Notes on the Significance of Biserial and Point Biserial Correlations

John C. Flanagan and Marion F. Shaycoft

In Five Years after High School many point biserial correlations are reported which, considered in terms of absolute value, are quite low. This has led some readers who are relatively unfamiliar with Project TALENT data to ask one or both of the following questions:

1. Are these low point biserials really significantly different from zero?
2. Even if they do differ significantly from 0, is the degree of relationship that they represent large enough to be important?

The answer to both these questions is yes. They will be discussed separately in this memo.

1. Testing the significance of a point biserial correlation

Two equivalent formulas for point biserial r are:

$$r_{pt.\ bis.} = \frac{\overline{X}_p - \overline{X}}{s} \cdot \sqrt{\frac{p}{q}} \qquad (1)$$

$$r_{pt.\ bis.} = \frac{\overline{X}_p - \overline{X}_q}{s} \cdot \sqrt{pq} \qquad (2)$$

In these formulas, p and q are the proportions of the total group (which may be considered a finite population) that are respectively members and nonmembers of the special subgroup (e.g., an occupational group) under consideration. $\overline{X}_p$ and $\overline{X}_q$ are the means of the members and nonmembers respectively, on the continuous variable. $\overline{X}$ and s are respectively the mean and standard deviation of the total group (members and nonmembers combined).

It is clear that point biserial r is significant if and only if the difference $\overline{X}_p - \overline{X}_q$ is significant--in other words if and only if the difference between these two means is sufficiently large that it is unlikely that the two groups (i.e., members and nonmembers) are random samples from the same hypothetical infinite population. Applying the familiar t test we can determine what values of $(\overline{X}_p - \overline{X}_q)/s$ are significant at the .05, .01, .001, and .0001 levels. Results are shown in Table 1 for N = 14,000 (and also for N = 7,000) for subgroups of different sizes. (Most of the point biserial correlations presented in Five Years after High School are based on about 14,000 cases; some are based on only half the group--i.e., about 7,000 cases.) Note that the values reported in Table 1 are $(\overline{X}_p - \overline{X})/s$ rather than $(\overline{X}_p - \overline{X}_q)/s$. After the significant values of the latter were determined they were converted to the more convenient $(\overline{X}_p - \overline{X})/s$ values, using the following formula:

$$\frac{\overline{X}_p - \overline{X}}{s} = q \cdot \frac{\overline{X}_p - \overline{X}_q}{s} \qquad (3)$$

Note also that $(\overline{X}_p - \overline{X})/s$ is merely $\overline{X}_p$ expressed in standard score form. Thus we shall designate it by $\overline{Z}_p$ in the remainder of this memo. We see from Table 1 that when the size of the p group is at least 100, a group mean would be more than .257σ away (in other words more than a quarter of a sigma away) from the overall mean by chance alone only once in 100 times. If we examine Table 6-5 in Five Years after High School, bearing in mind that normalized scores within a quarter of a sigma of the mean have a stanine of 5, we see that for every occupation there are some mean stanines other than 5, and that the number of such stanines is far larger than could possibly be accounted for by chance.

Table 2 shows, for various N's, the values of point biserial r corresponding to the same four levels of significance as in Table 1 (.05, .01, .001, .0001). The Table 2 values are obtained by multiplying the Table 1 values by $\sqrt{p/q}$, in accordance with formula 1. It is noteworthy that the critical values of point biserial r depend only on N. Neither p nor N_p affects the critical value at all. In other words when the Table 1 value for N = 14,000 and N_p = 4,000 is multiplied by the corresponding value of $\sqrt{p/q}$ the product is exactly the same as when the Table 1 value for N = 14,000 and N_p = 10, or 2, or even 1, is multiplied by its $\sqrt{p/q}$ value. The formula for $r'_{pt.\ bis.}$ (the point biserial r corresponding to a specified t value, representing the desired level of significance) turns out to be:

$$r'_{pt.\ bis.} = \frac{t}{\sqrt{N - 2 + t^2}} \qquad (4)$$

Comparison of the zero-order point biserials reported in Five Years after High School with the Table 2 values should make it clear that the vast majority of the reported correlations are significant.

Note that the Table 2 values are applicable to zero-order correlations only; for multiple point biserial correlations the formula would have to be modified to take into account the reduced number of degrees of freedom.

Table N-3 in Appendix N of Five Years after High School represents an entirely different approach to determining how high a point biserial r must be in order to be statistically significant. The formulas used are entirely different from those used for Tables 1 and 2 of this memo, and the underlying statistical logic is also very different. And yet the results of the two tables (Table 2 of this memo and Table N-3 of the report) are remarkably compatible. Besides the different formulas used for it, Table N-3 differs from Table 2 in two major respects.

1. It provides values for multiple correlations as well as for zero-order correlations. (Columns for which $n>1$ are for multiple correlations.)
2. It gives more conservative estimates because it determines not merely that the absolute value of the correlation is significantly higher than zero, but also that it is significantly higher than the value that would shrink to zero if the Wherry shrinkage formula were applied. The Table N-3 point biserials with a single asterisk are the ones corresponding to the correlation that would shrink to 0. The point biserials with 2 and 3 asterisks are the ones that differ significantly from the one-asterisk point biserial, at the .05 and .01 significance levels. If the one-asterisk value is subtracted from the two- or three-asterisk point biserial, the difference is virtually identical with the corresponding

Table 2 value significant at the .05 or .01 level. To verify this fact, which demonstrates the compatability of the two tables, it is suggested that the reader compare the Table N-3 column for N = 14,123, n = 1 with the N = 14,000 data of Table 2; and the Table N-3 columns for N = 7,017, n = 1 with the N = 7,000 data of Table 2. For further discussion of Table N-3 see the discussion starting at the bottom of page 7-17 of Five Years after High School and continuing on page 7-18.

2. Are the point biserials large enough to represent an important relationship?

Point biserial correlations between a dichotomous variable and a continuous normally distributed variable usually don't give a true representation of the degree of relationship between the two variables, since the maximum possible value of such a correlation is always below 1--how far below 1 depending on how far from a .50/.50 split the p:q split is. The third column of Table 3 shows the maximum point biserial r for various splits. To understand how point biserial correlations can be modified to give a truer estimate of the amount of relationship, we should consider the formula for biserial correlation, and the underlying meaning of that formula. The formula is:

$$r_{bis} = \frac{\overline{X}_p - \overline{X}}{s} \cdot \frac{p}{y} \qquad (5)$$

$$= \overline{z}_p \cdot \frac{p}{y} \qquad (6)$$

where y is the normal ordinate corresponding to the point of split between p and q.

The value of y/p (the reciprocal of part of the formula) is the mean deviation of the tail portion of a unit normal distribution. Thus if all cases in the sample are above all other cases in the population, $\overline{z}_p = y/p$ and therefore $r_{bis} = 1.00$.

Formula 1 (for point biserial correlation) may also be expressed as follows:

$$r_{pt.\ bis.} = \overline{Z}_p \sqrt{\frac{p}{q}} \qquad (7)$$

It will be noted that as p approaches 0 and q approaches 1.00, $r_{pt.\ bis.}$ approaches 0.00. Since the maximum value for $\overline{Z}_p$ is always y/p if the sample scores are distributed normally (note that this has nothing whatsoever to do with whether a normal distribution underlies the dichotomy), it is clear that the maximum value for $r_{pt.\ bis.} = \frac{y}{p}\sqrt{\frac{p}{q}} = \frac{y}{\sqrt{pq}}$.

To transform a point biserial correlation to a biserial correlation it is merely necessary to multiply the point biserial coefficient by $\sqrt{pq}/y$ which is precisely the same operation as dividing the point biserial correlation coefficient by the maximum value it can achieve.

Another way of describing the biserial correlation coefficient is that it is the ratio of the mean in the sample group to the mean for a sample which represented perfect prediction by the independent variable. Thus:

$$r_{bis} = \frac{\overline{Z}_p}{y/p} = \overline{Z}_p \cdot \frac{p}{y} \qquad (8)$$

When there are 10,000 cases in the total population, a sample of 25 will have a mean value of 3.11 if the correlation is perfect. The mean value which could be expected as often as one time in 10,000 samples is .78. Therefore the maximum expected value of the biserial by chance for a sample of 25 cases is .78/3.11 = .25, though of course the odds against getting a value that high by chance alone are enormous (about 10,000 to 1). For a sample N of 100, the corresponding "maximum expected chance value" of biserial r is .125, and for a sample N of 400 it is .0625.

Table 3 shows the biserial correlations corresponding to the point biserials of Table 2. These biserials have been obtained by multiplying the point biserials by $\sqrt{pq}/y$ and they are of course significantly different from 0 to exactly the same degree that the corresponding point biserials are. Of course the biserials that are reported in <u>Five Years after High School</u> are on the whole considerably higher than the critical values shown in Table 3. Inspection of the biserials in the report makes it clear that they are not only significant but that many of them are very sizable, representing an important degree of relationship. And, as the foregoing discussion has suggested, the biserial r's give more and better information about the degree of relationship between dichotomies of the type under consideration and normally distributed variables than do point biserial r's.

TABLE 1. Critical values of $\overline{Z}_p$* for various levels of significance

N	N_p	p	Value $\|\overline{Z}_p\|$ must exceed for significance at: .05 level	.01 level	.001 level	.0001 level
14,000	12,500	.8929	.006	.008	.010	.011
	7,000	.5000	.017	.022	.028	.033
	4,000	.2857	.026	.034	.044	.052
	2,500	.1786	.036	.047	.060	.070
	1,500	.1071	.048	.063	.080	.095
	1,000	.0714	.060	.078	.100	.119
	800	.0571	.067	.088	.113	.134
	600	.0429	.078	.103	.131	.155
	400	.0286	.097	.127	.162	.192
	200	.0143	.138	.181	.231	.273
	100	.0071	.195	.257	.328	.387
	50	.0036	.277	.364	.464	.549
	25	.0018	.392	.515	.657	.777
	10	.0007	.620	.814	1.040	1.229
	5	.0004	.876	1.152	1.471	1.739
	2	.0002	1.386	1.821	2.326	2.750
	1	.0001	1.960	2.575	3.289	3.889
7,000	1,500	.2143	.045	.059	.075	.089
	800	.1143	.065	.086	.109	.129
	600	.0857	.077	.101	.128	.152
	400	.0571	.095	.125	.160	.189
	200	.0286	.137	.179	.229	.271
	100	.0143	.195	.256	.326	.386
	50	.0071	.276	.363	.463	.548
	25	.0036	.391	.514	.657	.776

*$\overline{Z}_p$ is the mean of the p group, expressed in terms of standard score units.

TABLE 2. Critical values of zero-order point biserial r for various levels of significance

N	Value pt. bis. r must exceed* for significance at .05 level	.01 level	.001 level	.0001 level
14,000	.017	.022	.028	.033
7,000	.023	.031	.039	.046
400	.098	.128	.163	.191
200	.138	.180	.228	.266
125	.174	.226	.284	.331
62	.250	.325	.408	**
30	.361	.463	.570	**
20	.444	.561	.679	**
10	.632	.765	.872	**
5	.878	.959	.991	**

* Note that these values depend only on N (the total number of cases). Neither the proportion nor the number of cases in the selected subgroup (i.e., the "p group") has any effect at all on the significance of a point biserial correlation.

** Not computed; table of t not readily available for .0001 level of significance.

TABLE 3. Critical values of zero-order biserial r for various levels of significance

N	p	Max.* possible value of pt.bis.r	Value bis. r must exceed for significance at .05 level	.01 level	.001 level	.0001 level
14,000	.50	.798	.021	.027	.035	.041
	.20 (or .80)	.700	.024	.031	.040	.047
	.10 (or .90)	.585	.028	.037	.048	.056
	.05 (or .95)	.473	.035	.046	.059	.069
	.02 (or .98)	.346	.048	.063	.080	.095
	.01 (or .99)	.268	.062	.081	.104	.123
	.005 (or .995)	.205	.081	.106	.136	.160
	.002 (or .998)	.142	.117	.153	.196	.232
	.001 (or .999)	.107	.155	.204	.261	.309
7,000	.50	.798	.029	.038	.049	.058
	.20 (or .80)	.700	.033	.044	.056	.066
	.10 (or .90)	.585	.040	.053	.067	.079
	.05 (or .95)	.473	.049	.065	.083	.098
	.02 (or .98)	.346	.068	.089	.114	.134
	.01 (or .99)	.268	.087	.115	.147	.173
	.005 (or .995)	.205	.114	.150	.192	.227
	.002 (or .998)	.142	.165	.217	.277	.328
	.001 (or .999)	.107	.220	.289	.369	.436

---

* The maximum possible value of pt. bis. r = $\frac{y}{\sqrt{pq}}$ where y is the normal ordinate corresponding to the p:q split. This formula for the maximum requires the assumption that the continuous variable is normally distributed.

Formulas and tables to accompany:

"The Standard Point Biserial Correlation":
A Solution to the Biserial-versus-Point-Biserial Dilemma*

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

Notation for correlation coefficients

r_{bis} = biserial correlation coefficient

r (or $r_{pt\ bis}$) = point biserial correlation coefficient

r' (or $r'_{pt\ bis}$) = standard point biserial correlation coefficient

$$r_{pt\ bis} = \frac{\bar{X}_p - \bar{X}}{s} \sqrt{\frac{p}{q}} \qquad (1)$$

where X = continuous variable

N = total number of cases

N_p = number of cases in the smaller of the two groups

p = proportion of cases represented by N_p

$q = 1 - p$ (2)

$\bar{X}$ = mean of total distribution on variable X

$\bar{X}_p$ = mean of the N_p cases on variable X

s = sample standard deviation for total group on variable X

$$s = \sqrt{\frac{\Sigma (X - \bar{X})^2}{N}} \qquad (3)$$

(Note that divisor is N, not N-1.)

* Paper presented at American Educational Research Association Convention, in New York, 7 February 1971.

The "standard point biserial correlation coefficient" $\left(r'_{pt\ bis}\right)$ is defined as what point biserial r would equal if the dichotomy were exactly a 50-50 split.

$$r'_{pt\ bis} = \frac{r}{\sqrt{(1 - 2p)\left[2p\left(\frac{s_p^2}{s^2}\right) + r^2\right] + 2p}} \qquad (4)$$

r = original point biserial correlation
(= $r_{pt\ bis}$)

s_p = sample standard deviation for the N_p cases on variable X

$$s_p = \sqrt{\frac{\Sigma(X - \bar{X}_p)^2}{N_p}} \qquad (5)$$

(Note that the divisor is N, not N-1.)

$r'_{pt\ bis}$ = standard point biserial correlation

All other notation is the same as for $r_{pt\ bis}$.

If $\frac{s_p^2}{s^2}$ (needed for formula 5) is unknown, it may be estimated from the following formula provided the same assumptions required for biserial correlation hold:

$$\frac{s_p^2}{s^2} = 1 + r^2\left[\frac{yz}{p} - \left(\frac{y}{p}\right)^2\right] \qquad (6)$$

y = normal ordinate corresponding to point of split between p and q.

z = standard score corresponding in a normal distribution to the point of split between p and q. (In other words z is the abscissa corresponding to the ordinate y.)

All other notation is the same as above.

$$r_{bis} = r\,\frac{\sqrt{pq}}{y} \qquad (7)$$

TABLE 1. Relation between biserial and point biserial correlations, and various associated statistics*

(1)	(2)	(3) Location of split between p and q	(4) Means on dichotomous variable: For larger group	(5) Means on dichotomous variable: For smaller group	(6) Ratio of biserial to pt. bis. r	(7) Ratio of pt. bis. to bis. r	(8) Maximum pt. bis r**
p	q	z	y/q	y/p	$\frac{\sqrt{pq}}{y}$	$\frac{y}{\sqrt{pq}}$	$\frac{y}{\sqrt{pq}}$
.50	.50	.00	.798	.798	1.25	.798	.798
.20	.80	.84	.350	1.400	1.43	.700	.700
.10	.90	1.28	.195	1.755	1.71	.585	.585
.05	.95	1.64	.109	2.063	2.11	.473	.473
.02	.98	2.05	.049	2.421	2.89	.346	.346
.01	.99	2.33	.027	2.665	3.73	.268	.268
.005	.995	2.58	.015	2.892	4.88	.205	.205
.002	.998	2.88	.006	3.170	7.05	.142	.142
.001	.999	3.09	.003	3.367	9.39	.107	.107
.0005	.9995	3.29	.0018	3.554	12.58	.079	.079
.0002	.9998	3.54	.0008	3.789	18.66	.054	.054
.0001	.9999	3.72	.0004	3.958	25.26	.040	.040

Notation

p = proportion of cases in smaller group
$q = 1 - p$ = proportion of cases in larger group
z = point of split between p and q, expressed in standard deviation units
y = normal ordinate corresponding to z

*The basic assumptions underlying columns 3-8 in this table are (1) that the continuous variable is normally distributed, (2) that the dichotomous variable is an artificial dichotomy, i.e. a dichotomy which has a hypothetical continuous variable underlying it, and (3) that this hypothetical continuous variable is normally distributed.

**The value in column 8, which is the same as that in column 7, is the upper limit of a point biserial correlation assuming that the continuous variable is normally distributed.

TABLE 2. Conservative estimates of minimum values of zero-order and multiple correlation coefficients that are significant for samples of various sizes with various numbers of independent variables and various p:q splits

			N = 100		N = 1000			N = 10000			N=7017
			n = 1	n = 10	n = 1	n = 10	n = 100	n = 1	n = 10	n = 100	n = 1****
R''^2 or r''^2 (See Note 1)			.01010	.10101	.00100	.01001	.10010	.00010	.00100	.01000	.00014
R'' or r''			±.101	.318	±.032	.100	.316	±.010	.032	.100	±.012
R''' or r''' (See Note 2)			.342***	.512***	.113***	.180***	.392***	.036***	.057***	.126***	.043***
			.290**	.477**	.093**	.161**	.372**	.030**	.051**	.120**	.035**
POINT BISERIAL CORRELATION (See Note 3)											
	p	q									
	.50	.50	.322***	.474***	.106***	.160***	.334***	.038***	.051***	.105***	.040***
			.269**	.428**	.087**	.141**	.314**	.028**	.045**	.099**	.033**
			.080*	.247*	.025*	.080*	.246*	.008*	.025*	.080*	.010*
	.20	.80	.322***	.451***	.103***	.151***	.301***	.033***	.048***	.096***	.039***
			.262**	.405**	.084**	.132**	.283**	.027**	.042**	.090**	.032**
			.070*	.218*	.022*	.070*	.217*	.007*	.022*	.070*	.008*
	.10	.90	.310***	.423***	.100***	.140***	.264***	.032***	.044***	.084***	.038***
			.252**	.377**	.080**	.120**	.246**	.025**	.038**	.078**	.030**
			.059*	.184*	.019*	.058*	.183*	.006*	.019*	.058*	.007*
	.05	.95	.299***	.396***	.096***	.129***	.232***	.030***	.041***	.073***	.036***
			.240**	.340**	.077**	.109**	.213**	.024**	.035**	.067**	.029**
			.048*	.149*	.015*	.047*	.149*	.005*	.015*	.047*	.006*
	.02	.98	.287***	.364***	.093***	.117***	.193***	.029***	.037***	.060***	.035***
			.230**	.308**	.073**	.097**	.173**	.023**	.031**	.054**	.028**
			.035*	.110*	.011*	.035*	.109*	.003*	.011*	.035*	.004*
	.01	.99	.280***	.340***	.090***	.108***	.169***	.028***	.034***	.053***	.034***
			.222**	.285**	.071**	.089**	.149**	.022**	.028**	.046**	.027**
			.027*	.085*	.008*	.027*	.085*	.002*	.008*	.027*	.003*
	.005	.995	.274***	.330***	.088***	.102***	.151***	.028***	.032***	.046***	.033***
			.216**	.267**	.069**	.083**	.130**	.022**	.026**	.040**	.026**
			.021*	.065*	.006*	.021*	.065*	.002*	.006*	.021*	.002*
	.002	.998	.269***	.310***	.086***	.096***	.130***	.027***	.030***	.132***	.032***
			.210**	.248**	.067**	.077**	.110**	.021**	.024**	.034**	.025**
			.014*	.045*	.004*	.014*	.044*	.001*	.004*	.014*	.002*

TABLE 2 continued

Miscellaneous notation

N = no. of cases
n = no. of independent variables
p = proportion of cases in criterion group
q = 1 - p
y = normal ordinate corresponding to the split between p and q.

Note 1. R''^2 or r''^2 is value of R^2 or r^2 which, when Wherry shrinkage formula (Wherry, 1939) is applied, shrinks to zero.

$$R'' = \sqrt{\frac{n}{N-1}}$$

Note 2. R''' (or r''') is the value a correlation coefficient must exceed in order to differ significantly from R'' (or r''), using two-tail test. (Fisher's z was used.)

Note 3. The point biserial correlations followed by a single asterisk are the values corresponding to R'' (or r'') when the criterion value (group membership) is assumed to be a normally distributed variable with an arbitrary dichotomy imposed on it. Similarly, the point biserial correlations followed by a double (or triple) asterisk are the lowest values that differ significantly at the .05 (or .01) level from the point biserial corresponding to R'' or r''.

$$R''_{pt\ bis} = \frac{y}{\sqrt{pq}} R''$$

*$R_{pt\ bis}$ corresponding to R''

**Significant at .05 level (2-tail test)
***Significant at .01 level (2-tail test)

$$z'''_k = z'' + \frac{k}{\sqrt{N-n-2}}$$

where z'' = Fisher's z corresponding to R'' or to $R''_{pt\ bis}$.
z''' = Fisher's z corresponding to R''' or to $R'''_{pt\ bis}$.
k = critical ratio = $\frac{z''' - z''}{\sigma_{z''}}$

****This column is included solely for use in conjunction with Table 4 (to determine the significance of the correlations in that table).

TABLE 3. Corresponding values of point biserial r, biserial r, and standard point biserial r, for various values of p and various standard deviation ratios

Point Bis r		s_p/s	Biserial r and standard point biserial r´ p = .01	p = .02	p = .05	p = .10	p = .20	p = .50
.60	r_{bis}	→	2.240	1.735	1.268	1.026	.857	.752
	r´	*	.971**	.945**	.878**	.797**	.697**	.600**
		.8	.967	.937	.865	.781	.684	.600
		.6	.974	.949	.888	.812	.716	.600
		.4	.979	.959	.906	.837	.742	.600
		.2	.982	.964	.918	.853	.759	.600
		.0	.983	.966	.921	.859	.764	.600
.40	r_{bis}	→	1.493	1.157	.845	.684	.572	.501
	r´	*	.915**	.849**	.717**	.597**	.485**	.400**
		.8	.919	.856	.728	.610	.496	.400
		.6	.933	.878	.761	.644	.524	.400
		.4	.943	.895	.787	.673	.547	.400
		.2	.949	.906	.804	.692	.563	.400
		.0	.951	.909	.810	.698	.568	.400
.20	r_{bis}	→	.747	.578	.423	.342	.286	.251
	r´	*	.719**	.592**	.426**	.324**	.248**	.200**
		.8	.747	.623	.455	.346	.263	.200
		.6	.777	.659	.487	.372	.280	.200
		.4	.801	.688	.516	.394	.294	.200
		.2	.817	.707	.535	.410	.304	.200
		.0	.822	.714	.542	.415	.307	.200
.10	r_{bis}	→	.373	.289	.211	.171	.143	.125
	r´	*	.452**	.338**	.225**	.165**	.125**	.100**
		.8	.486	.367	.245	.179	.134	.100
		.6	.521	.397	.266	.194	.143	.100
		.4	.551	.424	.285	.207	.150	.100
		.2	.572	.442	.298	.216	.155	.100
		.0	.579	.449	.303	.219	.157	.100
.05	r_{bis}	→	.187	.145	.106	.085	.071	.063
	r´	*	.244**	.176**	.114**	.083**	.062**	.050**
		.8	.267	.193	.125	.091	.067	.050
		.6	.291	.211	.136	.098	.072	.050
		.4	.313	.227	.146	.105	.075	.050
		.2	.328	.239	.154	.110	.078	.050
		.0	.334	.243	.156	.111	.079	.050
.02	r_{bis}	→	.075	.058	.042	.034	.029	.025
	r´	*	.100**	.071**	.046**	.033**	.025**	.020**
		.8	.110	.078	.050	.036	.027	.020
		.6	.121	.086	.055	.039	.029	.020
		.4	.130	.093	.059	.042	.030	.020
		.2	.137	.098	.062	.044	.031	.020
		.0	.140	.100	.063	.045	.032	.020

* Value obtained by making the same assumptions that are required for biserial r.
** r´ obtained by using the estimated ratio of standard deviations.

TABLE 4. Correlations between "career plan clusters" and predictive indexes

Based on 5-year follow-up of 12th grade boys in TALENT sample (N = 7017)

Code	Career plan****	N_p	Correlation with linear composite***** r **	r_{bis} **	r'*** **	Correlation with propinquity index***** r	r'
A	Airplane Pilot	49	.045	.192	.260	.040**	.271**
B	Business and Industry	1468	.240	.339	.290	.148**	.191**
C	Architect	29	.063	.332	.443	.052**	.458**
D	Engineering and applied phys. sci.	436	.334	.658	.589	.196**	.450**
E	Quantitatively oriented profs.	103	.181	.585	.607	.105**	.487**
F	Biological sciences	75	.112	.407	.480	.036**	.208**
G	"People-oriented" sci. professions	215	.263	.656	.620	.148**	.486**
H	Professions in social sciences	327	.269	.582	.532	.148**	.404**
I	College professor: English	28	.109	.582	.656	.058**	.528**
J	Clergyman	88	.182	.624	.639	.069**	.348**
K	Teaching, etc. (non-science)	365	.166	.347	.355	.083**	.205**
L	H.S. Math teacher	31	.058	.298	.402	.044**	.373**
M	H.S. Science teacher	25	.026	.143	.211	.034**	.321**
N	H.S. Phys. Ed. teacher	28	.054	.289	.395	.034**	.297**
O	Misc. skilled occupations	355	.108	.228	.241	.102**	.259**
P	Technician	244	.130	.309	.336	.068**	.190**
Q	Misc. "blue collar" jobs	629	.316	.557	.501	.211**	.405**
R	Farming	186	.300	.785	.698	.140**	.468**
S	Protective	53	.045	.189	.254	.034**	.205**
Y	(Undecided)	401	.093	.188	.197	.037**	.083**
Z	(All other cases: Misc.)	1882	.149	.201	.168	.034*	.039*
	TOTAL	7017					

Determining the significance level of correlation coefficients:

Point biserial r values are compared with the last column of Table 2 to determine their significance. Biserial r or point biserial r´ has the same significance level as the point biserial r from which it was derived.

* Significant at .05 level.
** Significant at .01 level. A double asterisk at the top of the column applies to all correlations in the column.
*** Formula 6 was used, in conjunction with Formula 4.
**** NEC stands for the "Not Elsewhere Classified".
***** Both the linear composite and the propinquity index are composites of scores on variables in the TALENT battery.

REFERENCES

Shaycoft, Marion F. A New Multivariate Index for Use in Educational Planning. (Paper presented at the American Psychological Association Convention, September 2, 1969).

Shaycoft, Marion F. Propinquity Analysis: A New Guidance Tool. (Paper presented at the Annual Meeting of the National Council on Measurement in Education, March 1970).

Wherry, R. J. "A New Formula for Predicting the Shrinkage of the Coefficient of Multiple Correlation." Annals of Mathematical Statistics, 2: 440-457 (1939).

SELECTED REFERENCES ON PROJECT TALENT

[1] Flanagan, J.C., Dailey, J.T., Shaycoft, Marion F., Gorham, W.A., Orr, D.B., & Goldberg, I. Design for a Study of American Youth, 248 pp. Boston: Houghton Mifflin, 1962.

[2] Flanagan, J.C., Dailey, J.T., Shaycoft, Marion F., Orr, D.B., & Goldberg, I. Studies of the American High School. (Cooperative Research Project No. 226) Project TALENT Office, 1962. 375 pp.

[3] Flanagan, J.C., Davis, F.B., Dailey, J.T., Shaycoft, Marion F., Orr, D.B., Goldberg, I., & Neyman, C.A., Jr. The American High School Student. (Cooperative Research Project No. 635) Project TALENT Office, 1964. 738 pp.

[4] Flanagan, J.C., Cooley, W.W., Lohnes, P.R., Schoenfeldt, L.F., Holdeman, R.W., Combs, Janet, & Becker, Susan. Project TALENT One-Year Follow-Up Studies. (Cooperative Research Project No. 2333) Project TALENT Office, 1966. 348 pp.

[5] Shaycoft, Marion F., Dailey, J.T., Orr, D.B. Neyman, C.A., Jr., & Sherman, S.E. Studies of a Complete Age Group--Age 15. (Cooperative Research Project No. 566) Project TALENT Office, 1963. 370 pp.

[6] Shaycoft, Marion F. The High School Years: Growth in Cognitive Skills. (Cooperative Research Project No. 3051) Project TALENT Office, 1967. 376 pp.