The Elements of Mental Tests

Second Edition

John D. Mayer

MOMENTUM PRESS HEALTH

MOMENTUM PRESS, LLC, NEW YORK

The Elements of Mental Tests, Second Edition

First published in 2015 by the author
Durham, NH

ISBN-13: 978-1-60650-759-9 (paperback)
ISBN-13: 978-1-60650-760-5 (e-book)

Momentum Press Psychology Collection

Cover and interior design by Exeter Premedia Services Private Ltd., Chennai, India

First edition: 2015
Second edition: 2016

10 9 8 7 6 5 4 3 2 1

Printed in the United States of America.

average (GPA), the highest correlations are found among grade school children—who have a broad range of abilities. The correlation between IQ and GPA drops in college from high school because only the top two-thirds or one-half of all students go on to college—which results in a more restricted range of ability. McNemar (1969) offers a correction to the correlation coefficient due to such restriction of range.

The Strength of a Correlation Coefficient

If a test does predict a real-life criterion, what is the strength of the prediction and its practical application? For example, the SATs predict first-year college GPA at various colleges at about an $r = .40$ level. What does that level of correlation mean, exactly? There are at least four ways to view the level of a correlation coefficient.

Using the Level of the Correlation Itself

The most obvious way to describe the strength of a correlation is simply to report the correlation coefficient itself. Recall that the closer the correlation is to 1.0 (or −1.0), the stronger the relationship; conversely, the closer the correlation is to 0, the weaker the relationship. If the correlation is close to or exactly 1.0 or −1.0, the interpretation is fairly obvious: The relationship is perfect, or just about so. If the correlation is 0, the interpretation is also easy: there is no relationship. It is the in-between values that pose a greater issue of interpretation.

Intermediate values of the correlation coefficient neither correspond to anything (like percentages) that are easy to understand nor increase or decrease in equal intervals. Consider the $r = .40$ correlation between SAT scores and first-year college GPA. It is not possible to say that $r = .40$ means that 40 percent of the people are classified well. Nor can we say that a .80 correlation is twice as strong as a .40 correlation. So, other ways of communicating the intermediate values of correlations are used.

The Squared Correlation as an Index of Variance Explained

One approach to characterizing the strength of the relationship is to use the squared correlation as an index of the variance explained in the criterion. Statisticians agree that a correlation between two variables X and Y means that X can be said to explain a specific amount of the variance of the other variable, Y (and vice versa). The exact percentage of the variance explained in one variable by the other is given by taking the correlation coefficient, squaring it, and multiplying it by 100.

So, a correlation of .30 between variables X and Y means that variable X explains 9 percent of the variance (.30 squared times 100, or .09 × 100). That 9 percent is both the amount of variance that X explains of Y and that Y explains of X (explaining in either direction always is equivalent).

Given the SAT's $r = .40$ prediction of first-year-college GPA, the SAT accounts for 16 percent of the variance of that GPA (.40 squared is .16, and multiplied by 100 is 16 percent). This still leaves the question of what it means, exactly, to predict a percent of the variance.

Here is one way to think about it: First, each person's score varies from the mean to a certain degree. In variance terms, the average person varies from the mean in units that have been squared, summed, averaged, and then square-rooted, which results in the standard deviations.

If we think of an individual's deviation from the mean in terms of these squared, summed, averaged, and then square-rooted units, we can say that 16 percent of such a deviation is accounted for by the SAT. Imagine a college where the average first-year GPA was a 3.0, with a standard deviation of .6, and a variance of .36. If Jane obtained a first-year GPA of 3.72, then her deviation from the mean GPA is .72, or two times the variance. On average, we could estimate, 16 percent of that .72 upward shift, or about .115, can be explained by her SAT score.

Restated in general terms, the percentage of variance a test explains of a criterion can be made more concrete by thinking about individual cases as examples. In an individual's case, the percent of variance of the criterion explained can be thought of as the percent of the person's deviation from the mean on the criterion, as accounted for by the predictor variable. This accounting is only *on average* because the predictor variable may be more or less potent for a specific individual or subgroups of individuals.

If you find this a bit hard to follow, you are not alone ... which brings me to an alternative way to interpret the strength of a correlation coefficient: by graphing it.

Graphing the Correlation Coefficient

Another way to assess the strength of a correlation coefficient is to graph it and look at the points according to where they fall on the graph. If the points form a group in the shape of a line (or close to it), that is a strong relationship. To the degree the points vary from the line, the correlation is less strong. Upward sloping lines (from lower left to upper right) represent positive correlations; downward sloping lines (from upper left to lower right) represent negative correlations.

Graphing is a generally good approach to representing the strength of correlations. One drawback is that for the range of many real-life predictions—which is to say, between $r = .02$ and .40 (e.g., Mayer, 2001), such graphs often look more or less like a large jumble of points—and yet the relationships may nonetheless be very important. Graphing provides a good index (and is useful in explaining how a correlation coefficient works), but leaves open the question of how to discuss the practicality of the prediction, particularly at lower levels of correlations.

The Binomial Effect Size Display

The final approach I'll describe is referred to as a Binomial Effect Size Display (BESD, shown in Table 7.5). In the testing context, the BESD table divides people and their outcomes into those who scored high and low on the test (the rows), and those who were high and low on the outcome measure (the columns).

In Table 7.5, scores on a hypothetical *First Year Predictor Test* are divided (at the mean) into high- and low-scoring groups. The high test score row indicates the above average scorers on the First Year Test. The row beneath it indicates the below average scorers.

Table 7.5 A Binomial Effect Size Display Table for r = .00

		First-year college GPA		
		Low	High	Total
Scores on the First-Year Predictor Test	High	50	50	100
	Low	50	50	100
	Total	100	100	200

By convention, the BESD describes a sample of 200 people overall: 100 high scorers and 100 low scorers on the test. Because the sample is assumed to be perfectly normal, there always are equal numbers of participants above and below the test mean.

The columns of the BESD represent performance below or above a criterion, *first year college GPA* in this example. The participants are divided into those who were *low* and those who were *high* on the criterion.

Table 7.5 represents a baseline case: a BESD illustrating a zero correlation ($r = 0$) between the test and performance at the criterion. Because there is no relationship between the test and the criterion, participants are evenly distributed across the four cells. That is, people with high test scores were evenly divided between those who had high and low first-year college grades. Similarly, low test-scorers also were evenly divided. In this case, the test would not be useful as a tool to select people.

As the correlation rises from zero to a higher value, the cells in a BESD table change in a systematic way. If the correlation between the scores on the First-Year Predictor Test and first-year college grades was $r = .10$, the table would change slightly, with more high scorers showing high grades, and low scorers showing slightly lower grades (Table 7.6).

Table 7.6 A Binomial Effect Size Display Table for r = .10

		First-year college GPA		
		Low	**High**	**Total**
Scores on the First-Year Predictor Test	High	45	55	100
	Low	55	45	100
	Total	100	100	200

To adjust the table to reflect the correlation, these steps are employed:

1. Draw the standard BESD (already depicted in Table 7.5).
2. Find the correlation between the test and the criterion.
3. Divide the correlation in half.
4. Multiply it by 100.
5. Add and subtract the result from the middle four cells of the standard zero-correlation-based BESD, such that (assuming the correlation is

positive) there are more people who scored high on the test who succeeded and more people who scored low on the test who failed.

In the example just given, where $r = .10$, the .10 correlation is divided in half to yield .05. Next, .05 is multiplied by 100 to yield 5. In the final step we add 5 to the high and high and low and low cells, and subtract 5 from the high and low and low and high cells to yield the result.

This means that, of 100 high scorers on the First-Year Predictor Test, 55 will be judged as performing high on first-year grades, whereas only 45 of the low scorers on the test will be judged similarly.

Table 7.7 shows what happens for a correlation of $r = .40$ between scores on the First-Year Predictor Test and actual first-year GPAs at some colleges. The middle four cells are 30, 70, 70, and 30.

I obtained the numbers shown in Table 7.7 by taking the .40 correlation (for SATs), dividing it in half to get .20, multiplying it by 100 to get 20, and adding the resulting value (20) to the high and high and low and low cells (which started at 50) to get 70, and subtracting 20 from the high and low, and low and high cells (also starting at 50), to get 30. As mentioned earlier, this is near the actual level of prediction for the SAT.

A college that employed the SAT would be engaging in a far more successful selection process than one that used no predictive measures. If the college could admit only those students who scored high (i.e., above average) on the SAT, then 70 students of the incoming class would score above (the present) average, and only 30 below average, relative to a general sample (across all comparable colleges). This would allow more talented students to take advantage of the education the college provides, and protect other students from unnecessary failure.

Table 7.7 A Binomial Effect Size Display Table for r = .40

		First-year college GPA		
		Low	High	Total
Scores on the First-Year Predictor Test	High	30	70	100
	Low	70	30	100
	Total	100	100	200

The Binomial Effect Size Display table, relative to other approaches, allows for a more practical, pragmatic approach to explaining correlation by estimating what would happen in a selection process, if a test correlated with a criterion at a particular level.

Correlations Are Central to Psychometric Theory

Correlations of the sort described in this chapter play a central role in psychometric theory. For example, psychometric theory begins with an examination of the test score itself: how it can be divided into parts and what each part means. Understanding test scores, their parts, and the relationships among those parts involves thinking about the correlations among the parts. How it all works begins to unfold in the next chapter, in which psychometric theory is described in greater detail.