QUASI-EXPERIMENTATION

René van Hezewijk
Psychology, Open University of the Netherlands, Heerlen, The Netherlands

Keywords: critical rationalism, empiricism, experimentation, external validity, generalizability, independent (or experimental) variables, internal validity, non-equivalent groups designs, pre-experimental designs, quasi-experimental designs, reactive measurements, time-series designs, true experimental designs, valid reasoning

Contents

Summary

This article presents background and basic ideas on a number of research designs called quasi-experimental designs. Donald T. Campbell introduced the term to indicate a number of research designs that are useful in education, ecology, organizational studies, and in other situations where, for one reason or another, the ideal experimental designs cannot be attained. The analysis of these quasi-experiments is related to the threats to internal and external validity that were also identified by Campbell. Quasi-experiments differ from experiments and among themselves in the way (and the degree to which) threats to validity are handled.

This article first presents a brief overview of the history of quasi-experimentation and the theory of quasi-experimental designs as developed especially by Donald T. Campbell and his co-workers. It makes some preliminary philosophical remarks and discusses causal relations, experiments, and quasi-experiments in the behavioral, social, and life sciences. Threats to the internal and external validity of experiments in these sciences, that can be identified *a priori*, are briefly discussed. The next section discusses three pre-experimental designs, three true experimental designs, and a number of quasi-experimental designs, and introduces especially the non-equivalent groups designs and time-series designs. The following section introduces some principles of statistical analysis used in quasi-experimental design. The last section draws some conclusions.

1. Introduction

The history of the growth of knowledge about nature—including the starry heavens, the living systems, and human behavior and experience—has made much use of experiments and quasi-experiments. Once a farmer must have discovered that milk stored in the removed stomach of a dead cow became cheese. He or another farmer must have found by testing alternatives that the substance left in the stomach of the cow—rennet—is a necessary element in producing cheese. Whether this was an experiment or a quasi-experiment can no longer be reconstructed. However, it must have been some kind of implicit experimental reasoning.

As has been pointed out by the historian of classic science G.E.R. Lloyd, Greek scientists occasionally experimented in order to find out what really happens in the human body (e.g. whether liquid passes through the lungs) or in nature (e.g. Ptolemy's investigations of the refraction of light passing from air through water and glass). However, according to Lloyd, Greek speculation about the physical world not so much lacks empirical research or inadequately debated theories as it fails to link their theories to experimental results.

The central role that experiments play in today's science, including biology and psychology, has only been gradually attained. It took learned persons such as Francis Bacon, Galileo Galilei, René Descartes, and Robert Boyle to go beyond providing demonstrations of a certain assertion, formula, hypothesis, or deduction to testing hypotheses. Hypothesis testing remains the experiment's main function today. A demonstration only illustrates what a theory implies, whereas a true experiment severely tests a theory's implications. These earlier scientists not only used experiments as the precursors of severe tests, they also reasoned about the way proper experiments should be conducted and what could be concluded from them. In other words, they were at the cradle of the theory of experiments.

2. Experimentation and Quasi-Experimentation

2.1. History and Theories of Experimentation and Quasi-Experimentation

The development of the theory of the experiment—including the match between empirical research and theoretical reasoning—has been important in the development of the theory of conditionals in logic. Reasoning in terms of "If P, then Q" is only one or two steps away from acting in terms of "If I do this, I'll get that." Acting in terms of "If P, then Q" with the purpose of finding out if you are right is experimenting. So the logic of conditionals can be seen as the theory of acting to produce *q*s by manipulation of *p*s.
Philo, one of the Stoics in Greek classical philosophy, gave an account of conditionals that already contained the three most important argument forms. The first is the affirmative argument *modus ponendo ponens*, mostly known as *modus ponens*:
If P, then Q.
P.
Therefore, Q.

The second is called *contraposition* or *modus tollendo tollens* (better known as *modus tollens*):
If P, then Q.
Not Q.
Therefore, not P.

The Stoics also suggested a third argument, called *modus tollendo ponens* or the *disjunctive syllogism*.
If P or Q.
P.
Therefore, not Q.

Note that the "or" in this syllogism is the exclusive disjunction ("either P or Q, not both") instead of the inclusive disjunction that modern parlance usually implies with an "or" (i.e. "and/or").

The development of the logic of conditionals goes beyond the scope of the present article. Nevertheless, these argument forms play a major role in the fundamentals of the theory of experiments. It is by these standards that the validity of (experimental and other) reasoning is to be judged, and that the analysis of the approximations to valid reasoning called quasi-experiments can be introduced.

Much of the history of experimentation and the theory of experiments will be undistinguishable from the history of (the theory of) quasi-experimentation. This is due to some paradoxes in the history and philosophy of science that must be pointed out briefly. Much of the modern, critical realist view of science can be attributed to the critical rationalist philosophy of Sir Karl Popper and a number of his "sophisticators," for example, Imre Lakatos, John Watkins, David Miller, and Alan Musgrave. All of them have elaborated on Popper's falsificationist theory of knowledge and of science, thereby stressing—in the eyes of some perhaps even overstressing—the importance of theories and their critical inspection by experiments. As has been observed by a few, for example Franklin, the emphasis on the fallibility of knowledge, measurement, and observation has lead to a failure to appreciate the positive role of experiments in the history and growth of science.

Although Donald T. Campbell was one of those "sophisticators" of Popper's philosophy, he has emphasized two other aspects of Popper's work: its evolutionary aspect and—in association with that—the implication of the fallibility of knowledge; to wit, the approximation to truth as the only reasonable conclusion of fallibility. Thus, Campbell concluded in 1981 that knowledge is always tentative and can only be incrementally more certain with increasing data or evidence. With this in mind, it is understandable why especially Campbell has contributed much to the theory of quasi-experiments.

Campbell has always emphasized, enthusiastically and sympathetically, that the aim of science cannot be to publish only definite results, but to publish the best approximation to truth available at that moment and in that situation. That is, one always has to take into account the imperfections of the measuring devices, the achievable experimental designs, the fact that one must work with groups of participants who have other concerns than being experimental subjects, the fact that the scores will be different next time even if the same test is administered to the same person, and so on. Therefore, one should use multiple approaches that—though they may be imperfect taken individually—may complement each other when combined. Campbell applied this idea to measurement theory, as well as to experimental design.

Consequently, although the experiment with full and neat control of all parameters and with all conditions in optimal order is the ideal, one need not despair when the ideal cannot be obtained. One of Campbell's greatest contributions is the theory of quasi-experimental designs to be discussed shortly.

However, although Campbell's contributions can be seen as an elaboration on Popper's work, there are two other elements that make it necessary to give Campbell full credit as a *critical* follower of Popper. One is that Campbell has emphasized and worked on the growth of knowledge in, and about, practical issues. Popper's pursuits are about the growth of knowledge under the assumption of an ideal science. Campbell—perhaps because he was a psychologist—emphasized that much can be gained from practically constrained situations, *and also* that much of what we can know should and could be used in the field.

A second interesting difference related to this is that Popper has focused—as it were—on the modus tollens. If an expected consequence is found not to be the case then the argument must be wrong. Campbell—as it were—suggested that much could still be learned from the argument, although one is not completely sure about the truth of the consequence. He emphasizes the modus ponens (to be used in the field when certain treatments are called for) *and* the disjunctive syllogism (to be used to draw optimal if not maximum conclusions from near to ideal research situations). He managed to do this, according to the present author, by trying to see all situations in which truth or validity may be at stake, as situations in which approximating truth is to be preferred to not bothering to argue rationally at all.

2.2. Preliminary Philosophical Remarks

The main reason to design experiments in the sciences is to test ideas about the causes of the phenomena one studies. An experiment is set up to demand an answer from nature about the truth of a suspected cause of an event or state of affairs. Unlike the atheoretical empiricists who seem to accept raw observation as sufficient, critical realists force nature to give answers to questions about causes that cannot be obtained by reasoning or observation alone. Although nature does not speak a natural language—so to speak—it is organized such that it gives a signal that can be interpreted as a “yes” or “no” if the questions are put properly.

A critical realist perspective amounts to the position that causal relations are real. They are real in the sense that indeed nature “works” thanks to the real existence of causal agencies. That is, nothing would happen in nature unless there is a cause that makes it happen. Moreover, it is a cause that can be identified if it can be manipulated, or that we can hope some day to be able to manipulate, or that can be manipulated in principle.

However, we have not been given the competence to see these causes immediately and directly. Our perception of causal relations is imperfect. Our competence to approach truth with certainty is imperfect. Human knowledge is fallible. Instead of drawing the conclusion that therefore we can know nothing—the skeptical position—or that all knowledge depends on the perspective taken—the relativist position—many rational philosophers of science and rational scientists search for best approximations of truth. If absolute truth and perfect causal knowledge cannot be obtained, try to eliminate as much falsity as possible.

Campbell is one of these critical rationalists. He concluded that, if perfect experimental designs can be developed but not be applied for pragmatic, ethical, or other reasons, find the next best solutions. This is often the case in social and behavioral sciences, as well as in other life sciences. Some experiments are impossible because of ethical or judicial objections; other experiments are impossible because of social, historical, or pragmatic impediments.

2.3. Causal Relations and Experiments in Psychology and Other Life Sciences

In psychology and other life sciences, the aim of researchers is to discover the causal relations that govern the behavior and experiences of living systems. In order to test the hypotheses that theorists have developed, researchers need to conduct experiments. Ideally experimenters do three things: 1) make sure the independent (or experimental) variables (the cause(s)) are manipulated under precise control of the experimenters; 2) make sure other variables do not change and thereby confound the experimental results; 3) measure the precise effects on the dependent variable—the consequence(s). To the degree they are certain that no other but the experimental variable has had an effect on the dependent variables, they are certain that there is a causal relationship between cause and effect. In other words, the experiment is internally valid if it is certain that the experimental variable is the only variable that has an effect on the dependent variable(s).

Compared with the natural sciences (in which the experimental material is supposed to be non-living matter—that is it does not change on its own), the problem in the life sciences is that living matter changes no matter what control experimenters have or want. All participants in a psychological experiment “have a mind of their own.” They have a history (including a history of testing), they mature, they decide to participate or to stop participating in the experiment, and the like, thereby influencing the outcomes of experimental testing in a way that will not be found in the natural sciences. Therefore, some special measures must be taken to approximate the best validity attainable. The “theory of experimental and quasi-experimental designs” offers solutions for problems like these.

The external validity (or representativeness, or generalizability) of research in the life sciences—psychology, in particular—partly depends on the internal validity. If we are not sure whether the effect found is due to no other than the experimental variable (internal validity), we cannot be sure whether the effect found can be generalized to the population (external validity). Internal validity is a necessary condition (not a sufficient condition) for external validity. External validity is concerned with the question whether the effect found in the (ideal) experiment can be generalized to the population, to other settings, or to other resembling variables (resembling under a theory of resemblance or similarity). The theory of experimental and quasi-experimental designs offers solutions for that as well.

2.4. Threats to Internal Validity

In experiments with living systems the subjects, especially human participants, can be expected to react to many influences from inside or outside the subject, apart from the experimental variables under study. Both internal and external validity are thereby threatened. Most of these threats can be indicated theoretically in advance. That is, they can be known to influence participants in principle, regardless of the hypotheses tested. Consequently we can control for them as much as possible. The following briefly discusses these threats.

2.4.1. History

An important threat to internal validity is the “history” participating subjects undergo. This concerns the influences of external events *during the empirical study*. For instance, measuring the effect of an anti-ethnocentrism program can be disturbed when, between the program (X) and the measurement (O), ethnocentric disturbances come to a sudden political crisis. This may be due to an external event nobody can control and that triggers the change in political behavior (e.g. a sudden influx of illegal immigrants from another continent). It is impossible to isolate participants from events like these in a field experiment, so it is necessary to control for them. Thus, events participants get involved in that are beyond the experimental control should be controlled for in another way, for instance by measuring attitudes before and after the experimental manipulation, comparing them, and excluding a possible alternative explanation by controlling for their effect.

2.4.2. Maturation

A second type of confounding variable concerns the effects of the continuation of life during the experimental phase. Subjects grow older, get hungry, become tired, or gain competence in dealing with certain aspects of the experimental treatment that are not controlled for. In education and in organizational psychology the maturation factor is a variable to be taken seriously.

2.4.3. Testing

A third source of invalidity points to the mere fact of testing itself. A stone is not affected by measuring its speed, width, or weight. However, subjects—especially if they are human—are influenced by the mere fact of testing. Specifically, tests that involve skills implicated in the hypothesis tested may change the subject. The first measurement of a reading test will be different from the third, whatever happens in between the tests. I.Q. tests are another example. Subjects learn to score better. It is estimated that as much as five I.Q. points improvement from first to second testing can be attributed to the testing itself.

2.4.4. Instrumentation

Whenever measurement tools (instruments) are used, the using of the tool creates its own issues. This is called instrumentation. Instrumentation is a factor known in physics as well: the wear and tear of a measuring device (or instrument) might lead to a mistaken attribution to the experimental variable of the difference between first and second measurement. This is the reason why I.Q. tests have to be updated. In the life sciences, this is a serious factor. When panels of observers or interviewers are used to measure change, for example, it might very well be that a change from O_1 to O_2 is due to the observers' fatigue, but is mistakenly attributed to X. Campbell would call this "instrumentation," and it threats the internal validity of experiments.

2.4.5. Regression

Suppose researchers need to know how people within a wide range of I.Q. scores react to a new treatment to "boost your I.Q." This is the pretest (Test 1), the result of which is used to form five groups (very high I.Q., high, medium, low, and very low). These groups are subjected to the I.Q. boosting technique (we're sorry to say this is an imaginary example), and are then tested again (Test 2). It is predicted that the effect will show specifically in the group with the lower scores (the high-scoring individuals have reached a ceiling, says the imaginary theory). Suppose that the low-scoring individuals have migrated to the medium group. We are inclined to conclude that the I.Q. booster had a positive effect. However, the results do not justify the conclusion. The extreme scores on Test 1 will tend to the mean because circumstances responsible for the extreme nature of the scores in Test 1 are not likely to recur in Test 2. The change in the scores of the extreme group may, once again, be mistakenly attributed to X.

2.4.6. Selection

When participants are assigned to experimental group and control group(s) it often cannot be avoided that the recruitment is not random or the groups are not equivalent in the theoretically relevant way. Thus, sometimes participants will be found in the experimental group that are more favorable of research in general, and therefore produce a better result on the experimental treatment. What is called the Hawthorne effect is a good example.

In a large company, the effect of lighting on the efficiency of switchboard personnel was measured. Control groups were used. Unexpectedly, any change in lighting whatsoever had a positive effect on the experimental group, even when lighting dropped well below normal. The mere fact of being selected for an experiment resulted in scores that could not be distinguished from the experimental variable, in this case lighting. The influence on the experimental results need not be elaborated here.

2.4.7. Mortality

Related to the selection effect there may be a non-random dropping out of participants. For instance, teachers with a low achievement motivation will have a higher probability of dropping out of the experimental group in which a new teaching method is tested. (Note that achievement motivation is not a factor that *typically* affects the efficacy of teaching methods.

Participants of experiments other than those testing teaching methods may differ in the inclination to achieve, thus affecting outcomes in many types of experiments by dropping out sooner (low achievement motivation) or later (high motivation). If there were a direct effect on teaching this would threaten external validity, as it would directly affect the experimental variable *X* by the interaction with selection.)

2.4.8. Interaction of Mortality, and/or Selection, and/or Maturation, Etc.

The above threats to internal validity may have an effect not only purely by themselves, but also in interaction with each other. For instance, any changes in the participants due to what has happened to them during the period in which the experiment is conducted may lead to an increase in their tendency to drop out.

-
-
-

Bibliography

Campbell D.T. (1957). Factors relevant to the validity of experiments in social settings. *Psychological Bulletin* **54**, 297–312. [First publication by Campbell on factors threatening the valid conclusions from experiments.]

Campbell D.T. (1974). Evolutionary epistemology. *The Philosophy of Sir Karl Popper* (ed. P.A.Ed. Schilpp), pp. 413–463. La Salle, Ill.: Open Court. [Campbell discusses what he learned from Popper about the growth of knowledge as an evolutionary process, and how truth is only approximated.]

Campbell D.T. and Fiske D.W. (1959). Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychological Bulletin* **56**, 81–105. [Campbell develops a theory of measurement by which the whole of several fallible, imperfect, or incomplete measurements can be more than the sum of the individual measurements.]

Campbell D.T., Riecken H.W., Boruch R.F., Caplan N., Glennan T.K., Pratt J., Rees A., and Williams W. (1974). Quasi-experimental designs. *Social Experimentation: A Method for Planning and Evaluating Social Intervention* (ed. H.W. Riecken and R.F. Boruch), pp. 87–116. New York: Academic Press. [Summary of ideas on (quasi) experimentation in social settings, and its relation to action research.]

Campbell D.T. and Ross H.L. (1968). The Connecticut crackdown on speeding: time series data in quasi-experimental analysis. *Law and Society Review* **3**, 53. [Example showing the usefulness of good (quasi) experimentation in social and legal matters.]

Campbell D.T. and Stanley J.C. (1963). *Experimental and Quasi-Experimental Designs for Research*, 10th edn., 84 pp. Chicago: Rand McNally. [The original (small) book publication that contained the basis of the theory of quasi-experimentation.]

Campbell D.T., Webb E.T., Schwartz R.D., Sechrest L., and Grove J.B. (1981). Approximations to knowledge. *Nonreactive Measures in the Social Sciences* (ed. E.T. Webb, D.T. Campbell, R.D. Schwartz, L. Sechrest, and J.B. Grove), pp. 35–77. Boston: Houghton Mifflin. [Chapter indicating the idea of approximation to truth as the non-relativistic, non-skeptical, rationalist alternative for imperfect experimentation and measurement.]

Cook H.J. (1990). The new philosophy and medicine in seventeenth-century England. *Reappraisals of the Scientific Revolution* (ed. D.C. Lindberg and R.S. Westman), pp. 397–436. Cambridge: Cambridge University Press. [Chapter on the history of the development of experiments.]

Cook T.D. and Campbell D.T. (1979). *Quasi-Experimentation: Design and Analysis Issues for Field Settings*, 405 pp. Dallas, Ill.: Houghton Mifflin. [Second, much larger book publication on quasi-experimentation, with more sophisticated designs as well as chapters on the statistical analysis of quasi-experiments.]

Franklin A. (1986). *The Neglect of Experiment*, 290 pp. Cambridge: Cambridge University Press. [A book publication containing arguments about the proper place of experiments in the growth of knowledge.]

Lloyd G.E.R. (1964). Experiment in early Greek philosophy and medicine. *Proceedings of the Cambridge Philological Society NS*, **10**, 50–72. [Article on the history of experimenting in classical Greece.]

Lloyd G.E.R. (1979). *Magic, Reason and Experience: Studies in the Origin and Development of Greek Science*, 335 pp. Cambridge: Cambridge University Press. [Book publication on the history of ideas in relation to observation, research, and experiments.]

Neale J.M. and Liebert R.M. (1986). *Science and Behavior: An Introduction to Methods of Research*, 3rd edn., 324 pp. Englewood Cliffs, N.J.: Prentice-Hall. [Book publication with a balanced introduction and overview of methodological issues in behavioral sciences.]

Sanford D.H. (1989). *If P, Then Q: Conditionals and the Foundations of Reasoning*, 265 pp. London: Routledge. [Book publication on the logic of conditional reasoning.]

Biographical Sketch

Professor René van Hezewijk is a theoretical psychologist and dean of the Faculty of Psychology at The Open University of the Netherlands. He also is chief editor of the leading Dutch journal for psychology. He has published in various journals on theoretical issues in psychology.