \( D^2 \) is defined by formula 7). The first pair, which has very substantial overlap, has a \( D^2 \) of only .32. It seems undesirable to combine jobs that have a \( D^2 \) much above 1.50, because the people in them are too different to be lumped together in one heading. It was therefore decided to apply this rather stringent criterion to our empirical data, in determining what career groups to merge. There didn't seem to be any compelling reason for forcing every job to be combined in a "family" with other jobs if there were some that didn't fall into natural clusters.

So much for the methodological decisions on grouping. Now let's get to our actual empirical study, applying the methods we decided were most suitable for the grouping of jobs.

**General procedures in the empirical study**

The grouping study was based on the test scores and other data collected on 14,123 grade 12 boys who responded to the follow-up questionnaire sent them five years after high school. To get orthogonal variables for the total group, two principal components analyses were carried out -- one for the cognitive variables and one for the noncognitive. As many principal components were obtained as there were variables in the battery -- 64 for the analysis of cognitive variables and 45 for the noncognitive. All of these are being used, since there seems to be no advantage to dimension reduction in this situation and rather substantial disadvantages in
terms of the potential loss of information that
could result from reducing 64 or 45 variables to
a substantially smaller number. Preliminary
results are presented in this report.

After the respondents to the follow-up question-
naire were classified into the a priori catego-
ries on the basis of their career plans, the
groups were "purified" by eliminating cases
where the alleged career plan seemed to have
little basis in reality. For instance anyone who
indicated five years after high school that he
intended to become a physician was excluded
from the purified group of prospective physi-
cians if he had not even entered college yet.
Objective criteria were set up in advance for
each career category, defining what kinds of
responses, if any, would result in exclusion
from the purified group.

Tentative job families were established on the
basis of hierarchical analysis of the cognitive
data and eliminated or modified if the analysis
of the noncognitive data didn't confirm them.
(Actually as things turned out, the cognitive
and noncognitive results agreed very well.
Hardly any groups had to be eliminated or
changed on the basis of different findings
from the two analyses.)

In establishing tentative groups on the basis
of results of the hierarchical analyses, some
liberties had to be taken with the hierarchical
model, because the data appeared not to cluster
in groups which fitted this model very closely.
There appeared to be only a very small number
of nuclei of clusters and before these nuclei
acquired a large number of "satellite groups"
in the hierarchical development they tended to
coalesce. Thus, depending on where the merging
process was stopped we were presented with the
choice of either a small number of clusters,
most of them including only about three or four
groups each, supplemented by a very large number
of separate groups (i.e. single-group "clusters")
or, alternatively, one very large cluster,
which has swallowed up the few smaller
separate clusters and has also swallowed up most
of the separated groups. If we were to hold
strictly to the hierarchical model there didn't
seem to be any happy medium between these two
extremes. However study of the hierarchical
data led to the conclusion that meaningful and
useful clusters could be established from the
long chain of careers groups that the hierarchi-
cal analysis tended after a while to yield, by
breaking the chain at carefully chosen points to
split it into several sections. In a few in-
stances there was some ambiguity as to the
exact point at which it would be best to break
the chain, because the career group in the
vicinity of the proposed split seemed to fit
equally well on either side. In such cases,
rather than make an arbitrary decision, the
career group in question was included in both
clusters.

After clusters based on the hierarchical analy-
sis had been tentatively determined, a final
check was made on the basis of $D^2$ matrices
(formula 7 data). Because the values appeared to
be somewhat unstable for very small groups, the
$D^2$ matrices were limited to career groups con-
taining at least 50 cases, supplemented by a
few smaller groups that on the basis of the
hierarchical analysis appeared to cluster
with them. This resulted in limiting the
number of groups in the $D^2$ matrix to 93.

Empirical results

The modified hierarchical procedure described in
the previous section, in conjunction with the $D^2$
matrix procedure (formula 7) reduced the 93
career plan groups that were included (plus a
94th group: "undecided") down to 19 categories
(plus a 20th for the "undecided" group). The
categories are summarized in Table 2. Of the 19
categories, only 11 were clusters containing
at least four groups. One contained just two
groups and each of the remaining seven consisted
of just a single career group that didn't
cluster with anything else. (Among these seven
unique and relatively homogeneous groups were
architect and clergyman.)

As the opposite side of the same coin we have the
handful of jobs that seemed to cluster naturally
with more than one job family. A case in point
is computer programer, which was the only career
group falling in three separate job families.
This unusual multiplicity of categories can
probably be attributed to the fact that there
are so many different kinds and levels of
programmers that one might almost say that the
term "computer programer" doesn't denote any
one job category.

Since we started with 173 career groups (in the
hierarchical analysis) and ended with only 93
going into the 19 categories, what happened to
the other 80? The answer is that each of these
80 groups had fewer than 50 cases, and some had
fewer than 10. These groups, then, because of
their small size, probably had rather unstable
centroids. Consequently it still isn't entirely
clear whether they are unique, clustering with
no other group, or whether more data would fix
their centroids so that it would become apparent
that they belong in a cluster with other groups.

Table 3 shows the composition of each of the
clusters, and Table 4 shows how homogeneous each
cluster is, by presenting the within-cluster
range of $D^2$ values, separately for the cognitive
and noncognitive variables.

It is interesting to observe that clusters that
were relatively homogenous in terms of the
cognitive variables turned out to be fairly
homogenous on the noncognitive variables too.
As a further check, it is planned to investigate
whether the Grade 11 data confirm the clusters
established on the basis of the Grade 12 data.

But what significance are we to attribute to the
fact that there were so many jobs that didn't
fall in tight clusters and so few that did?
Probably the basic significance of this outcome
lies in all that is implied by the apparent
nonexistence of a clear hierarchical structure
underlying the career groups. The centroids of
these career groups are to a great extent scat-
tered widely in n-dimensional space rather than
falling neatly in tight little clusters. The
patterns of aptitudes and abilities that char-
acterize various jobs are perhaps almost as
diverse as the corresponding patterns for people.
Therefore in selecting a career field, there
should be less need for a square peg -- or even
a scalene triangular peg -- to force himself in-
to a round hole than there would otherwise be.
The full range of jobs should include lots for
scalene triangles of different shapes and sizes.

REFERENCES

[1] Flanagan, J. C., Dailey, J. T., Shaycoft,
Marion F., Gortham, W. A., Orr, D. B., &
Goldberg, I. Design for a Study of American Youth,

[2] Flanagan, J. C., Dailey, J. T., Shaycoft,
Marion F., Orr, D. B., & Goldberg, I. Studies of the American High School.

[3] Flanagan, J. C., Davis, F. B., Dailey, J. T.,
Shaycoft, Marion F., Orr, D. B., Goldberg,

[4] Flanagan, J. C., Cooley, W. W., Lohnes, P. R.,
Schoenfeldt, L. F., Holdeman, R. W., Combs,


Newman, C. A., Jr., & Sherman, S. E. Studies

[7] Shaycoft, Marion F. The High School Years:

of an Hierarchical Grouping Procedure to a

[9] Ward, J. H., Jr., Buckhorn, Janice, and Hall,
Kathleen, Introduction to PERSUB, PRL-TR-67-3 (I), Lackland AFB, Texas: Personnel Research Laboratory, Aerospace Medical Division, August 1967.

[10] Ward, J. H., Jr., Hall, Kathlen, and Buck-
horn, Janice, PERSUB Reference Manual,
PRL-TR-67-3 (II), Lackland AFB, Texas:
Personnel Research Laboratory Aerospace
Medical Division, August 1967.

NOTES

1 This research was conducted as a part of Project TALENT, a project conducted jointly by the American Institutes for Research and the University of Pittsburgh, pursuant to a grant from the U.S. Office of Education, Department of Health, Education, and Welfare. The opinions expressed herein, however, do not necessarily reflect the position or policy of the U.S. Office of Education, and no official endorsement by the U.S. Office of Education should be inferred.

2 All formulas in this paper are in the Appendix, which also contains a section defining all the notation used.

3 The author is indebted to Robert A. Bottenberg
and Joe H. Ward, Jr., with whom she discussed the proposed analyses, for their helpful advice concerning hierarchical analysis; to Dr. Bottenberg for making available a copy of USAF Personnel Laboratory's GROUP 4 hierarchical analysis program [8,9,10]; and to Bradford W. Wade, who wrote a versatile hierarchical program (HIER) that used GROUP 4 as a starting point but was specifically designed for Project TALENT's needs, and who also wrote the series of auxiliary programs needed for preparing input to the hierarchical analyses and for computing the subsequent formula 7 matrices.

4 All of the alternative formulas that were tried out are incorporated as options in the HIER program referred to in Note 2 above, as are many other formulas.

5 Of the 4 alternative formulas, formula 14 gave results closest to formula 5 (and 7). Formula 13 gave the most dissimilar and least meaningful results. It did not work well with these data.

APPENDIX

I. NOTATION

\[ n = \text{no. of variables (} = \text{no. of dimensions); } \]
\[ g = \text{no. of groups; } \]
\[ N_j = \text{no. of cases in group } j; \]
\[ N = \text{total no. of cases; } \]
\[ g = \sum_{j=1}^{N} N_j \]  \hspace{1cm} (1)
\[ y_{ijk} = \text{score of individual } k \text{ in group } j \text{ on variable } i \]

\[ i = 1, 2, 3, \ldots n \]
\[ j = 1, 2, 3, \ldots g \]
\[ k = 1, 2, 3, \ldots N_j \]
\[ \overline{y}_{ij} = \frac{\sum_{k=1}^{N_j} y_{ijk}}{N_j} \]  \hspace{1cm} (2)

\[ s_{ij} = \text{sample standard deviation of variable } i \text{ for group } j \]
\[ s_{ij} = \sqrt{\frac{\sum_{k=1}^{N_j} (y_{ijk} - \overline{y}_{ij})^2}{N_j}} \]  \hspace{1cm} (3)

\[ \sigma_{ij} = \text{estimate of population standard deviation of variable } i \text{ for group } j \]
\[ \sigma_{ij} = s_{ij} \sqrt{\frac{N_j}{N_j - 1}} \]  \hspace{1cm} (4)

\[ F_{ijk} \text{ = special case of } y_{ijk}, \text{ where the variables are principal components.} \]

\[ d_{AB}^2 = \text{squared distance between centroids of groups A and B.} \]

\[ d_{(AB)C}^2 = \text{squared distance between centroid of combined groups A and B, and centroid of any other group } C. \]

\[ D_{AB}^2 = \text{square of distance between centroids of groups A and B, where each dimension is scaled so that the standard deviation of the distance between a point in group A and a point in group B is uniform for all dimensions. (This scaling has the effect of changing elliptical configurations to circular ones.)} \]

\[ \sigma_{AB}^2 = \frac{n}{N_A + N_B - 2} \left( \frac{\sum_{i=1}^{N_A} s_{1A}^2 + \sum_{i=1}^{N_B} s_{1B}^2}{N_A + N_B - 2} \right) \]  \hspace{1cm} (10)

\[ \sigma_{d_{AB}}^2 = d_{AB}^2 + \sum_{i=1}^{n} (s_{1A}^2 + s_{1B}^2) \]  \hspace{1cm} (11)

\[ \sigma_{A+B}^2 = \frac{1}{N_A + N_B - 1} \left[ (N_A + N_B - 2) \frac{\sigma_{AB}^2}{N_A + N_B} + \frac{N_A N_B}{N_A + N_B}
\right. \]

\[ + \left. \frac{N_A N_B}{N_A + N_B} \frac{d_{AB}^2}{N_A + N_B} \right] \]  \hspace{1cm} (12)

\[ D_{AB}^2 \#1 = \frac{\sigma_{d_{AB}}^2}{2 \sigma_{AB}^2} \]  \hspace{1cm} (13)

\[ D_{AB}^2 \#4 = \frac{d_{AB}^2}{\sigma_A^2 + \sigma_B^2} \]  \hspace{1cm} (14)
Figure 1. The Grouping Problem: What Groups Get Combined?

Note:
Center of circle or ellipse represents centroid of group. Circle or ellipse is assumed to enclose 95 percent of the group.
Figure 2. A Pitfall of the Ultrametric Maximum Method of Hierarchical Analysis: Empire-Building by Small Clusters at the Expense of Large Clusters

Figure 2a. A 14-point grouping problem

All line segments shown on this diagram are 2 units long. Except for the J and H points, all points are uniformly spaced, in a two-dimensional reticular pattern.

Figure 2b. Two intermediate stages in hierarchical analysis

Figure 2c. Final grouping that would result from Ultrametric Maximum Method

Figure 2d. Final grouping, using $d^2$ as the criterion

TABLE 1. Distance Matrix Corresponding to Figure 2

<table>
<thead>
<tr>
<th></th>
<th>C1</th>
<th>C2</th>
<th>C3</th>
<th>D1</th>
<th>D2</th>
<th>E1</th>
<th>E2</th>
<th>F1</th>
<th>F2</th>
<th>G</th>
<th>H</th>
<th>J1</th>
</tr>
</thead>
<tbody>
<tr>
<td>C1</td>
<td>2.00</td>
<td>2.00</td>
<td>3.46</td>
<td>4.00</td>
<td>2.00</td>
<td>3.46</td>
<td>2.00</td>
<td>2.00</td>
<td>2.00</td>
<td>3.86</td>
<td>6.70</td>
<td></td>
</tr>
<tr>
<td>C2</td>
<td>2.00</td>
<td>2.00</td>
<td>2.00</td>
<td>2.00</td>
<td>2.00</td>
<td>4.00</td>
<td>3.46</td>
<td>3.46</td>
<td>3.86</td>
<td>8.70</td>
<td></td>
<td></td>
</tr>
<tr>
<td>C3</td>
<td>2.00</td>
<td>3.46</td>
<td>3.46</td>
<td>4.00</td>
<td>3.46</td>
<td>2.00</td>
<td>4.00</td>
<td>2.00</td>
<td>7.89</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>D1</td>
<td>2.00</td>
<td>4.00</td>
<td>3.46</td>
<td>5.29</td>
<td>4.00</td>
<td>5.29</td>
<td>2.83</td>
<td>9.85</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>D2</td>
<td>3.46</td>
<td>2.00</td>
<td>6.00</td>
<td>5.29</td>
<td>5.29</td>
<td>4.79</td>
<td>10.70</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>E1</td>
<td>2.00</td>
<td>3.46</td>
<td>4.00</td>
<td>2.00</td>
<td>5.46</td>
<td>7.89</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>E2</td>
<td>5.29</td>
<td>5.29</td>
<td>4.00</td>
<td>5.82</td>
<td>9.85</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>F1</td>
<td>2.00</td>
<td>2.00</td>
<td>4.79</td>
<td>4.70</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>F2</td>
<td>3.46</td>
<td>2.83</td>
<td>5.96</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>G</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>5.82</td>
</tr>
<tr>
<td>H</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>8.56</td>
</tr>
</tbody>
</table>
Figure 3. The Dumbbell Configuration: A Pitfall of the Ultrametric Minimum Method.

Applying the ultrametric minimum method of hierarchical analysis to the configuration shown at the right would result in the two clusters shown below. Note that the two ends of the "dumbbell" are in the same cluster.
Figure 4: Amount of overlap corresponding to various values of $d^2$

<table>
<thead>
<tr>
<th>$d/\sigma_{ij}$</th>
<th>.8</th>
<th>1</th>
<th>$\sqrt{2}$</th>
<th>$\sqrt{3}$</th>
<th>2</th>
<th>3</th>
<th>4</th>
</tr>
</thead>
<tbody>
<tr>
<td>$D^2$</td>
<td>.32</td>
<td>.50</td>
<td>1.00</td>
<td>1.50</td>
<td>2.00</td>
<td>4.50</td>
<td>8.00</td>
</tr>
</tbody>
</table>

Note: Center of circle represents centroid of group. Radius of circle = $2\sigma_{ij}$

---

<p>| TABLE 2. Job Families for Males*: Partial list (Based on 5-year follow-up of grade 12 boys) |</p>
<table>
<thead>
<tr>
<th>No. of career groups included</th>
<th>No. of cases included</th>
</tr>
</thead>
<tbody>
<tr>
<td>A. Airplane pilot</td>
<td>1</td>
</tr>
<tr>
<td>B. Business and industry</td>
<td>19</td>
</tr>
<tr>
<td>C. Architect</td>
<td>1</td>
</tr>
<tr>
<td>D. Engineering and applied physical sciences</td>
<td>12</td>
</tr>
<tr>
<td>E. Math and physical science: Quantitatively oriented professions</td>
<td>5</td>
</tr>
<tr>
<td>F. Biological sciences: Theoretical and applied</td>
<td>6</td>
</tr>
<tr>
<td>G. &quot;People-oriented&quot; professions in the sciences</td>
<td>4</td>
</tr>
<tr>
<td>H. Professions in the social sciences</td>
<td>4</td>
</tr>
<tr>
<td>I. College professor: English</td>
<td>1</td>
</tr>
<tr>
<td>J. Clergyman</td>
<td>1</td>
</tr>
<tr>
<td>K. Teaching and other &quot;people-oriented&quot; professions (non-science)</td>
<td>6</td>
</tr>
<tr>
<td>L. High school math teacher</td>
<td>1</td>
</tr>
<tr>
<td>M. High school science teacher</td>
<td>1</td>
</tr>
<tr>
<td>N. High school physical education teacher</td>
<td>7</td>
</tr>
<tr>
<td>O. Miscellaneous skilled occupations</td>
<td>6</td>
</tr>
<tr>
<td>P. Technician</td>
<td>18</td>
</tr>
<tr>
<td>Q. Miscellaneous &quot;blue-collar&quot; jobs</td>
<td>2</td>
</tr>
<tr>
<td>R. Farming</td>
<td>4</td>
</tr>
<tr>
<td>S. Protective</td>
<td></td>
</tr>
<tr>
<td>T. (Undecided)</td>
<td></td>
</tr>
<tr>
<td>**</td>
<td>100**</td>
</tr>
</tbody>
</table>

*Based on two hierarchical analyses of $d^2$ matrices (combining groups with the smallest $d^2$), with the resulting groups modified on the basis of the corresponding matrices of $D^2$ values. One of the $d^2$ matrices on which a hierarchical analysis was based was in 64 dimensions (principal components of 64 cognitive variables) and the other was in 45 dimensions (principal components of 45 noncognitive variables).

This table includes all groups having at least 50 cases (in the five-year follow-up of grade 12 boys). Some smaller groups are also included which fit into a family defined by the larger groups.

**Includes 7 duplications; therefore 93 separate categories.

***Includes 756 duplications; therefore 10494 separate cases.
<table>
<thead>
<tr>
<th>Job Family</th>
<th>Career Code</th>
<th>Career</th>
<th>No. of cases</th>
</tr>
</thead>
<tbody>
<tr>
<td>A. 841</td>
<td>Airplane pilot</td>
<td></td>
<td>99</td>
</tr>
<tr>
<td>B. 112</td>
<td>In business for self (NEC)</td>
<td></td>
<td>278</td>
</tr>
<tr>
<td>120</td>
<td>Industry or business (NEC)</td>
<td></td>
<td>141</td>
</tr>
<tr>
<td>711(2)</td>
<td>Banking and finance</td>
<td></td>
<td>116</td>
</tr>
<tr>
<td>716</td>
<td>CPA</td>
<td></td>
<td>280</td>
</tr>
<tr>
<td>717</td>
<td>Accountant, auditor, comptroller (exc. CPA)</td>
<td></td>
<td>202</td>
</tr>
<tr>
<td>723</td>
<td>Efficiency expert, industrial engineer, production management</td>
<td></td>
<td>111</td>
</tr>
<tr>
<td>730</td>
<td>Business management, business administration (NEC)</td>
<td></td>
<td>511</td>
</tr>
<tr>
<td>731</td>
<td>Manufacturing management</td>
<td></td>
<td>95</td>
</tr>
<tr>
<td>732</td>
<td>Wholesale or retail trade management; marketing</td>
<td></td>
<td>195</td>
</tr>
<tr>
<td>746</td>
<td>Insurance salesman</td>
<td></td>
<td>61</td>
</tr>
<tr>
<td>748</td>
<td>Salesmen (NEC)</td>
<td></td>
<td>350</td>
</tr>
<tr>
<td>749</td>
<td>Sales manager</td>
<td></td>
<td>73</td>
</tr>
<tr>
<td>726</td>
<td>Personnel administration</td>
<td></td>
<td>81</td>
</tr>
<tr>
<td>335</td>
<td>Pharmacist</td>
<td></td>
<td>91</td>
</tr>
<tr>
<td>222*</td>
<td>Computer programe</td>
<td></td>
<td>121</td>
</tr>
<tr>
<td>642*</td>
<td>U.S. Armed Forces: Officer</td>
<td></td>
<td>108</td>
</tr>
<tr>
<td>747**</td>
<td>Auto salesman</td>
<td></td>
<td>22</td>
</tr>
<tr>
<td>743**</td>
<td>Stockbroker</td>
<td></td>
<td>23</td>
</tr>
<tr>
<td>745**</td>
<td>Real estate sales</td>
<td></td>
<td>26</td>
</tr>
<tr>
<td>C. 250</td>
<td>Architect</td>
<td></td>
<td>58</td>
</tr>
<tr>
<td>D. 240</td>
<td>Engineer (NEC)</td>
<td></td>
<td>167</td>
</tr>
<tr>
<td>241</td>
<td>Civil and/or hydraulic engineer</td>
<td></td>
<td>105</td>
</tr>
<tr>
<td>242</td>
<td>Electrical and/or electronic engineer</td>
<td></td>
<td>243</td>
</tr>
<tr>
<td>243</td>
<td>Mechanical or automotive engineer</td>
<td></td>
<td>143</td>
</tr>
<tr>
<td>231*</td>
<td>Chemist</td>
<td></td>
<td>81</td>
</tr>
<tr>
<td>642*</td>
<td>U.S. Armed Forces: officers</td>
<td></td>
<td>108</td>
</tr>
<tr>
<td>222*</td>
<td>Computer programer</td>
<td></td>
<td>121</td>
</tr>
<tr>
<td>244**</td>
<td>Aeronautical engineer</td>
<td></td>
<td>37</td>
</tr>
<tr>
<td>245**</td>
<td>Chemical engineer</td>
<td></td>
<td>58</td>
</tr>
<tr>
<td>234**</td>
<td>Geologist</td>
<td></td>
<td>17</td>
</tr>
<tr>
<td>235**</td>
<td>Meteorologist</td>
<td></td>
<td>10</td>
</tr>
<tr>
<td>230**</td>
<td>Scientist (NEC)</td>
<td></td>
<td>18</td>
</tr>
<tr>
<td>E. 211</td>
<td>Mathematician</td>
<td></td>
<td>46</td>
</tr>
<tr>
<td>232</td>
<td>Physicist</td>
<td></td>
<td>64</td>
</tr>
<tr>
<td>231*</td>
<td>Chemist</td>
<td></td>
<td>81</td>
</tr>
<tr>
<td>462*</td>
<td>College professor: Science</td>
<td></td>
<td>68</td>
</tr>
<tr>
<td>461**</td>
<td>College professor: Math</td>
<td></td>
<td>28</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Job Family</th>
<th>Career Code</th>
<th>Career</th>
<th>No. of cases</th>
</tr>
</thead>
<tbody>
<tr>
<td>F. 310</td>
<td>Biologist, zoologist, botanist, etc.</td>
<td></td>
<td>32</td>
</tr>
<tr>
<td>314</td>
<td>Specialist in fish, wildlife, forestry, conservation, etc.</td>
<td></td>
<td>49</td>
</tr>
<tr>
<td>313**</td>
<td>Specialist in agricultural science</td>
<td></td>
<td>35</td>
</tr>
<tr>
<td>316**</td>
<td>Microbiologist</td>
<td></td>
<td>15</td>
</tr>
<tr>
<td>317**</td>
<td>Biochemist</td>
<td></td>
<td>9</td>
</tr>
<tr>
<td>322**</td>
<td>Surgeon</td>
<td></td>
<td>14</td>
</tr>
<tr>
<td>G. 329</td>
<td>Physician (NEC)</td>
<td></td>
<td>229</td>
</tr>
<tr>
<td>332</td>
<td>Dentist</td>
<td></td>
<td>97</td>
</tr>
<tr>
<td>360*</td>
<td>Psychologist</td>
<td></td>
<td>106</td>
</tr>
<tr>
<td>462*</td>
<td>College professor: Science</td>
<td></td>
<td>68</td>
</tr>
<tr>
<td>H. 360*</td>
<td>Psychologist</td>
<td></td>
<td>106</td>
</tr>
<tr>
<td>393</td>
<td>Lawyer</td>
<td></td>
<td>393</td>
</tr>
<tr>
<td>463</td>
<td>College professor: social studies</td>
<td></td>
<td>85</td>
</tr>
<tr>
<td>460</td>
<td>College professor (NEC)</td>
<td></td>
<td>179</td>
</tr>
<tr>
<td>I. 464</td>
<td>College professor: English</td>
<td></td>
<td>57</td>
</tr>
<tr>
<td>J. 521</td>
<td>Clergyman</td>
<td></td>
<td>176</td>
</tr>
<tr>
<td>K. 400</td>
<td>Teaching (NEC)</td>
<td></td>
<td>423</td>
</tr>
<tr>
<td>420</td>
<td>High School Teacher (NEC)</td>
<td></td>
<td>69</td>
</tr>
<tr>
<td>423</td>
<td>High School Teacher: Social Studies</td>
<td></td>
<td>94</td>
</tr>
<tr>
<td>450</td>
<td>School administration (principal, etc.): except college</td>
<td></td>
<td>43</td>
</tr>
<tr>
<td>370</td>
<td>Social work</td>
<td></td>
<td>60</td>
</tr>
<tr>
<td>412</td>
<td>Teaching elementary school</td>
<td></td>
<td>45</td>
</tr>
<tr>
<td>L. 421</td>
<td>High school math teacher</td>
<td></td>
<td>62</td>
</tr>
<tr>
<td>M. 422</td>
<td>High school science teacher</td>
<td></td>
<td>50</td>
</tr>
<tr>
<td>N. 429</td>
<td>High school physical education teacher</td>
<td></td>
<td>57</td>
</tr>
<tr>
<td>O. 861</td>
<td>Printing trades</td>
<td></td>
<td>106</td>
</tr>
<tr>
<td>758</td>
<td>Computer operator, etc.</td>
<td></td>
<td>78</td>
</tr>
<tr>
<td>798</td>
<td>Miscellaneous administrative</td>
<td></td>
<td>38</td>
</tr>
<tr>
<td>797</td>
<td>Miscellaneous clerical</td>
<td></td>
<td>112</td>
</tr>
<tr>
<td>864</td>
<td>Draftsman</td>
<td></td>
<td>101</td>
</tr>
<tr>
<td>738</td>
<td>Supervisor (in a business)</td>
<td></td>
<td>87</td>
</tr>
<tr>
<td>661*</td>
<td>Police (public)</td>
<td></td>
<td>151</td>
</tr>
</tbody>
</table>

Continued on page 13
TABLE 3 (continued)

<table>
<thead>
<tr>
<th>Job Family Code</th>
<th>Career Description</th>
<th>No. of cases</th>
</tr>
</thead>
<tbody>
<tr>
<td>P.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>811</td>
<td>Electronic technician</td>
<td>212</td>
</tr>
<tr>
<td>125</td>
<td>Electronics (NEC)</td>
<td>156</td>
</tr>
<tr>
<td>299</td>
<td>Lab technicians, research assts., etc. (in physical sciences)</td>
<td>78</td>
</tr>
<tr>
<td>222*</td>
<td>Computer programe</td>
<td>121</td>
</tr>
<tr>
<td>347**</td>
<td>Medical and dental technicians; technicians in biol. and clinical sciences</td>
<td>16</td>
</tr>
<tr>
<td>348**</td>
<td>Medical and dental technologists</td>
<td>27</td>
</tr>
<tr>
<td>Q.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>102</td>
<td>Foreman (NEC)</td>
<td>49</td>
</tr>
<tr>
<td>810</td>
<td>Electrician (NEC)</td>
<td>168</td>
</tr>
<tr>
<td>812</td>
<td>Appliance repair</td>
<td>55</td>
</tr>
<tr>
<td>820</td>
<td>Mechanic (NEC)</td>
<td>116</td>
</tr>
<tr>
<td>821</td>
<td>Auto mechanic</td>
<td>74</td>
</tr>
<tr>
<td>822</td>
<td>Airplane mechanic</td>
<td>68</td>
</tr>
<tr>
<td>828</td>
<td>Machinist</td>
<td>176</td>
</tr>
<tr>
<td>832</td>
<td>Carpenter</td>
<td>55</td>
</tr>
<tr>
<td>833</td>
<td>Metal trades</td>
<td>92</td>
</tr>
<tr>
<td>834</td>
<td>Bricklayer, mason, roofer, printer, plasterer, etc.</td>
<td>62</td>
</tr>
<tr>
<td>835</td>
<td>Plumber, pipefitter</td>
<td>72</td>
</tr>
<tr>
<td>837</td>
<td>Misc. building and construction</td>
<td>71</td>
</tr>
<tr>
<td>899</td>
<td>General labor (unspecialized)</td>
<td>277</td>
</tr>
<tr>
<td>853</td>
<td>Auto, bus, and truck drivers</td>
<td>85</td>
</tr>
<tr>
<td>824**</td>
<td>Industrial machine repair</td>
<td>21</td>
</tr>
<tr>
<td>813**</td>
<td>Phone installation, repair, maintenance</td>
<td>39</td>
</tr>
<tr>
<td>838**</td>
<td>Mining, quarrying, well-drilling</td>
<td>16</td>
</tr>
<tr>
<td>836**</td>
<td>Operating earthmoving equipment; roadbuilding</td>
<td>47</td>
</tr>
<tr>
<td>R.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>631</td>
<td>Farm or ranch owner</td>
<td>128</td>
</tr>
<tr>
<td>639</td>
<td>Farming: other and miscellaneous</td>
<td>244</td>
</tr>
<tr>
<td>S.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>661*</td>
<td>Police (public)</td>
<td>151</td>
</tr>
<tr>
<td>666</td>
<td>Fireman</td>
<td>41</td>
</tr>
<tr>
<td>640</td>
<td>U.S. Armed Forces (rank unspecified)</td>
<td>54</td>
</tr>
<tr>
<td>641**</td>
<td>U.S. Armed Forces enlisted personnel</td>
<td>13</td>
</tr>
<tr>
<td>T.</td>
<td>(Undecided)</td>
<td>803</td>
</tr>
</tbody>
</table>

*Included in more than one group.

**Tentatively included in the group, on the basis of subjective decision, since the N is too small for conclusive empirical data.
### TABLE 4. Range of Inter-career $D^2$ Values within Job Families

(Based on principal components*** of 64 cognitive variables or 45 noncognitive variables.)

<table>
<thead>
<tr>
<th>Job Family</th>
<th>No. of Career Groups</th>
<th>No. of within-family $D^2$ values per matrix</th>
<th>Range of $D^2$ values (Formula 7)</th>
<th>Cognitive</th>
<th>Noncognitive</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>1</td>
<td>0</td>
<td>----</td>
<td>----</td>
<td>----</td>
</tr>
<tr>
<td>B**</td>
<td>16</td>
<td>120</td>
<td>.29-1.55</td>
<td>.18-1.82</td>
<td>Business and industry</td>
</tr>
<tr>
<td>B**</td>
<td>19</td>
<td>171</td>
<td>.29-3.80</td>
<td>.18-3.05</td>
<td></td>
</tr>
<tr>
<td>C</td>
<td>1</td>
<td>0</td>
<td>----</td>
<td>----</td>
<td></td>
</tr>
<tr>
<td>D*</td>
<td>7</td>
<td>21</td>
<td>.46-1.33</td>
<td>.40-1.29</td>
<td>Engineering and applied physical science</td>
</tr>
<tr>
<td>D**</td>
<td>12</td>
<td>66</td>
<td>.46-6.15</td>
<td>.40-4.06</td>
<td></td>
</tr>
<tr>
<td>E*</td>
<td>4</td>
<td>6</td>
<td>.97-1.56</td>
<td>.83-1.31</td>
<td>Math and physical science</td>
</tr>
<tr>
<td>E**</td>
<td>5</td>
<td>10</td>
<td>.97-2.59</td>
<td>.83-1.87</td>
<td></td>
</tr>
<tr>
<td>F*</td>
<td>2</td>
<td>1</td>
<td>1.57</td>
<td>1.27</td>
<td>Biological sciences</td>
</tr>
<tr>
<td>F**</td>
<td>6</td>
<td>15</td>
<td>1.57-6.09</td>
<td>1.27-6.49</td>
<td></td>
</tr>
<tr>
<td>G</td>
<td>4</td>
<td>6</td>
<td>.79-1.30</td>
<td>.77-1.69</td>
<td>People-oriented, scientific Social sciences</td>
</tr>
<tr>
<td>H</td>
<td>4</td>
<td>6</td>
<td>.54-.96</td>
<td>.63-.90</td>
<td></td>
</tr>
<tr>
<td>I</td>
<td>1</td>
<td>0</td>
<td>----</td>
<td>----</td>
<td></td>
</tr>
<tr>
<td>J</td>
<td>1</td>
<td>0</td>
<td>----</td>
<td>----</td>
<td></td>
</tr>
<tr>
<td>K</td>
<td>6</td>
<td>15</td>
<td>.47-1.81</td>
<td>.38-1.32</td>
<td>People-oriented, non-science</td>
</tr>
<tr>
<td>L</td>
<td>1</td>
<td>0</td>
<td>----</td>
<td>----</td>
<td></td>
</tr>
<tr>
<td>M</td>
<td>1</td>
<td>0</td>
<td>----</td>
<td>----</td>
<td></td>
</tr>
<tr>
<td>N</td>
<td>1</td>
<td>0</td>
<td>----</td>
<td>----</td>
<td></td>
</tr>
<tr>
<td>O</td>
<td>7</td>
<td>21</td>
<td>.62-1.39</td>
<td>.59-1.59</td>
<td>Miscellaneous skilled</td>
</tr>
<tr>
<td>P*</td>
<td>4</td>
<td>6</td>
<td>.34-1.07</td>
<td>.25-1.65</td>
<td>Technician</td>
</tr>
<tr>
<td>P**</td>
<td>6</td>
<td>15</td>
<td>.34-4.32</td>
<td>.25-2.76</td>
<td></td>
</tr>
<tr>
<td>Q*</td>
<td>14</td>
<td>91</td>
<td>.41-1.60</td>
<td>.29-1.41</td>
<td>Miscellaneous &quot;blue-collar&quot; jobs</td>
</tr>
<tr>
<td>Q**</td>
<td>18</td>
<td>153</td>
<td>.41-4.03</td>
<td>.29-3.06</td>
<td></td>
</tr>
<tr>
<td>R</td>
<td>2</td>
<td>1</td>
<td>.45</td>
<td>.20</td>
<td>Farming</td>
</tr>
<tr>
<td>S*</td>
<td>3</td>
<td>3</td>
<td>1.03-1.37</td>
<td>.84-1.64</td>
<td>Protective</td>
</tr>
<tr>
<td>S**</td>
<td>4</td>
<td>6</td>
<td>1.03-3.90</td>
<td>.84-3.94</td>
<td></td>
</tr>
<tr>
<td>T</td>
<td>1</td>
<td>0</td>
<td>----</td>
<td>----</td>
<td></td>
</tr>
</tbody>
</table>

*Excluding careers marked with double asterisk (***) in Table 3.

**Including careers marked with double asterisk (***) in Table 3.

***The principal components are based on scores of 14123 grade 12 boys who responded to the 5-year follow-up questionnaire. The information about careers was provided in the responses to that questionnaire.