

## To Pretest or Not to Pretest

Robert Pasnak\*

Department of Psychology, George Mason University, USA

Received: April 23, 2018; Published: June 08, 2018

\*Corresponding author: Robert Pasnak, George Mason University, Fairfax VA 22030



### Abstract

Pretests decrease the internal validity of experiments (i. e., render them incapable of completely proving what they are intended to prove). The loss of internal validity may be great in some cases, and little or no loss in others, but just how much has been lost cannot be proven. Pretests are sometimes credited with demonstrating that randomization has produced reasonably equivalent groups, but that is indexed by  $p$ , with equal accuracy whether or not pretests are employed. There are usually many sources of differences between subjects which could result in different outcomes on the dependent variable(s). The only valid index of whether an initial inequality of groups has produced a spurious result is  $p$ . This value specifies the probability that differences in the outcome for various levels of the independent variable(s) were produced by a failure of randomization to create equality in whatever pretests might measure and any other source of differences. Pretests can increase the power of experiments to detect small differences in outcomes, but only if pretest and posttest scores are strongly correlated. Conditions under which accepting the confound introduced by pretesting is most likely to be profitable are identified.

**Keywords:** Pretests; Internal Validity; External Validity; Power

### Introduction

There are three great goals for all research: internal validity, external validity, and power. An unfortunate misunderstanding that has become increasingly prevalent is the conception that pretesting increases internal validity. In fact, pretesting has the opposite effect; it reduces internal validity. What pretesting does increase is power if pretest and posttest scores are strongly correlated. Small effects may be detected if pretests are administered, although there are issues in statistical analysis after pretesting. This may reflect displacement of conventional methodology courses by courses in advanced correlational methods, which have great value in their own right and have engaged a generation of young scholars. This brief treatise is an effort to clarify the purpose and interpretation of pretests. To make it easier to follow, the discussion will be in terms of one independent and one dependent variable, although good research often involves more than one of each, and the measurements taken for the dependent variable will be defined as participants scores, although the same considerations hold for other measurements.

### Internal Validity

Internal validity expresses the extent to which we can be certain that, within the conditions of the experiment, the effects attributed to the independent variable are indeed due to that variable and not some other. Any variable whose effects might be mistaken for those of the independent variable is termed a "confound" or confounding variable. Confounds are sometimes inherent in the design of an experiment – as when pretests are employed – or are sometimes produced when an experiment is not conducted properly. When unconfounded designs are used and experiments are conducted

properly, there will be no confounds whose effects could be mistakenly attributed to the independent variable. A common cause of mistakes is the failure of random assignment of participants to levels of the independent variable to accomplish what it is supposed to accomplish – equality in all causal variables, known and unknown. The extent to which a failure of randomization could have produced the differences observed on the dependent variable is expressed by  $p$ . Pretests neither increase nor decrease the validity or accuracy of  $p$  as a measure of the failure of randomization to produce initial equality in the characteristics of participants in educational research. In essence a  $p$  of 0.2 means exactly the same thing whether or not there were pretests.

We can have equal confidence that there is a 2 per cent chance that the difference observed was not due to an initial inequality of the participants, whether there was or whether there was not a pretest. The only gain for a researcher who administered a pretest is that  $p$  may be smaller than it would otherwise be, because differences between individuals can be removed. Absent the pretest,  $p$  might have been .03 or even  $>.05$ . In the latter case, the effect of the independent variable would not be recognized, by conventional standards. However, the statistical analysis required when pretests are administered is less powerful than when they are not. This is because the pretest and posttest scores are correlated, and accounting for the correlation, which is a statistical necessity, sacrifices degrees of freedom, typically one for each pair of correlated scores, rendering the statistical analysis less powerful and therefore less able to detect effect of the independent variable. This is

not often understood, because the error term (measure of chance differences) is smaller. This decrease may or may not be offset by the loss of degrees of freedom, depending on whether pretest and posttest scores are strongly or weakly correlated.

Note that for any data set, applying an analysis that ignore correlation between correlated scores will produce a statistical analysis that is more significant, but spurious. Note also that the advantage of pretesting when there is one is that effects of the independent variable are more likely to be detected, not that spurious findings are more likely to be avoided. This advantage in detection actually comes at the cost of an increased likelihood of spurious findings, as we shall see.

### External Validity

External validity expresses the extent to which the results of an experiment can be generalized beyond the conditions of the experiment. This usually means to different populations measured under different conditions, sometimes with different operational definitions of the variables. It would logically be impossible to have higher external validity than internal validity. In fact, there is little or no scientific basis for estimating the external validity of the great majority of experiments. If a sample were randomly and independently drawn from a specified population, one could lawfully generalize results to that population, but such sampling is seldom done, except in survey research. (Random sampling is not to be confused with random assignment to experimental conditions.) Generalization to unmeasured populations and conditions is usually done on faith alone. Fortunately, it seems to work pretty well. If manipulations designed to improve health, for example, are basically similar, they usually produce effects in the same direction for 12-year-old city girls as for 12-year-old rural girls, although very often to a greater or lesser extent, and there are plenty of exceptions.

It is indeed fortunate that many experiments in many disciplines have sufficient, albeit unknown, external validity to be generally replicable with different populations and subtle or not so subtle differences in experimental conditions; otherwise, scientific research in these disciplines would have little value. High external validity is often the result of [1] robust experimental manipulations that produce big effects and [2,3] wide variability in the characteristics of the participants in educational research. Unfortunately, manipulations of the independent variable often have small effects, probably because the independent variable usually expresses only one of the myriad sources of individual differences in what is measured by the dependent variable. And, wide variability in participants' characteristics reduces the power the experiment to detect whatever effects the independent variable has. To overcome these two difficulties, some experiments must be designed that have high power.

### Power

The power of an experiment is, strictly speaking, its ability to demonstrate the effects of the independent variable under given conditions. Some of the characteristics of experiments designed to have high power, especially having homogeneous participants and extreme values of the independent variable, tend to reduce external

validity. However, such experiments may be useful for testing theory, or for identifying variables that affect scores, even if those effects are small. However, pretesting, the main subject of this pretest, involves a different compromise. Pretests may increase power, but do so at the cost of reducing the internal validity of experiments. That is to say, pretests can improve our ability to detect small effects of the independent variable, but reduce our ability to be certain that the effects we identify are actually those of the independent variable, because they create a confound.

### Pretest-treatment Interactions

A classic methodology text (Kerlinger 1964, pp. 339) identified that confound long ago: the test-treatment or sensitization-treatment interaction. The pretest may cause the experimental manipulation to produce an effect it would otherwise not have produced. Or it may cause different levels of the independent variable to have effects greater or smaller than they would have had if not preceded by a pretest. This is not the effect of testing itself; if all participants are pretested, the effect of the pretest per se will on average be the same for all. Pretesting will be a constant like all the other things held constant in a good experiment, not a variable whose effects could be mistaken for those of the independent variable. The confound is the interaction of the pretest with experimental manipulation (the different levels of the independent variable). It is not difficult to recognize that being pretested could make a participant more responsive or less responsive to a manipulation. Children pretested on mathematics, even if given no feedback, may recognize what they need to learn in order to do better next time, and consequently respond well to experimental instruction. "For instance, in education a pretest can be the impetus for learning the correct answers to items and thus increase the posttest level of performance" (Cook and Campbell 1979, pp.102).

This is true whether or not the pretest has two different forms. Children in a control group, however, have no opportunity to apply the insight they gained from the pretest. Even though the pretest is a constant, it can potentiate responses at different levels of the independent variable differently, because those levels are different. That is the nature of an interaction between variables. The interaction could also be negative. If children deduced from the pretest that they were poor at math (rightly or wrongly) they might disengage and not respond to experimental instruction which would otherwise have been effective. It is also easy to recognize, of course, that there might be times when a test-treatment interaction would be exceedingly unlikely to occur, even though participants were pretested. The pretest might, for example, be covert observations that participants could not detect. Such observations could not be expected to have any impact on the participants' behavior, much less different impacts at different levels of the independent variable. The problem with the test-treatment confound is that its effect cannot be measured or disproven absent Solomon-Campbell designs (Campbell & Stanley 1963). These designs are so inefficient that the author has never seen one outside of textbooks in more than 40 years.

The upshot is that if there are pretests, Tom can argue that the test-treatment interaction had a big effect, accounting for all of the

differences that would otherwise be attributed to the independent variable; Dick can argue that it had no effect at all, and that the differences observed are entirely due to the independent variable, and Harry can argue that both the confound and the independent variable had effects, combining to produce the participants' scores. None of these scientists can prove the other wrong, they can only argue, as is the case with any other confound. And science hates argument, it wants proof.

### When to Pretest and When Not

It should be obvious that one should avoid pretesting in situations where a knowledgeable scientist would admit that there is a reasonable probability of a test-treatment interaction. These include situations where a pretest might guide participants to the purpose of the research or give them information about what was expected on outcome measures, or motivate or demotivate them, or in any way enable those receiving to the experimental treatment to profit from the testing more than those in the control group. Often these will be cases in which the behavior measured is relatively labile. If it is going to take a year of instruction to produce differences between participants in the different experimental conditions, a 20 minute pretest is unlikely to have a big impact on their final score. But again, one can only speculate, not prove, that the confound made no difference. Another reason for avoiding pretests that is not so critical in a scientific sense, but may be of great practical significance, is the time required of participants and the investments of researcher's resources. This should not be an issue when there is one brief pretest, but it may be an issue, when there are many dependent variables, as there often are, and tests are longer or involve scarce resources. This is when the mistaken belief of too many researchers and reviewers that research cannot be trusted if there are not pretests is particularly damaging. The unfounded belief results in less research, when pretesting is expensive or involves long delays or intricate analyses by specialists who are hard-pressed to accomplish all that is asked of them.

So when should a researcher pretest? The short answer is that the design weakness of pretesting often must be accepted when high power is needed. Three conditions usually signal the researcher that pretests are desirable, and that arguments over whether the pretests interacted with the independent variable must be dealt with, even though they cannot be defeated. These conditions are:

- a) the researcher has reason to believe that the effect of the independent variable will be small;
- b) the number of participants is small; or
- c) the participants are heterogeneous in relevant characteristics. Any of these conditions alone, and certainly two or more in combination, may result in a Type II error, the failure to detect the effect of the independent variable when it has one. Note that they do not affect the probability of a Type 1 error, an erroneous rejection of the null hypothesis. Hence, absent any of these conditions, one should not pretest.

### Acknowledgement

The research reported here was supported by the Institute of Education Sciences, U.S. Department of Education, through Grant 305A170114 to George Mason University. The opinions expressed are those of the authors and do not represent views of the Institute or the U.S. Department of Education.

### References

1. Campbell, DT & Stanley J (1963) Experimental Designs and quasi-Experimental Designs for Research. Skokie, IL: Rand McNally, USA.
2. Cook, TD & Campbell, DT (1979) Quasi-Experimentation. Design & Analysis Issues for Field Settings. Chicago, IL: Rand McNally College Publishing Company.
3. Kerlinger FN (1973) Foundations of Behavioral Research (2<sup>nd</sup> edn.). Holt, Rinehart, and Winston, Inc., New York, USA.



This work is licensed under Creative Commons Attribution 4.0 License

Submission Link: <https://biomedres.us/submit-manuscript.php>



### Assets of Publishing with us

- Global archiving of articles
- Immediate, unrestricted online access
- Rigorous Peer Review Process
- Authors Retain Copyrights
- Unique DOI for all articles

<https://biomedres.us/>