

Experimental Ethnography: The Marriage of Qualitative and Quantitative Research

Lawrence W. Sherman, University of Pennsylvania, and
Heather Strang, Australian National University

Background

Building ethnographic methods into the separate branches of randomized controlled trials (RCT) could substantially increase the range of conclusions that can be produced by experimental research designs as well as by ethnographic methods.

- Experimental designs offer greater internal validity for learning the effects of a social program.
 - Ethnographic methods offer greater insight into why the effects were produced.
 - The prospects for such integration depend on the capacity of two different communities within social science to work together.
 1. Some conclusions are hard to accept when they are stated only in terms of numbers.
 2. If the numbers are linked to stories about program effects they may be more understandable
 3. Social scientists themselves may be more likely to discover the truth about program effects, and to reach more understanding about why effects may vary across individuals exposed to a program.
-

Proposal

The authors propose that social science unite the insights of stories and numbers by introducing ethnographic methods into the best-known research design.

Note: This work could be called “experimental ethnography.” This strategy can achieve experiments that create both a strong “black box” test of cause and effect, and a rich distillation of how those effects happened, person by person, case by case and story by story.

History

The branch of social science that had first advocated the integration of stories and numbers was program evaluation.

- During the 1960s program evaluation scientists created textbook doctrine about the integration of *process* evaluations and *impact* evaluations.
 1. Process evaluations were designed to tell the story about how a program was or was not implemented as planned.
 2. Impact evaluations were designed to analyze numbers measuring the effects that the program caused to occur.
 - In practice, so many programs in the 1960s broke down during implementation that there was only a story to tell, but no numbers.
 1. This experience led to a rising distaste for impact evaluations among the more story-centric process evaluators.
-

Restorative Justice Example

One account that illustrates people’s interest in stories as well as numbers is found in a series of tests of a program called “restorative justice (RJ).”

- Restorative justice typically consists of a two to three hour conference between an offender and the victim led by a police officer or other trained facilitator.
 1. The agreements reached at these conferences may:
 - Take the place of formal court sentencing as a diversion from court,
 - Inform court sentencing prior to judges’ deciding sentence, or
 - Follow court sentencing as a supplement to prison or probation.
 - The story describes the way the program works, and sets the stage for proposing the structure of an experimental ethnography research design
 - The research question is whether restorative justice can change people’s lives for the better and can the effects persist over the life-course, rather than a few months or years.

Note: This research can be done both quantitatively and qualitatively, although research funding tends to be limited to quantitative measures.

Uniting Stories and Numbers

It is only possible to draw conclusions about the effects on victims involved in restorative justice by comparing their outcomes with those who volunteer for the program, but are not randomly selected to participate.

- Ethnographic studies of programs rarely make these comparisons, qualitatively or quantitatively – they focus on the before and after comparison interpreted through the language of a participant.
- The resulting hypotheses cannot be tested rigorously within the cases participating in the program because there is no comparison to those who do not participate.

- Only a comparison of what happens in the absence of a program can rule out the possibility that a benefit would have occurred anyway.
 - The great strength of an RCT design is that it allows us to estimate the average effect across the entire range of cases in treatment and control groups and nothing in the design requires that results be measured in numbers.
 - It is the logic of the RCT design, and not any particular method of data collection that gives it internal validity, as long as most cases are treated as randomly assigned, and as long as all cases are analyzed by intention to treat rather than by treatment actually received.
 - In prospective ethnography it is possible to both develop and test grounded theory as the research progresses.
-

Comparisons

What RCTs usually fail to do is to make enough comparisons. In the case of the restorative justice story there would be three possible comparisons.

1. Between all volunteers and all refusers.
2. Between those volunteers who attended RJ conferences and those who did not volunteer.
3. Between volunteers assigned to the control group and refusers

Note: In the case of restorative justice, the fact that both offenders and victims must give consent multiplies the possible comparisons. All three would be biased by self-selection, and would lack the internal validity of a randomized design. But the comparisons would yield insights from qualitative exploration that could be important in understanding the limits, or even the potential of the program being tested.

How Many Cases?

The question of “how many cases” bedevils qualitative social science.

- The direct tradeoff between depth and breadth is unavoidable, given a fixed amount of resources.
- The tradeoff is either very detailed insights on a small, perhaps atypical, sample of people versus less detailed insight on a larger, and perhaps more typical, cross-section of people relevant to the research question.

- In the case of experimental ethnography, the question of how many cases can become more sharply focused
 1. The sampling frame and the universe are identical for experimental ethnography, a truly rare opportunity for ethnography to be based on systematic sampling methods.
 2. The challenge of creating probability samples remains immense, even for relatively small experiments.
 3. The ratio of potential cases to available resources may result in a sample that is too small to create a stable estimate of population parameters.
-

The Solution

The solution to this problem of how many cases is to focus entirely on the cases in the randomly assigned groups.

- By design, five offenders could come from the randomly assigned group and five from the control group of each ethnographer's sample, allowing each ethnographer to make "grounded theory" comparisons.
 - The cost-benefit ratio would be especially high for evaluating a treatment like RJ in which the accumulation of findings shows highly diverse and unpredictable responses of different kinds of people to the same treatment.
-

Pros and Cons

There are several pro and con arguments for focusing on the randomly assigned experimental sample.

- The main argument is that it would increase the internal validity of the test.
 1. The combination of the exterior "black box" (quantitative) ruling out rival hypotheses and qualitative descriptions within the "black box" would be unprecedented at the level of cases as randomly assigned.
 - The argument against focusing only on the experimental sample is that it would limit the external validity that could be gained from experimental ethnography.
 - The external validity argument is simultaneously strengthened and challenged in situations of high refusal rates.
 - A justification of ethnography would be the generative function of identifying research questions.
 1. What the ethnography could do is explore how the kinds of people who refused, their circumstances, and the circumstances of their offense differed from the kinds of people and circumstances where consent was forthcoming.
-

Another Option - Memoirs	One option for adding qualitative information to the universe of cases in the experimental pipeline would be to ask each operational person seeking consent to prepare a short memoir of each case. The article describes the procedure and questions involved in producing a memoir.
Systematic Reviews	<p>A further justification for focusing ethnography on the experimental sample is the increasing use of systematic reviews of randomized trials.</p> <ul style="list-style-type: none"> • With each experiment seen as merely one tree in a forest of evidence, the importance of knowing how each tree compares to all others becomes even greater. • Rather than writing off the average effect of a series of RCTs as negligible or negative, it may be more useful to isolate the one or two most successful RCTs and determine how they differed from the majority. • It may be chance, but there may also be substantive differences in the sample or the way the treatment was delivered that could help explain the difference in outcomes. • Those substantive differences may even point to refinement of the treatment for future RCTs, treating the accumulated knowledge as a trial-and-error process of invention rather than as a verdict based on the average result.
New Methods to Accommodate Data	<p>At present, the methods of systematic reviews do not accommodate the kinds of data that ethnography could generate. Yet there is no reason such methods could not be developed.</p> <ul style="list-style-type: none"> • RCTs that enjoy the benefit of experimental ethnography could be coded in very rich ways, such as the tendency of the control group to resent that they were not randomly assigned to the treatment group. <p>Note: Since the issue of control group reactivity is one of the most fundamental threats to the external validity of RCTs, this task should be reason alone for funding experimental ethnography.</p>
Getting Systematic Reviews Right	<p>The rising power of systematic reviews to dominate the policy conclusions about innovative ideas makes it all the more important to get them right.</p> <ul style="list-style-type: none"> • While some quantitative measures can examine the consistency of black box causal mechanisms, few would doubt the increased insight that could be gained from ethnographic materials.

- Some may argue about how large a sample would be needed to generate reliable insights about the causal processes operating with the randomly assigned group.
 1. The possibility of placing an ethnographer with every research participant should cut short that argument.
 2. Even with a fifty percent sample in RCTs of fifty cases per treatment group, many would place great credence in consistent conclusions found across a research team of five or ten ethnographers, each studying both experimental and control cases.
 - The more detailed the descriptive material published about each RCT, the greater the possibility for systematic reviewers to comb the details looking for ground theory about why some RCTs produced better outcomes in the treatment group than other RCTs did.
 - Given the greater likelihood that more rigorous methods will find that promising ideas are ineffective, the risk of prematurely rejecting promising treatments may rise along with evaluation rigor.
-

Conclusion

This proposal requires two necessary and sufficient conditions to become a reality.

1. Ethnographers and experimentalists must be willing to work together.
2. There must be sufficient funding.

Note: Those who attack RCTs say they cannot tell why something worked. Those who attack non-experimental evaluations say they cannot tell what worked. Rather than choosing between the two, experimental ethnography could provide the win-win answer that works.