Metodología de las Ciencias del Comportamiento 3(2), 283-290

Observational studies and quasi-experimental designs: Similarities, differences and generalizations

Salvador Chacón-Moscoso* †

Universidad de Sevilla

William Shadish

University of Memphis

Resumen

En este artículo se diferencia entre los cuasiexperimentos ('Q-E') y estudios observacionales ('OS'). En los cuasiexperimentos se implementan intervenciones intencionalmente para observar sus efectos sin asignar aleatoriamente las unidades a las condiciones del estudio. En los estudios observacionales se observa la ocurrencia de comportamientos, de tal forma es posible obtener registros formales y cuantitativos de dichos comportamientos. Por lo que los 'OS' observan el comportamiento natural sin manipularlo. No obstante, cuando se trabaja en investigación de campo, como en el caso de la evaluación de programas, esos límites son a veces confusos. Este artículo presenta un análisis de algunas de las similitudes y diferencias entre 'Q-E' y 'OS' con respecto a la obtención de resultados válidos y generalizables. Describiremos comparativamente las características estructurales de los diseños en 'OS' y 'Q-E'; y también se tratará la generalización de resultados obtenidos desde 'OS' y 'Q-E'. Para ambos análisis se utilizará como referente principal el trabajo presentado en Shadish, Cook y Campbell (2002).

PALABRAS CLAVE: *cuasi-experimentos, estudios observacionales, generalización, evaluación de programas.*

Abstract

In this article, we differentiate between quasi-experiments (Q-E) and observational studies (OS). In Q-E, an intervention is deliberately introduced to observe its effects, but units are not assigned to conditions randomly. In OS, a study observes the occurrence of behaviors, sometimes also allowing them to be formally recorded and quantified; that is, OS observe natural behavior without manipulating it. But, when we work in field research, as in program evaluation, those boundaries are sometimes blurred. This paper presents an analysis of some similarities and differences between Q-E and OS with respect to obtaining valid and generalizable results. We will describe comparatively the structural design characteristics of OS and Q-E; and we will also discuss the generalization of results obtained from OS and Q-E. In doing both analyses, we borrow from prior work presented in Shadish, Cook and Campbell (2002).

KEY WORDS: *quasi-experiments, observational studies, generalization, program evaluation.*

The original idea for this paper was born in the city of Seville, in September of 2000, during a seminar about 'causation, validity and randomized experiments', given by William Shadish (University of Memphis). This seminar mainly consisted of previewing advances in experimental and quasi-experimental designs now published in Shadish, Cook and Campbell (2002). At that seminar, the first author speculated that some of the ideas in the latter book could be applied to understanding some of the similarities and differences between observational studies and quasi-experiments. This paper presents some initial ideas toward that end. A prior version of this work

*The authors would like to thank Thomas D. Cook and M.T. Anguera-Argilaga for providing helpful comments on previous drafts of the paper.

†Dirección postal del primer autor: Departamento de Psicología Experimental. Facultad de Psicología, Camilo José Cela s/n, 41018-Sevilla (Spain). E-mail: <schacon@psicoexp.us.es.>

was presented in the 6*th* Conference of the *European Association of Psychological Assessment* (Chacón-Moscoso, Cook, Anguera-Argilaga, Pérez-Gil, and Shadish, 2001, September). A more extensive description of what we are going to outline in this paper will be published in Chacón-Moscoso, Shadish, Cook, Anguera-Argilaga, and Sánchez-Meca (in preparation).

A classic analysis of the difference between quasi-experiments and observational studies might refer only to the degree of control of the intervention that the researcher has. In that analysis, a quasi-experiment (Q-E) is a study in which an intervention is deliberately introduced to observe its effects, but in which units are not assigned to conditions randomly (Shadish, Cook and Campbell, 2002). An observational study (OS) observes the occurrence of behavior, allowing it to be formally recorded and quantified, but without manipulating it (Anguera, in press). This is similar to the criterion of 'degree of control over scheduling intervention' used to classify these studies by Arnau, Anguera and Gómez (1990) and Moreno, Martínez and Chacón (2001). Following this criterion, OS might be called *low intervention* studies (observers do not have any control over the situation and they just observe behaviors that appear depending on subjects' desire). Q-E might be called *medium intervention* studies (researchers have certain degree of control over the situation, without being able to assign subjects randomly to conditions, and can provoke subjects' behaviors manipulating variables). Randomized experimental studies might be called *high intervention* studies (researchers have a high control over the situation, and provoke behaviors). Of course, the distinction among low, medium, and high intervention studies is merely one of convenience, as some quasi-experiments can have high control over assignment to the intervention, as with regression discontinuity designs (Shadish et al., 2002). Nonetheless, the distinction becomes a matter of degree rather than a matter being absolute.

However, this classic analysis may lead to such clear differentiations when we go into field research areas, such as program evaluation, making it difficult to assign study labels in such contexts. When we are in such contexts, we try to implement the most precise and high quality study for that particular situation, without paying much attention to study labels. Then we find that in such contexts, the distinction between Q-E and OS is more diffuse, and they share many similarities.

Scientific validity is a property of knowledge claims, not methods. No method guarantees validity. If so, then we can use the same validity logic to judge knowledge claims when they come from either Q-E or OS (Shadish, 1995). Of course, these two methods often focus on different objects of study (that is, different focal questions of interest), with Q-E being oriented to study causal inferences while OS is mainly focused on descriptions.

In the analysis that follows, we follow the validity framework of analysis used by Shadish, Cook and Campbell (2002), also outlined in Corrin and Cook (1998) and Shadish and Cook (1999). This validity framework was originally designed for causal inference generalization. Obviously this is not the objective of all OS, but, excluding internal validity, the rest of the types of validity (statistical conclusion, construct validity and external validity) can be applied to OS just as much as they can to QE. We will also borrow from the 'utos' system of structural design dimensions first introduced by Cronbach (1982) in order to compare both kind of studies. Finally, having this comparative analysis as a point of reference, we will discuss implications for generalizing results from Q-E and OS using a theory proposed originally by Cook

(1990), and also presented in Cook (1993, 2000), Shadish (1995) and Shadish, Cook and Campbell (op. cit.).

Structural design dimensions related to validity framework

This section of the paper introduces basic structural design dimensions that we will use to distinguish between OS and Q-E. We will also show how those dimensions might relate to the four validity types.

The structural design dimensions are adapted from Cronbach (1982), who said that studies consist of units (often persons), treatments, outcomes, and settings. Taking the first letter from each of these four words, he defined the acronym *utos* to refer to the "instances on which data are collected" (Cronbach, 1982, p. 78)- to the actual people, treatments, measures, and settings that were sampled in the experiment. He then defined two problems of generalization: (a) generalizing to the "domain about which [the] question is asked" (p. 79), which he called *UTOS*; and (b) generalizing to "units, treatments, variables, and settings not directly observed" (p. 83), which he called **UTOS.*

Cook and Campbell (1979) and Shadish, Cook and Campbell (2002) analyze experiments in terms of four validity types: statistical conclusion validity, internal validity, construct validity, and external validity. Next, in table 1, we define the four types of validity.

<u>Statistical Conclusion Validity</u>: The validity of inferences about the correlation (covariation) between treatment and outcome. <u>Internal Validity</u>: The validity of inferences about whether observed covariation between A (the presumed treatment) and B (the presumed outcome) reflects a causal relationship from A to B, as those variables were manipulated or measured. <u>Construct Validity</u>: The validity of inferences about the higher-order constructs that represent sampling particulars. <u>External Validity</u>: The validity of inferences about whether the cause-effect relationship holds over variation in persons, settings, treatment variables, and measurement variables.

Table 1. Four types of validity

These four validity types could map onto Cronbach's system as follows (although this mapping is an adaptation of Cronbach's system, and differs from his system in ways we will point out):

1. Statistical Conclusion Validity refers to the study of the correlation between "t" (treatment) and "o" (outcome) in "utos". Its aim is to analyze if and how much the treatment, as implemented, correlates with the measured outcomes. For example, in a neuropsychological rehabilitation program intended to increase visospatial coordination in traumatic brain injury (TBI) patients, statistical conclusion validity seeks to answer the question of whether and how much the increase in visospatial coor-

dination (outcome) is correlated with the implementation of the neuropsychological rehabilitation program (treatment).

2. Internal Validity refers to the question of a causal relationship between "t" (treatment) and "o" (outcomes). Its aim is to analyze whether some or all of the observed correlation between treatment and outcome was the result of treatment, or would have been displayed in the absence of the program. Following the previous example, to study internal validity means analysing whether the neuropsychological rehabilitation program (treatment) caused a some or all of the increase in the visospatial coordination (outcome) in the patients, or whether visospatial coordination would have increased (outcome) in whole or in part if the program had not been implemented.

3. Construct Validity involves making inferences about the constructs that research operations represent, based on the particular sample of "utos" used in the intervention. Continuing with our example, construct validity concerns questions that are like the following: a.*units*, do the units in the study have the characteristics of TBI patients, and not other characteristics; b.*treatments*, does the treatment actually studied include the required components of a neuropsychological rehabilitation program, and not of other treatments; c.*outcomes*, is error frequency in a visospatial coordination task well-represented by the label "coordination"; and d.*settings*, is the setting in the study representative of rehabilitation centers, and not on other kinds of settings. In Cronbach's terms, this is closest to his generalization from *utos* to *UTOS*, though it is not identical with it.

4. External Validity might be thought of as whether a causal relationship observed in *utos* holds over variations in units, settings, treatments, and outcomes. In Cronbach's terms, this is closest to his generalization from *utos* to **UTOS*, though again it is not identical because Cronbach limited **UTOS* to things **not** studied, but external validity concerns generalizations over variations that may or may not have been studied. In our example, a possible external validity study would analyze whether the same rehabilitation program could be used in different rehabilitation centers (S*), or on patients suffering traumatic brain injury of a different nature than those in the study (U*).

This same framework can be used to describe structural design dimensions in OS. The main object of OS is to create descriptions and analyze regularities in units, settings and times. OS is concerned with causal connections much less. Following this reasoning we can analyze both OS and QE as to their structure (via the *utos* system).

In the following, table 2, next page, we elaborate some ways in which these basic structural dimensions can be implemented in both OS and QE; and in subsequent pages we provide more details.

1.Units: We can distinguish between unit selection (into the study) and assignment (into conditions). In all OS, selection and assignment criteria are mostly unknown; they are usually also unknown in QE, but some QE designs like regression discontinuity are an exception to this. Another differentiation is that comparison groups in Q-E usually refer to a specific, identifiable person or group of persons, while in OS such groups must be created from available data, and are often not an identifiable group other than by virtue of having been statistically aggregated by the researcher.

U	UNITS (Usually persons)	Selection criteria
		Assignment criteria
		Comparison Groups
T	TREATMENT/ PROGRAM ACTIONS	Intervention level (degree of control over scheduling of interventions)
		Intervention level changes
O	OUTCOMES/INSTRUMENTS	Type of data
		Data quality recording
		Instruments
		Type of instruments
		Instruments changes
S	SETTING CONTEXT	Feasilibity conditions
		Modulation of contextual variables

Table 2. Structural design dimensions of OS and QE

2. Treatment: Regarding the degree of control over scheduling of intervention, Q-E can manipulate variables in order to analyze their effects on outcome (they can have a medium level of intervention). On the other hand, OS tend to have little or no control over scheduling of the interventions (they have a low degree of intervention). In both cases, however, it is important to analyze the intervention level changes carefully. In a QE where the intervention context is continuously changing, at the beginning of the study we can decide to manipulate a certain variable, but perhaps during its implementation it changes; or, we can design an OS and later decide to manipulate some variables.

3. Observations/outcomes: Sometimes in Q-E we can have standardized instruments, as we have a certain degree of control over the situation under study. We have far less control of observations in OS, that is, over data types, quality of recording, and types of instruments, which is why it is especially important to explore data quality in OS (for example, using reliability analyses between and within observers). And, just as the intervention level can change, so also we can have instrument changes during a study, in both OS and QE.

4. Setting: We have distinguished feasibility conditions and modulation of contextual variables. OS have few or no feasibility restrictions as to the setting of study, as this is the feature that makes them so flexible. Q-E are subject to more feasibility restrictions, as they have to find a setting that will allow treatment manipulation and measurement of results with certain instruments.

In summary, then, Q-E and OS differ mainly based on the degree of control over scheduling of interventions. Less consistently, they can also differ with QE having somewhat greater control of observations, somewhat less feasibility in terms of access to a wide array of settings, and sometimes a greater degree of control over unit selection and assignment.

Generalizing results from observational and quasi-experimental studies using a grounded theory

The object of this section is to describe the logic of generalized causal inference from both OS and Q-E. Formal sampling methods are of limited utility for generalizing results in field settings; they can sometimes be used for sampling persons, very rarely for sampling settings, and never for sampling treatments and outcomes. Consequently, we are going to present a theory based on five principles originally proposed by Cook (1990), and also presented in Cook (1993, 2000), Shadish (1995) and Shadish, Cook and Campbell (2002). Because validity is a property of knowledge claims, not methods, then the five principles can be applied to claims about construct validity and external validity generalizations generated by Q-E or OS (Shadish, 1995).

In the following, we are going to summarize five principles of the logic of causal generalization, in the process shedding light on some similarities and differences between OS and Q-E. The five principles are:

a. *Surface Similarities:* Scientists generalize most confidently to applications where the targets of generalization are most similar to those in the original research. The strategy is to assess the match of the operations in the concrete study to the targets of generalization, mainly to the prototypical features of the construct to which we want to generalize. This principle can be applied in both OS and Q-E. In both cases, we may judge whether the units, treatments, outcomes or settings in a study appear to be similar to the intended target of generalization, especially in prototypical features.

b. *Ruling out irrelevancies:* Scientists generalize by identifying those attributes of units, settings, treatments and outcome measures that are irrelevant because they do not change a generalization. The strategy here is to identify irrelevancies, and where possible including diverse array of them in the research so as to demonstrate generalization over them. QE have the advantage that they can more easily introduce such variations deliberately, but OS have the advantage of being able to take advantage of naturally-occurring variations in irrelevancies.

c. *Making discriminations:* Scientists generalize by making discriminations that limit generalization. The strategy is to demonstrate that the findings hold for the operations or constructs as specified, not for alternative operations or constructs. Here again, OS take advantage of naturally-occurring variations in discriminating variations, but QE can introduce those variations deliberately.

d. *Interpolation and Extrapolation:* Scientists generalize by interpolating to unsampled values within the range of the sampled units, settings, treatments and outcomes, and much more difficult, by extrapolating beyond the sampled range. Again, OS and Q-E can both use this principle, but OS has less opportunity to introduce changes into the context of study than Q-E and depends on the natural occurrence of a range of variations on the continuum over which generalization is desired.

e. *Causal explanation:* Scientists generalize by developing and testing explanatory theories about the target of generalization, aiming to specify which parts of the treatment affect which parts of the outcome through which mediating process. When carrying out an OS, researches have a greater degree of difficulty than in QE in following up processes in different variables without introducing modifications into the study context. For example, it is more difficulty for OS to introduce measures of mediating processes.

These five principles of generalization are related to both construct and external validity. These principles oblige us to make explicit our definition of study units, treatments, outcomes, and settings, in their three levels (sample, population, and differential population). Then, this process helps us to specify our constructs of reference in order to design our study, or to criticize our study retrospectively. It is also useful to analyze systematically which kind of generalization we are trying to carry out, in the sense of analyzing which study components or part of those components we want to make the focus of our generalizations.

Conclusions

In this paper we have tried to present a view of Q-E and OS that describes their similarities and discrepancies within some common frameworks. Both types of studies are often considered to come from very different worlds. But in field research like program evaluation, sometimes the distinction between these designs is not so clear. Practitioners need to have a clear framework of analysis from which to design the highest quality designs as possible, independently of which label we use to classify them. In this paper, we have shown that Q-E and OS have some differential features, mainly related to the degree of control over scheduling of intervention, but that they also have much more in common than is often realized. When analyzing their structural design dimensions and how the five principles of generalization might apply to them, we have found more similarities that might be expected. We hope these ideas will eventually help to lead to a common framework of analysis from which to design practical research.

References

Anguera-Argilaga, M.T. (in press). Observational methods (General). In Fernandez-Ballesteros, R. (Ed.). *Enciclopedia of Psychological Assessment.* London: Sage.

Arnau, J.; Anguera, M.T. y Gómez, J. (1990). *Metodología de la investigación en ciencias del comportamiento.* Murcia: Servicio de Publicaciones de la Universidad de Murcia.

Campbell, D. and Stanley, J. (1966). *Experimental and quasi-experimental designs for research.* Chicago: Rand McNally Company.

Chacón-Moscoso, S.; Cook, T.; Anguera-Argilaga, T.; Pérez-Gil, J.A. and Shadish, W. (2001, September). *Convergence between quasi-experimental and observational designs to enhance evaluative results generalization.* Paper presented at the 6^{th} Conference of the European Association of Psychological Assessment. Aachen. Germany.

Chacón-Moscoso, S.; Shadish, W.; Cook, T.; Anguera-Argilaga, T. and Sánchez-Meca, J. (in preparation). Generalización de resultados evaluativos en estudios observacionales y cuasi-experimentales. En Anguera-Argilaga, M.T. and Chacón-Moscoso, S. (Eds.) *Evaluación de programas sociales y sanitarios. Un abordaje metodológico.* Madrid: Síntesis.

Cook, T. D. (1990). The generalization of causal connections: Multiple theories in search of clear practice. In L. Sechrest, E. Perrin and J. Bunker (Eds.) *Research methodology: Strengthening causal interpretations of nonexperimental data* (DHHS Publication No. PHS 90-3454, pp 9-31). Rockville MD: Department of Health and Human Services.

Cook, T. D. (1993). A quasi-sampling theory of the generalization of causal relations. In L. Sechrest and A.G. Scott (Eds.) *Understanding Causes and Generalizing about them* (pp. 39-82). San Francisco: Josey-Bass Publishers.

Cook, T. D. (2000). Towards a practical theory of external validity. In L. Bickman (Ed.) *Validity and Social Experiments: Donald Campbell's Legacy* (Vol.1, pp. 3-43). Thousand Oaks, Ca: Sage.

Cook, T. D. and Campbell, D.T. (1979). *Quasi-experimentation. Desings and analyses uses for fields settings.* Boston: Houghton Mifflin.

Corrin, W. and Cook, T.D. (1998). Design elements of quasi-experiments. In H.J. Waldberg, A.J. Reynolds and H.J. Walberg (Eds.) *Advances in Educational Productivity, 7,* 35-37.

Cronbach, L. J. (1982). *Designing evaluations of educational and social programs.* San Francisco: Jossey-Bass.

Cronbach, L. J.; Ambron, S.R.; Dornbush, S.M.; Hess, R.D.; Hornik, R.C.; Philips, D. C.; Walker, D.F. and Weiner, S.S. (1980). *Toward reform of program evaluation.* San Francisco: Josey-Bass.

Moreno, R.; Martínez, R.J. and Chacón, S. (2001). *Fundamentos metodológicos en Psicología y ciencias afines.* Madrid: Pirámide.

Shadish, W.R. (1995). The logic of generalization: Five common principles to experiments and ethnografies. *American Journal of Community Psychology, 23,* 419-428.

Shadish, W.R. and Cook, T.D. (1999). Design rules: More steps towards a complete theory of quasi-experimentation. *Statistical Science, 14,* 294-300.

Shadish, W.R.; Coók, T.D. and Campbell, D.T. (2002). *Experimental and quasi-experimental designs for generalized causal inference.* Boston: Houghton-Mifflin.

Original revisado: 23/10/2001
Versión final aceptada: 16/11/2001