Dealing with Problems of Causality in Impact Studies

Léonie J Rennie

Curtin University

L.Rennie@curtin.edu.au

Most dictionaries define impact (among other meanings) as "effect, influence", and so by definition, impact studies must measure an effect and attempt to demonstrate that it is a consequence of a particular cause. In educational and sociological research this is extremely difficult. The "gold standard" of a randomised pre-test-post-test control group research design is nigh impossible to implement; instead, researchers are required to establish by other means temporal precedence, consistent association of meaningful magnitude between supposed effect and cause, and eliminate alternative explanations for their correlation. This presentation reports examples of efforts to do this, particularly in non-school based research.

Increasingly, the providers of funded programs or events are expected to meet accountability requirement to demonstrate the social impact of their offerings. Most dictionaries define impact in terms of the "influence or effect exerted by a new idea, concept, ideology, etc." (The Macquarie Encyclopedic Dictionary, 1990). By definition, then, impact studies must measure an effect and attempt to demonstrate that it is a consequence of a particular cause. In the naturally occurring settings where many of the funded programs or events occur, this is extremely difficult. The so-called "gold standard" of randomly controlled trials offers the strongest grounds for establishing causality, but are virtually impossible to implement in most social contexts, because the researcher is not able to select subjects at random and/or determine who experiences the intervention (the program or event). So what alternative approaches are available to the researcher? In this paper, a research approach is suggested to explore a causal connection between an impact and the intervention that preceded it. The paper begins with a discussion of causality, outlines some questions the researcher must answer to establish causal connections, then describes two examples of impact studies in *ex post facto*, natural settings.

Causality in Impact Studies

Research based on explanatory theory is designed to test hypotheses. It is often expressed as "If X, then Y", where X represents the theoretical cause, and Y is its hypothesised effect. The reasoning is deductive, or *a priori*, from cause to effect. The researcher introduces the X in a controlled experiment and determines if Y follows. In contrast, in most impact studies in natural settings the X, an educational

intervention/program/event, has already been introduced and the impact (the Y) occurs after it, so the research question becomes "Did X cause Y?" The reasoning is inductive, or *a posteriori*, from effect to cause, because the researcher works retrospectively. However, the research question remains fundamentally the same: Is there a causal connection between X and Y? What is meant by a causal connection in this kind of *ex post facto* research?

Causality is a concept that cannot be observed, so in any causal research, the research task is to explore the set of conditions that enable the inference that a causal relationship exists. Many introductory texts to social research provide discussions about the nature of causality and using Punch (2009, p. 80) as an example, we can set out four conditions that must be met in order to infer a causal connection between two variables. First, there must be a consistent connection between the supposed cause (X) and its effect (Y). Second, X must occur before Y, that is, the cause must precede the effect. Third, there must be a plausible explanation for the causal connection, that is, how the X has resulted in the Y. Fourth, other plausible alternative explanations for the relationship between X and Y must be discounted.

Randomly controlled trials are powerful ways to meet these four requirements, but in naturalistic situations, they are rarely an option. The task is complicated further because the relationships between variables in social research are rarely unidirectional and most variables are not unidimensional. Our own social experiences tell us that multiple causes and multiple effects are involved in human relationships. Because of this, the term variable is not often used in naturalistic, qualitative research. Lincoln and Guba (1985) analysed the notion of causality in naturalistic inquiry, noting particularly the mutual and continual interaction between the variables or elements in a situation. They argued that the concept of causality be replaced by the concept of "mutual shaping" to reflect the continual contextual variation in which these elements may occur. Lincoln and Guba (1985) stated "explanations are at best 'here-and-now' accounts that represent a 'photographic slice of life' of a dynamic process that, in the next instant, might present a very different aspect" (p. 155). In comparing scientific (or positivistic) with naturalistic research approaches, Guba and Lincoln (1981) suggested that the question of causality "Can X cause Y?" becomes "Can X cause Y in a natural setting?" (p. 65). How can we find an answer to this question for impact studies?

One approach is to return to those four conditions for establishing cause-effect relationships and see how they might be helpful in natural settings. Those conditions can be rephrased as questions for the impact researcher to answer. For ease of expression in the following discussion, the supposed cause – the intervention, program, event or innovation – will be referred to as the intervention.

Question 1: Is there a consistent connection between the impact and intervention? The first step in answering this question is to demonstrate that an impact occurs. Sometimes this is straightforward: If an intervention is designed to increase enrolments in science, a change in enrolments is easy to measure. But in many public events, such as participation in a science festival or visiting an aquarium, the intended impact is less clear. Impacts on people usually result in changed behaviour of some kind; cognitive, affective, psychomotor, or social changes are all possible outcomes. The researcher must identify desired and unexpected outcomes and work out how to

measure them. If the impact is quantifiable, the connection might be demonstrated with a correlation coefficient of some kind. If there is a control group, a similar group who did not experience the intervention, then the impact might be demonstrated by group comparison. When there is no control group, then an impact might be detected by comparing before and after intervention measures. Where a pretest is not feasible, participants might be asked to reflect on, and possible rate, any changes they perceived to result from the intervention. Very often the impact is not quantifiable, and is best described qualitatively, so this will require a different kind of data collection. A mixed-method approach provides a broader based measurement and offers opportunities for complementarity and triangulation of data.

Question 2: Did the intervention happen before the impact? This is the easy question, because the answer is usually yes, and it requires no more attention.

Question 3: Is there a logical explanation for the intervention to have resulted in the impact? Funded interventions usually have explicitly stated reasoning to explain why the intervention should result in the intended outcome. If the intervention is a visit to an institution, then the researcher may have to use the institution's mission statement, or similar, as a starting point. Sometimes, researchers have to figure out the reasoning themselves to inform the development of their measuring instruments.

Question 4: Can other plausible explanations for the impact be discounted? This part of the research may require considerable creativity on the part of the researcher. It needs to be carried out in connection with Questions 1 and 3 so that the research design enables collection of all of the data necessary to test all plausible explanations.

Clearly, establishing a causal connection between an intervention and its impact in a natural setting is very challenging. Rennie and Johnson (2004) set out some of the research issues in measuring impact in such settings, particularly visits to places like museums. Because causal relationships are inferred, the research task becomes one of assembling sufficient evidence (much of it circumstantial, because direct links are rarely evident) to build a case that although incomplete is sufficiently compelling to be persuasive. Two examples of impact studies follow to indicate some of the research difficulties that may be encountered.

The Impact of a Practice Test in Computer Assisted Assessment

This doctoral study was carried out by Leith Sly (2000) who managed the Computer-Assisted Assessment (CAA) Laboratory at Curtin University. Her purpose was to investigate the impact of a practice test taken prior to an assessed test in a tertiary, computer-assisted assessment environment. Earlier research had demonstrated that undergraduate students who took a practice test (PT) prior to their assessed test (AT) tended to perform better on the AT than those who did not, so the Impact Question was: What caused the improved performance?

There were three logical explanations for improved performance: Students used the feedback provided by the PT to direct their subsequent learning resulting in improved performance; students new to the testing environment used the PT to familiarize themselves with the testing procedure, thus reducing anxiety when taking the AT; the more able students chose to take the practice test and this accounted for their superior performance. Other explanations related to an effect of the discipline area, the length of the test, or the sex of the student were less plausible.

Although it would be possible to assign a PT to a random half of the student cohort and compare the results of the PT and non-PT groups, this was not an ethical approach because it would deny some students the opportunity to experience something believed by their lecturers to be beneficial. However, in units where the PT was optional, the AT performance of self-selecting groups of PT and non-PT students could be compared.

The research design used 13 studies in five subject areas to compare the AT performance of the PT and non-PT groups. In some studies other variables could be investigated, including student anxiety (using a standard test to measure anxiety), the length of the PT compared with the AT, the amount of content in the AT covered in the PT, and student ability (using previous achievement). Some qualitative data about students' perceptions of taking and using a PT were gathered via an online survey.

The findings revealed that the PT group consistently outperformed the non-PT group on AT (effect size varied from .24 to .75), and where the PT covered only part of the content on the AT, students performed better on the common part. When students had previous experience of an AT, performance differences were less, and students were less likely to take a PT. Students reported in the online survey that the PT was useful to identify weak areas and revise, and to familiarise with test environment. No relationship with performance difference was found for prior ability, anxiety, length of test, sex of student, or subject area.

These findings suggested, and the students' survey responses supported this suggestion, that the PT had a formative role in directing students in further study and also a role in familiarizing them with the CAA testing procedure and the format and content of the tests. The possible causal influences of a range of other variables, such as anxiety or prior ability, were discounted.

The Impact of a Public Lecture Series on Human Genetics

In line with its mission statement (to increase interest and participation of the community in science and modern technology), a science centre offered a series of evening lectures, open to the public, and presented by experts in various genetic disorders. The purpose of the impact study (Rennie & Williams, 2000) was to assess the success of the lecture series in terms of the science centre's mission, but as this was rather vague, the Impact Question became: What is the impact of the lectures on attendee's perceptions, ideas and opinions about genetics?

Specific challenges arose because lectures are transitory events, there is little time to learn about genetics; people vary and the lectures are different in content and presentation; further, different people attend different numbers and combinations of lectures. Therefore, lecture attendees do not have a common experience, so are unlikely to have a common outcome. The lecture context also presented difficulties for data collection. Data had to be collected quickly, to place minimal demands on the attendees who came for their own agenda, not the researchers' agenda. For each attendee, a set of data had to be completed in one lecture, so a primarily closed response questionnaire was chosen because it enabled more data to be collected in a short time. Further the questionnaire had to be flexible for use at any lecture, with a variety of attendees. Finally, the science centre requested demographic and other data to assist their planning for further lecture series, placing more demands on data

collection. Because there was no comparison group, the impact was measured using a survey in a one-group pretest-posttest design.

The survey was designed in two-parts. Part A was a 14-item, 7-point semantic differential instrument about Human Genetics, with seven items focused on Research in Human Genetics and seven items focused on Learning about Human Genetics. It was used as a pretest and posttest for first time attendees (to control for the effect of attending multiple lectures). The survey was scrutinised for face and construct validity, and field-tested to ensure the instructions and the items were unambiguous and easily understood. Part B measured attendees' responses to the lecture and demographic information. It was used as a posttest for all lecture attendees.

The results suggested some impact. Attendees became more positive in terms of the practical applications of research in human genetics, their interest in the topic, their confidence in talking about it, thinking that benefits were likely to outweigh risks, believing that everyone needs some knowledge about human genetics, but more convinced that genetics knowledge was certain. According to the attendees' own opinions, their interest in, and knowledge of, genetics increased substantially.

Summary and Discussion

Table 1 summarises the outcomes of the two studies in terms of three causal questions (temporal precedence is ignored) and demonstrates that the study investigating the impact of taking a practice test was more successful in established a consistent relationship between taking a PT and higher performance on the assessed test. Further the co-causes (use of feedback and familiarity with the testing context) were satisfactory explanations of this impact, and the findings were sufficiently compelling to recommend broadening the use of practice tests to other units.

The investigation of the impact of the genetics lectures was more challenging. It required the development of measuring instruments tailored to the context, and although some changes were found in attendees' thinking about genetics, the most compelling data were their positive perceptions about their increased interest in, and knowledge about, genetics. Importantly, the science centre found the broader findings and demographic data helpful for planning their next lecture series. Both studies achieved some success, and this was enhanced by the mixed method approach.

Table 1
Summary of the Impact Studies in Terms of the Causal Questions

Causal question	Impact of practice test	Impact of genetics lectures
Connection between Y and X?	Established by repeated studies with consistent results	Pretest-posttest comparisons revealed some changes consistent with a connection
Logical explanation for Y to cause X?	Yes: feedback has a formative role, and familiarity with test environment	Exposure to information and opportunity to discuss it during lecture is consistent with changes
Discounted other logical explanations?	Yes, ability and anxiety discounted	Limited opportunity due to lack of control group.

References

Guba, E., & Lincoln, Y. (1981). *Effective evaluation: Improving the usefulness of evaluation results through responsive and naturalistic approaches*. San Francisco, CA: Jossey-Bass Publishers.

Lincoln, Y., & Guba, E. (1985). *Naturalistic inquiry*. Newbury Park, CA: Sage Publications Inc.

Punch, K. F. (2009). *Introduction to research methods in education*. Los Angeles: Sage.

Rennie, L. J., & Johnston, D. J. (2004). The nature of learning and its implications for research on learning from museums. *Science Education, 88*(Suppl. 1), S4-S16.

Rennie, L. J., & Williams, G. F. (2000). The nature and measurement of learning from a lecture series in human genetics. In P. J. Fensham & R. Cross (Eds.), *Social responsibility of science and citizenship: Implications for education* (pp. 17-34). Melbourne: Melbourne Studies in Education (Special Issue).

Sly, J. L. (2000). *Computer managed learning assessment in higher education: The effect of a practice test*. Unpublished Doctoral Dissertation. Curtin University, Perth, Western Australia.