

Eclectic Research and Construct Validation

J. Scott Armstrong
The Wharton School, University of Pennsylvania

Consider the following situation: You have a fixed budget and would like to measure causal relationships in a study involving buyer behavior. How would you go about allocating the budget for this study?

This paper outlines two possible research strategies – intensive research and eclectic research. Each strategy utilizes the budget in a different manner. The intensive approach involves allocating the budget to a single study, and the eclectic approach divides the budget among a series of smaller-scale studies that differ markedly from one another.

Intensive research is called for when problems of reliability are of utmost concern; eclectic research is called for when problems of construct validity are paramount. Since we believe that problems of construct validity deserve more attention than they currently receive for problems in non-experimental research, we advocate stronger emphasis on eclectic research.

Reliability and Validity: Some Working Definitions

It is important to differentiate the concept of construct validity from that of reliability. Construct validity asks essentially whether the operational measure of a relationship measures what it purports to measure. Reliability, on the other hand, asks whether the operational measure produces similar results when the measurement is replicated.¹ In practice, the distinction is not always easy to make. One way of making this distinction is adapted from D. T. Campbell and Fiske (1959).² They expanded on a definition that apparently drew from Thurstone (1937). Thus *reliability* is seen as the agreement among two or more operational measures of a relationship through *approaches that are as similar as possible*. In other words, reliability asks for replication when no relevant factors are altered. *Construct validity* is seen as the agreement between two or more operational measures of a relationship through approaches that are *as different as possible*.

These concepts are illustrated in Figure 1. Approaches x_1 and x_2 are similar. Agreement between these approaches would be termed as evidence of reliability. The same is true for the agreement between approaches x_1 and x_4 . Approaches x_1 and x_3 , are rather different, however, as are x_2 and x_3 , or x_2 and x_4 . The agreement among these "different" approaches is termed evidence of construct validity.

"Shotguns versus Rifles"

In the social sciences, it is usually difficult to devise realistic experiments. As a result, there is a heavy dependence on the use of nonexperimental data. Results from nonexperimental data are, however, subject to interpretation by any one of a large number of hypotheses. In terms of Figure 1, there is only a weak correspondence between any one estimate (e.g., x_1) and X , the true relationship. If the correspondence between x_1 and X is weak, a highly reliable estimate of x_1 , may be of little value. This could yield a highly reliable estimate of the *wrong* value.

¹ See, for example, Selltitz, Jahoda, Deutsch, and Cook (1959) for a more complete discussion of reliability and construct validity.

² This paper is concerned with the measurement of *relationships* between variables, rather than the measurement of variables.

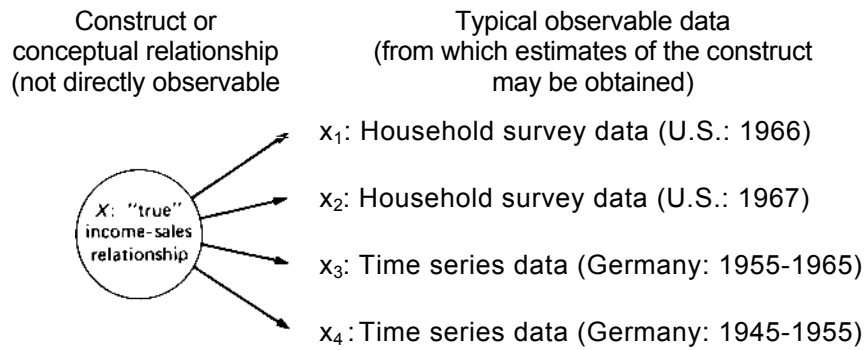


Figure 1. Measurement of a Hypothetical Construct

A familiar analogy may help to clarify the above. Assume that a hunter is about to shoot a bird, but he cannot see his target. However, he does know the general location of the target because he saw the bird go into a bush, can hear the bird, and can see branches moving. Since it is getting late, he decides to try to shoot the bird. The question he now faces is whether to use his rifle or his shotgun.

If he uses the rifle, there appears to be a very strong likelihood that he will miss his target altogether and the bird will fly away. If he were to hit the bird, of course, it is very likely that the shot would be fatal. If he loads the shotgun with buckshot, he is more likely to hit the bird, but the blast may not kill the prey. By wounding the bird, however, he will get further information on its location and can then move in to capture it.

The rifle is analogous to the use of intensive research. When aimed precisely, it does the job and with little waste. Poorly aimed, however, it does little good at all. The shotgun, like eclectic research, is likely to do some good if aimed in the *general* direction of the target. It may down the bird or at least slow it and allow for a second shot.

The Need for Eclectic Research in the Social Sciences

Social scientists often use the physical sciences as an ideal in considerations of research methodology, but this approach has shortcomings. In realistic experiments performed in the physical sciences, the correspondence between the operational measure and the construct can be rather close. The "target" is easier to see than it is in the social sciences. In this situation, the intensive approach (the rifle) is useful.

Seldom is the correspondence between operational measure and construct so clear-cut in the social sciences. The social sciences must utilize nonexperimental data, which are subject to a substantial amount of measurement error. Even economics, one of the more highly developed of the social sciences, must draw heavily on nonexperimental data of low reliability (e.g., see Morg; 1963).

The lack of correspondence between a construct and its measure implies that construct validation should receive much attention in the social sciences. Cronbach and Meehl (1955) presented what is now considered to be a classic on construct validity in psychological tests. In a somewhat different context, Payne (1951) pointed out that problems with (construct) validity seem to be far more important than problems with reliability in the wording of survey questions. More recently, Webb et al. (1966) based their arguments for the selection of "unobtrusive measures" on the notion of construct validity. Mayer and Brown (1965) attempt to direct attention toward "non-sampling errors" (related to our use of construct validity) in sample surveys. S. W. Cook and Selltitz (1964) and Curtis and Jackson (1962) discuss the assessment of construct validation in attitude measurement.

Problems with construct validation, then, have been noted in various ways by numerous social scientists. If we view eclectic research at one end of a continuum and intensive research at the other, this concern argues for an

increased emphasis on eclectic research in the social sciences.³ The same concern seems to be less evident in the physical sciences; in fact, the opposite approach was advocated by Bridgman (1927) : "If we have more than one set of operations, we have more than one concept, and strictly there should be a separate name to correspond to each different set of operations."

The State of the Art

Despite the concern of numerous writers about the importance of assessing construct validity, most published studies today are intensive rather than eclectic. My impression is that the reward structure in our research institutions is strongly oriented toward intensive research.

Intensive research calls for the specialist, who typically has expensive equipment, elaborate experiments, or large samples. His team is composed of specialists in the same field. The specialist is the man who understands and can argue for expensive equipment or large-scale surveys. He finds it easier to publish his results, since journals are usually organized by rather narrow technical areas, and editors appear to judge the reliability rather than the validity of a finding.

Eclectic research calls for the generalist, who has an interdisciplinary staff and spreads his budget to get by with makeshift equipment, crude experiments, or small samples.

Although the individual scientist may be motivated to emphasize the reliability of a relationship rather than its construct validity, the history of science depends quite heavily on the use of construct validity. A researcher will start with a relationship and demonstrate its applicability in a very limited context. Later researchers may show that the relationship has more generality by finding evidence for it in other situations. Alternatively, the additional research may serve to define the limits to which the relationship may be generalized; for example, Newtonian concepts of time and space are useful on earth but alternative theories prove more valuable elsewhere. Or perhaps the additional studies show that the original relationship was spurious.

Thus, while the demonstration of construct validity is regarded as something that is good for *science*, it does not seem to be regarded as a good strategy for the individual *scientist*. The emphasis among individual scientists is on demonstrating reliability rather than construct validity.

Velikovsky: A Critical Case for Eclectic Research

To demonstrate the value of eclectic research when non-experimental data are involved, a critical case is presented. The case is "critical" in that the theories generated by the eclectic research *differ drastically from those generated by intensive research*. In addition, this case is well documented, and more recent evidence allows us to compare the predictive validity of the theories generated by each approach.

The eclectic research was carried out by Immanuel Velikovsky (1950), whose theories demonstrated high construct validity but not high reliability. The existing theories, on the other hand, demonstrated high reliability and low construct validity. Velikovsky used historical papers and cultural artifacts to develop theories about the history of the earth. These documents, such as the Bible, are regarded by most scientists as being so unreliable that they are unworthy of use in the development of scientific theory. For example, any passage in the Bible is subject to almost as many interpretations as there are interpreters. Since somewhat different explanations are given by each interpreter, the source itself may be said to have low reliability.

³ D. T. Campbell and Fiske (1959) make a strong argument for an eclectic approach. They refer to this strategy as "multiple operationalism," "methodological triangulation," "convergent operationalism," "operational delineation," or "convergent validation."

Velikovsky's theories about the history of the earth are supported by evidence from the Bible, Greek mythology, ancient calendars, and geological findings. His approach uses unreliable data and encompasses many different scientific disciplines.

Velikovsky proposes that there were some catastrophic events in the history of the world. These events, such as the near-collision of Venus and Earth about 1500 B.C., had such a great effect on our planet that there was no need for highly reliable instruments to record the events. Everyone on earth was aware of what was happening. Velikovsky (1950) summarizes his argument:

If a phenomena has been similarly described by many peoples, we might suspect that a tale, originating with one people had spread around the world, and consequently there is no proof of the authenticity of the event related. But just because one and the same event is embodied in traditions that are very different indeed, its authenticity becomes highly probable, especially if the records of history, ancient charts, sundials, and the physical evidence of natural history testify to the same effect [p. 308].⁴

Velikovsky's use of extremely different approaches, or eclectic research, to test his theory did not impress the scientific community. In fact, prominent scientists mounted an active and somewhat successful campaign to suppress the publication of Velikovsky's theories. The stated reasons for this attack involved primarily the theoretician's "lack of scientific method" and his failure to be an "expert" in each of the fields affected by his theories. The history of this case is summarized in *The Velikovsky Affair* (de Grazia et al., 1967).

The power of Velikovsky's approach, which we call eclectic research, has been demonstrated over the past two decades. Velikovsky made a series of very different predictions based on his theory. Most of these predictions were labeled impossible in the light of existing theories. Recent evidence from space probes and other sources has confirmed many of the predictions, however, and no prediction has yet been proved incorrect. A summary of the cases that support the predictive validity of Velikovsky's theories is presented in de Grazia et al. (1967). The predictions that have received support include the high temperature of Venus, the existence of electromagnetic fields in the solar system, the possibility that the length of the day (on Earth) may change, the likelihood of radio noises from Jupiter, the presence of hydrocarbons in the atmosphere of Venus, the rather recent age of oil deposits in the Gulf of Mexico, the likelihood that signs of advanced human culture would be found in uninhabited areas of northeastern Siberia, and the age of the ancient civilization of Mexico.

In summary, Velikovsky made dramatic use of "different approaches" or eclectic research. That eclectic research led to accurate predictions of phenomena that had not been previously obvious testifies to the power of this strategy (as well as to the usefulness of nonexperimental data and, of course, to Velikovsky himself). The history of this case also demonstrates that eclectic research is not a well-accepted research strategy.

Eclectic Research in a Study of Buyer Behavior

Most of what we know about buyer behavior comes from studies of nonexperimental data, and the vast majority of these studies are based on a single approach; in short, eclectic research seems to receive little use.

In this study, two simple but specific relationships were examined. The first was the relationship between buying power and the purchase of cameras, which is referred to as the *income-sales relationship*. The second considers the relationship between the price of cameras and the purchase of cameras, which is referred to as the *price-sales relationship*. These two studies are presented to indicate an operational procedure for interpreting the results of eclectic research and to examine the importance of eclectic research in the study of buyer behavior.

⁴ See also the last three paragraphs of Chapter 12 in Velikovsky (1955) for a similar argument.

The Income-Sales Relationship

Four different approaches were used to measure the income elasticity for camera sales. The approaches differed because each used a different type of data:

Subjective Data. Previous studies of other durable goods (e.g., automobiles and refrigerators) were used to provide a subjective estimate of the income elasticity of camera sales. This estimate was made prior to the analysis of the camera sales data gathered from this study. The subjective analysis serves to force the researcher to specify what he thinks he knows about the relationships that he will study. The subjective estimates vary from one researcher to another, but the researcher must start with a well-defined hypothesis.

Cross-Sectional Data (Differences Among 12,000 Households in the United States, 1960-1961). Data on income and dollar purchase of cameras were analyzed for households, which were grouped in six income categories. In these data, the camera supply factors were relatively constant for all households; that is, the price at which cameras were offered for sale did not vary across households. The model that was used to explain differences among households accounted for age and education as well as the income of the household head.

Cross-Sectional Data (Differences Among Countries). Data from 19 countries were examined. The model for explaining differences among countries used not only a measure of buying power but also eight other variables, including the price of cameras; the population, and the climate. These data were based on averages from the period 1960-1965.

Longitudinal Data Across Countries. Data on the *changes* in sales rates from 1961 to 1965 were obtained for 21 countries. Differences among the various countries on the amount of change were explained by a model that incorporated measures of income, price, and population.

Each of the above types of data has its own advantages; for example, cross-sectional data over households are useful for assessing the income elasticity, since their incomes vary greatly. Rather than select the "best data," however, the emphasis was on using many types of data in studying the same parameter. Figure 2 presents the results of the four different approaches, as well as a measure of the confidence interval for each estimate. All estimates except the subjective estimate were obtained from multiple regression analyses.⁵ As indicated previously, an attempt was made to include all important variables in each regression model to measure the effect on sales due to income alone.⁶

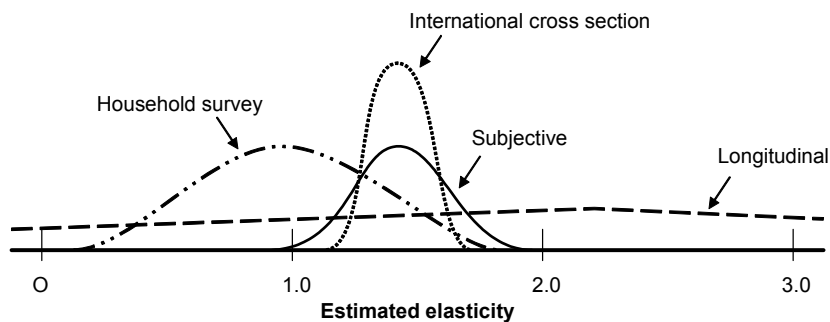


Figure 2. Confidence Intervals for Estimates of Relationship between Income and Sales

⁵ The details of these analyses, which are not important to the argument, can be found in Armstrong (1968). One defect of the eclectic approach, however, is that it would be rather difficult to describe a number of approaches in a single journal article.

⁶ See J. R. Meyer and Kuh (1957) for a discussion of the factors that may lead to differences between estimates obtained from time series and from cross-sectional data.

Each confidence interval was derived from the standard error of the estimate for the income coefficient, except of course for the subjective data. These standard errors correspond to measures of reliability.

Figure 2 shows a fair amount of agreement among the various estimates. For example, a "true" income elasticity of 1.3 would not be inconsistent with any of the estimated confidence intervals. Another way to examine the same data is presented in Table 1, which summarizes the measures of the confidence that one might have in each estimate of the income-sales relationship. The reliability estimates are again the standard errors around the income coefficients from each model. The construct validity estimate was obtained in a crude and simple fashion by taking the expected value (i.e., the regression coefficient for income) from each approach and estimating the standard deviation among these four values.

Table 1. Alternative Measure of Confidence for the Income-Sales Relationship

| Source of Estimate | Type of Measure | Standard Error |
|------------------------------------|--------------------|----------------|
| Subjective analysis | Reliability | 0.2 |
| Household survey data | Reliability | 0.2 |
| International cross section | Reliability | 0.1 |
| Longitudinal data over countries | Reliability | 1.0 |
| Comparisons among all of the above | Construct validity | 0.4* |

* Represents the standard deviation among the expected values from each of the four estimates.

The rather common practice of using the standard error (a reliability estimate) as *the* measure of confidence in a research report can be misleading if there is a high likelihood that the estimate lacks construct validity (i.e., is biased). The measure used for construct validity in Table 1 represents an attempt to assess the amount of bias in the estimates. Therefore, we suggest that that the research report contain *both* measures of reliability and measures of construct validity, the latter being of special importance.⁷

The results on the income-sales relationship provide evidence for high reliability and high construct validity. Further research in this area, if deemed necessary, might follow an intensive strategy (especially regarding the longitudinal data) or an eclectic strategy (to find still other types of data).

The Price-Sales Relationship

A similar analysis was used to examine the relationship between the price of cameras (an industry index) and the purchase rate of cameras. With one exception, which involved household survey data, the approaches used for the income-sales relationship were also useful for the price-sales relationship. (Since the price at which cameras were sold did not vary according to household, there was no way to obtain an estimate of price elasticity from these data.)

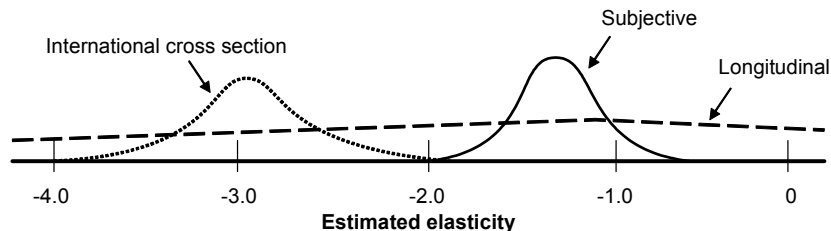


Figure 3. Confidence Intervals for Estimates of Relationship between Price and Sales

⁷ One may also desire an overall index of confidence, such as the mean square error. The *mean square error* makes an explicit attempt to include problems of both reliability (variance about the sample mean) and construct validity (bias).

The three different estimates for the price-sales relationship are presented in Figure 3. In contrast to the income-sales study, it is difficult to find a single estimate that would be consistent with each of the confidence intervals. Thus we would say that the estimates do not show evidence of high construct validity. The measures of confidence are also summarized in Table 2.

Table 2. Alternative Measure of Confidence for the Price-Sales Relationship

| Source of Estimate | Type of Measure | Standard Error |
|------------------------------------|--------------------|----------------|
| Subjective analysis | Reliability | 0.2 |
| International cross section | Reliability | 0.3 |
| Longitudinal data over countries | Reliability | 1.0 |
| Comparisons among all of the above | Construct validity | 1.7 |

Eclectic research led to conclusions quite different from those that would have been reached if intensive research had been used. The estimate of construct validity of the price-sales relationship indicates *much more* uncertainty about the true value than does any one of the reliability estimates. We are not confident about our ability to use the price elasticity estimate in generalizing to other situations (e.g., for forecasting changes in sales over time). Since the estimates themselves are biased, it is unlikely that the problems with the estimates of price elasticity could be resolved by improving the *reliability* of the existing estimates.

Since we are unable to *control* the sources of bias in nonexperimental data, we must find ways to compensate for their effects. Although a strategy of using more intensive research has some merit (e.g., to ensure that all relevant variables have been included in the international cross section), the most obvious course of action would be to employ more eclectic research. We should find still different approaches, which might include time series data from the United States, household expenditure data from Europe, consumer panel data from the United States, shopping experiments, or consumer intentions surveys.

Conclusions

The studies on income-sales and the price-sales relationships demonstrated the use of eclectic research. In particular, the agreement among the expected values of a given parameter as measured by different approaches was used to assess construct validity. The standard deviation among these values served as a summary statistic for construct validity.

These investigations, especially the study of the price-sales relationship, supported the notion that problems with construct validity are serious. Measures of reliability did not adequately reflect the uncertainty in the estimates. It is especially useful, then, if some measure of construct validity is provided in published studies on buyer behavior.

Summary

Intensive research is a strategy used to demonstrate that the findings from a study are reliable. *Eclectic research* is a strategy that tries to demonstrate that the findings from a study have construct validity. These strategies represent extreme points on a continuum, with the intensive side advocating that the research budget be spent on a single study and the eclectic side favoring a large variety of small-scale studies.

This paper argued that research in the social sciences, which must depend heavily on nonexperimental data, should place more emphasis on eclectic research than on intensive research.

The research of Velikovsky was mentioned to illustrate the power of eclectic research. This example provided a critical test for the eclectic approach, since the predictions generated by this approach differed greatly from those obtained by intensive research. Recent evidence suggests that the superiority of the eclectic approach was proved dramatically in this case.

Finally, eclectic research was applied to a study of buyer behavior. A simple operational procedure was suggested for using the results to assess construct validity. The results supported (1) the idea that problems with construct validity are of great importance, and (2) the argument that greater use should be made of eclectic research when nonexperimental data are involved. It was also suggested that published studies of relationships involving buyer behavior should present evidence of construct validity.

No references available.