When Life Gives You Lemons: What to Do When Something Goes Wrong in Your Carefully Planned Research and How to Avoid Disasters in the First Place

Sheila T. Murphy

Perhaps it is because my last name is Murphy and I have lived my entire life cursed by Murphy's Law that "Whatever can go wrong will" that I seem to be a magnet for "research gone wrong" scenarios. As a result, I have become skilled at planning research in such a way that even if I cannot avoid research catastrophes altogether, at least I can salvage something valuable from the smoldering ruins of what was a perfectly designed study. In this chapter, I draw on actual examples from myself and my EE colleagues to illustrate key lessons that I have learned over the years.

Lesson 1
Think ahead of all the possible outcomes (both intended and unintended) your intervention may have.

S. T. Murphy (✉)
Annenberg School for Communication and Journalism,
University of Southern California, Los Angeles, CA, USA
e-mail: smurphy@usc.edu

One of the earliest attempts at leveraging partnerships with people who write, produce and broadcast programs for entertainment-education (EE) purposes was a partnership between Population Control International, Televisa, and a very talented writer/producer, Miguel Sabido. When these parties first met in 1977, Miguel was already using a social content communication methodology that was based on Albert Bandura's Social Learning Theory (SLT) which was subsequently renamed Social Cognitive Theory (SCT). Social Cognitive Theory—which states that audience members are much more likely to engage in a behavior they have seen being performed by someone they like and/or who is similar to them— remains the theoretical backbone of EE (Bandura, 2004). Miguel was directly responsible for several telenovelas (dramatic televised stories) that significantly reduced the Mexican birthrate, while increasing the sales of condoms and oral contraceptives as featured in *Acompáñame*. This was the first of a string of successes for Miguel Sabido, who is sometimes referred to as the "grandfather of entertainment-education" (see Sabido, 2021).

But Miguel did produce one earlier storyline whose results were more mixed. In 1975, Miguel created a telenovela, *Ven Conmigo* (*Come With Me*), that aimed to promote adult literacy by featuring a story that revolved around characters enrolled in an adult literacy class at their local library. One episode mentioned the national distribution center that provided free literacy booklets. The very next day, over 25,000 people showed up to get their booklets, which ran out after the first thousand. No one, including Miguel, had foreseen this problem. This taught Miguel a valuable lesson— don't get audience members motivated to change without making the necessary resources available. Being disappointed is a negative experience that may undermine an entire EE campaign.

I have made similar miscalculations. Lourdes Baezconde-Garbanati and I thought we were brilliant when we designed a beautiful campaign *Es Tiempo* (It's Time) based on the blooming of the jacaranda tree to remind women to get screened for cervical cancer in East Los Angeles. The campaign worked a little too well and when the jacarandas bloomed, Latinas from all over Los Angeles overwhelmed local clinics. They ended up scheduling appointments up to 6 months in advance, by which time the jacarandas were bare and fewer women remembered or were motivated to keep their appointments.

Lesson 2
Are there any key variables (e.g., gender, age, marital status, health, current behaviors, etc.) or confounding variables (e.g., lack of insurance results in fewer pap tests) that may strengthen or undercut the impact of your campaign? If so, be sure to measure them!

For another project, my colleague Lourdes Baezconde-Garbanati and I set out to design and conduct a large-scale quasi-experiment or "clinical trial" that would directly test the relative efficacy of the same health-related information presented in either a narrative or nonnarrative format. To determine the relative power of narrative over nonnarrative we deliberately chose a story about a young girl's Quinceañera, or 15th birthday, traditionally celebrated in many Latinx households in Los Angeles where our study was conducted. Additionally, both the narrative film—*Tamale Lesson*—and nonnarrative film—*It's Time*—featured primarily Latina actors. As a result, we predicted that not all women in our study would be similarly impacted by the films. Rather we predicted that, particularly for the narrative, the Mexican American women in our sample would identify the most with the characters, be the most transported into the story and as a result would show the greatest impact in terms of shifts from pretest to 6-month-posttest in cervical cancer-related knowledge, attitudes and behavior. And they did (Murphy et al., 2015).

In this study we deliberately included ethnicity as a factor in our quasi-experimental design and had equal numbers (300 each) of African American, European American and Mexican American women. However, other things could have mattered as well. For example, what if education level mattered in how women reacted? What if insurance coverage made women less likely to pay attention to *Tamale Lesson* because they could not afford to go to a doctor? Or income?

To address such potential confounds you have to make sure to include any variable that might make a difference in your study at the start. These are often what I call "the usual suspects" or standard demographics like age, gender, education level, marital status, number of children, ethnicity and others that may be specific to your study like acculturation level. I can't tell you the number of times I've been brought into an evaluation of an EE campaign AFTER the data was already collected and asked to try to analyze the data. It never ends well.

Lesson 3
Work with a cultural advisor to help you develop your narrative, measure impact, and avert disaster.

Researchers and EE producers need to remember that for many EE projects you will work on, you are not a member of your target audience. For instance, even when I am designing an intervention for women in Los Angeles, I realize that my education level, ethnicity, age and socioeconomic status make it very unlikely that I can predict the impact of an intervention to increase the willingness of teenage Latinas from East Los Angeles to get vaccinated against HPV. Luckily, I am smart enough to realize I am clueless with respect to this population and need one or more cultural advisors, as the following example illustrates.

During our formative research for the *Tamale Lesson*, we conducted a survey that revealed that the two most frequently mentioned barriers for Mexican American women in Los Angeles were time and money. Some well-meaning soul at the National Cancer Institute decided that the obvious solution was to employ a fleet of medical trucks that would have the typical set up for a pap test inside. On the surface this made sense. If Latinas are too busy to come to a clinic to get screened for cervical cancer, then take the clinic to them. A woman could make an appointment, and the truck would pull up outside her house, and the woman could have her pap test there and then. NCI was rather proud of this solution and had made initial enquires into the purchase and outfitting of several medical vans to service East Los Angeles.

Fortunately, before they moved forward, I ran some focus groups of Latinas who had not been screened for cervical cancer in the past two years. At first, the women seemed fairly positive about the truck idea. But soon I began to notice snickering among the group which then erupted into uncontrollable laughter. I failed to see what was so funny. Finally, one embarrassed participant told me that one of the slang terms for a woman's genitals was taco. So what NCI was proposing was essentially a "taco truck," a term typically applied to the ubiquitous food trucks in LA. From that point on in the group, the participants became far less constrained in their opinions and revealed that while it may seem like a good solution, neither they nor their family and friends would ever consider entering such a truck for a whole host of reasons such as neighbors suspecting you had a sexually transmitted disease and explaining what the pap test actually entailed. And that was the end of NCI's taco truck.

Lesson 4
Pilot your materials with actual members of the intended target audience.

Even though we had a number of cultural advisors when developing two films to increase cervical cancer-related knowledge, attitudes and behavior, *Tamale Lesson* and the nonnarrative *It's Time*, it was still essential to pilot the intervention with actual members of the target audience. In this case our target audience was Latinas between the ages of 21–45 living in Los Angeles who had not had a cancer screening using a pap test (or *Papanicolaou*) in the past three years. We recruited several focus groups of eight to ten women each to watch the 11-minute films through and then discuss.

We discovered several problems with both our narrative and nonnarrative film that needed to be addressed. In *Tamale Lesson*, one of the main actresses used heavy eye liner. While this meant nothing to me or our cultural advisors, it was a clear sign of association with a local gang which made audience members wary and unlikely to identify with her. As EE experts you realize that if the audience fails to identify with the key character demonstrating the main behavior (here getting a pap test), the impact of the story is substantially weakened. As a result, we reshot with another actress before conducting the larger study.

We also realized a second problem during piloting. In the nonnarrative we used percentages to discuss the relative risk of cervical cancer among the different ethnic groups in our study. These formative focus groups revealed that some of the Latinas in our sample were unfamiliar with the concept of percentages, so we changed the script to avoid the term "50 percent" instead saying "almost half of women have HPV at some point in their lives." When measuring the impact of the films on normative beliefs we likewise avoided percentages, instead asking "Out of one hundred women like you, how many do you think would have her daughter get vaccinated against human papillomavirus (HPV)?"

Lesson 5
Be aware of numeric literacy, random assignment and witchcraft.

This unfamiliarity with percentages is part of a larger methodological issue known as numeracy or numeric literacy. Not only are percentages often problematic, particularly in developing countries, but so are the concept of interval scales which we in the West use every day (e.g., On a scale

from 1 to 10, where 1 means strongly disagree and 10 means strongly agree). This makes measuring the impact of an intervention challenging, to say the least. One workaround is using verbal labels for each response option (e.g., Would you say you strongly disagree, somewhat disagree, neither agree nor disagree, slightly agree or strongly agree).

In addition to having little familiarity with percentages and interval scales, individuals in other cultures may be completely unfamiliar with the concept of random assignment. Two colleagues, Paul Falzone and Paul Sparks, gave the following account of using dice to randomly assign participants to receive different versions of an intervention. They were attempting to pilot an early version of Wanji Games, an interactive narrative format to teach health and livelihoods skills in the Teso region of Uganda. They had more participants than they needed so one of the researchers, Paul Sparks, brought out dice to help with random selection of participants. After some discussion in their native language, villagers were now refusing to participate, and wanted the researchers to leave. Moreover, the villagers later became uncooperative with other researchers saying that the Wanji Games folks tried to use witchcraft on them (referring to the dice)! The moral of this story is to remember that something that makes perfect sense methodologically to you (e.g. using dice for random assignment) may seem odd or suspicious to others.

Lesson 6
Use a "control group" to account for historical confounds.

One of the most common critiques of entertainment-education is that many projects, particularly early attempts, lacked a rigorous evaluation of their impact by an unbiased researcher. In the early days of EE, if there was any evaluation of impact at all, it was often an afterthought done by the same team that designed the narrative intervention. More recent EE projects have made a strong, unbiased, quantitative evaluation a requirement for funding. However, many EE evaluations still rely on a posttest only design where a sufficiently large sample of the target audience is either exposed to the narrative (experimental group) or not (control group).

But is exposing the control group to nothing always fair? Sometimes not. As most textbooks will tell you, almost all experimental designs contain a control group to whom the intervention has not been given (or, if appropriate, a placebo version lacking the active ingredient). So, for

example, to test whether embedding information in a story or "narrative" leads to stronger, longer-lasting effects in knowledge, attitudes and behavior we had to create a control nonnarrative that had the same facts but presented them in a nonnarrative format. Control groups allow researchers to account for the effect of history or other uncontrollable events that might impact your outcome measures.

A control group becomes essential if you have a posttest only design. One project I was involved with tried to "normalize" condom use in India. This was particularly challenging because condoms were kept behind the counter and had to be requested from the pharmacist and, to make matters worse, just saying the word "condom" in India was extremely taboo. The BBC developed an award-winning campaign of small vignettes run as Public Service Announcements (PSAs). Each PSA in the series used humor to position a male condom user as smart and another man who is scandalized by hearing the word "condom" in public look foolish (these PSAs can be found on BBC Media Action's website). We carefully designed the impact study by releasing the series in certain television markets first, deliberately withholding the PSAs in other comparable areas to serve as our controls. Unfortunately for our experimental design, the PSAs became immediately popular, particularly one that allowed individuals to download a free "condom a cappella ringtone" which four million viewers did. The Indian government magnanimously decided that all of India should be allowed to see the PSAs immediately. There went our carefully constructed control group! We were forced to come up with a measure of how much young sexually active men were exposed to the campaign (how many they remembered, etc.). The research design was not ideal, but the campaign ultimately increased condom sales in India around 8 percent (Frank et al., 2012).

Lesson 7
Plan a pretest-posttest design (as opposed to a posttest only design), so you can salvage a study when something happens in the middle of your data collection.

The gold standard for establishing causality, however, is showing change at the individual level (in other words, surveying the same individuals twice—once at Time 1 to establish a pretest baseline measure and then again after the experimental intervention at Time 2 or posttest to assess the degree to which each person has changed their knowledge, attitudes and behavior). Although reaching the same person twice can be

challenging outside of the laboratory, it can control for the effects of "history" or the impact of something occurring outside of your intervention that may nevertheless impact your posttest measures.

Just such a "historical" effect occurred in June 2015, while my colleagues at Hollywood Health and Society (HH&S) Erica Rosenthal and Kate Folb and my student Traci Gillig and I were conducting a study designed to measure the impact of a transgender storyline on a medical show *Royal Pains*. The producers of *Royal Pains* alerted HH&S of an upcoming storyline that featured a transgender actress playing a transgender 16-year-old girl, Anna, who experiences health complications while self-administering estrogen in order to transition from male to female.

Approximately two weeks prior to the episode airing we collected pretest levels of attitudes toward transgender individuals, rights and policies such as sharing restrooms from 488 regular *Royal Pains* viewers. Our plan was to conduct a posttest immediately after the story aired on June 24th—with viewers who watched the episode the night it aired acting as our experimental group and those viewers who had not seen the episode serving as our control group.

Unfortunately for our well-laid plans, the Royal Pains transgender storyline aired the same month that former Olympic swimmer and reality TV star Caitlyn Jenner announced her transition (as described in more detail in Rosenthal & Folb, 2021). How could we ever disentangle the impact of the Jenner announcement from the effects of our *Royal Pains* storyline? After the initial panic subsided we realized that we had individual pretest data for each participant. Since we were going to analyze change in transgender attitudes from pretest to posttest, perhaps all was not lost. We added items to the posttest to measure the degree of exposure to Jenner's announcement and other transgender storylines (such as *Transparent*). Luckily, our results showed that the Jenner announcement had not swamped the impact of our storyline and we were able to separate out exposure to the Jenner announcement from our transgender episode. Viewers who saw Anna's transgender storyline reported more supportive attitudes toward transgender people and related policy issues (such as transgender high school students should be able to use the restroom that matches their gender identity) than those who did not see that episode. And interestingly, attitude change was cumulative across different transgender portrayals on different shows suggesting that the frequency of sympathetic portrayals matters (Gillig, Rosenthal, Murphy, & Folb, 2018). The moral of this story is that what at first glance appears to be a

methodological lemon is much easier to salvage with individual level pretest-posttest design that allows you to measure change before and after the intervention at the individual level (as well as exposure to any potential confounds here the Jenner announcement post).

Lessons Learned and Best Practices

In the early days of entertainment-education, resources were often funneled almost exclusively into making the best possible narrative intervention with little thought given to evaluating its impact. Everett Rogers and his former student Arvind Singhal were involved in early ground-breaking EE projects such as *Hum Log (We People)*, the first serial or "soap opera" broadcast in India beginning in 1984 and continuing for over 30 years. *Hum Log* revolved around a middle-class family's struggles and aspirations. But along the way, viewers learned about adult literacy, contraception and a legion of other social issues. When asked how they knew *Hum Log* was having the desired impact on viewers, the producers and research team pointed to the over 400,000 letters they had received. While this outpouring is incredibly impressive, it would not be sufficient for many current funders of EE projects. Today it is often required that an EE project should include a well thought out quantitative evaluation strategy, preferably conducted by independent researchers who will present an objective view of the project's impact. The previous "lessons learned" were designed to be one small step toward helping help EE practitioners do just that.

The "lessons learned" discussed above can essentially be divided into one of two buckets. The first bucket requires understanding what type of quantitative evidence funders and journal editors expect in order to demonstrate your intervention produced the intended impact. These include strong study design (lesson 7) involving a control group who did not receive the intervention when appropriate (lesson 6), avoiding—or at least accounting for—potential confounds (lesson 2), as well as measurement issues (lesson 5). These methodological and measurement and statistical issues are perhaps the most straightforward and easiest to learn. One could take online classes in research methods and statistics or identify successful EE projects by researching articles subsequently published in peer-reviewed journals that describe the project and intervention measures in detail. Failing this, if no one on your team has experience in methods, measurement and statistics, I strongly recommend bringing on a

well-respected individual or team to help oversee the evaluation and analyses. It is vital that the evaluation should be integral to the intervention and be one of the earliest things you focus on, not an afterthought. If you are uncertain that your evaluation captures key constructs you could ask someone whose work in EE you admire to look over your proposed measures before you go into the field with project. Remember that your evaluation *will* be critiqued at some point—it is up to you whether that critique is a biopsy that identifies and removes problems at an early stage or an autopsy when data is already collected and nothing more can be done.

The second bucket involves common sense, something that can be orthogonal to academic achievements. Thinking ahead about possible outcomes both intended and unintended of your intervention (lesson 1) requires viewing the intervention through your target audience eyes and situation. Because much EE research is funded by international agencies and conducted by researchers from other cultures, it is essential to not only acknowledge ignorance but actively fill any relevant knowledge gaps by hiring cultural advisors (lesson 3) and piloting your intervention and measurement materials with members of the intended target audience (lesson 4). Remaining humble and keeping a sense of humor always helps. After all, what can possibly go wrong? And on the bright side, at least your last name is not Murphy.

References

Bandura, A. (2004). Social cognitive theory for personal and social change by enabling media. In A. Singhal, M. J. Cody, E. M. Rogers, & M. Sabido (Eds.), *Entertainment-education and social change: History, research, and practice* (pp. 75–96). Routledge.

Frank, L. B., Chatterjee, J. S., Chaudhuri, S., Lapsansky, C., Bhanot, A., & Murphy, S. T. (2012). Conversation and compliance: Role of interpersonal discussion and social norms in public communication campaigns. *Journal of Health Communication, 17*(9), 1050–1067. https://doi.org/10.1080/10810730.2012.665426

Gillig, T. K., Rosenthal, E. L., Murphy, S. T., & Folb, K. (2018). Do transgender portrayals change hearts and minds? The influence of media narratives on viewers' attitudes toward transgender people. *Sex Roles, 78*, 515–527. https://doi.org/10.1007/s11199-017-0816-1

Murphy, S. T., Frank, L. B., Chatterjee, J. S., Moran, M. B., Zhao, N., Amezola de Herrera, P., et al. (2015). Comparing the relative efficacy of narrative vs nonnarrative health messages in reducing health disparities using a randomized

trial. *American Journal of Public Health, 105*(10), 2117–2123. https://doi.org/10.2105/AJPH.2014.302332

Rosenthal, E. L., & Folb, K. L. (2021). Entertainment-education, American style: Informing and studying Hollywood's portrayals of social issues. In L. B. Frank & P. Falzone (Eds.), *Entertainment-education behind the scenes: Case studies for theory and practice*. Palgrave Macmillan.

Sabido, M. (2021). Miguel Sabido's entertainment-education. In L. B. Frank & P. Falzone (Eds.), *Entertainment-education behind the scenes: Case studies for theory and practice*. Palgrave Macmillan.