

# Intercultural Communication Training: An Introduction

## The Evaluation of Training Programs

Contributors: Richard Brislin & Tomoko Yoshida  
Book Title: Intercultural Communication Training: An Introduction  
Chapter Title: "The Evaluation of Training Programs"  
Pub. Date: 1994  
Access Date: October 15, 2013  
Publishing Company: SAGE Publications, Inc.  
City: Thousand Oaks  
Print ISBN: 9780803950757  
Online ISBN: 9781452243603  
DOI: <http://dx.doi.org/10.4135/9781452243603.n6>  
Print pages: 144-172

This PDF has been generated from SAGE knowledge. Please note that the pagination of the online version will vary from the pagination of the print book.

<http://dx.doi.org/10.4135/9781452243603.n6>

[p. 144 ↓ ]

## Chapter 6: The Evaluation of Training Programs

### Reasons for Interest in Evaluation

Don Robbins was a professor in the Department of Management at a large state university. He taught courses in management development, human resources management in global perspective, and (through a cross-listed course in the Sociology Department) program evaluation. In addition to his teaching, research, and writing, he also carried out some consulting work with large firms in the city where he lived. He was also active in community activities, and many of these activities centered around his membership in the Rotary Club. Although enjoying the Rotarian lunches and club activities for their own sake, he also realized that other club members were prominent businesspeople who might have information about possible consulting jobs for him. One day, Don received a telephone call in his office. The conversation went like this:

**Caller:** Don? Jack Pierce from the Rotary Club. I'm new in town, but we met at the last meeting. Do you remember my telling you that I'm with Barnsworth Manufacturing? We have a new training program here, and I hope you can help us out with it.

[p. 145 ↓ ] **Don:** What are the goals of the training program?

**Jack:** We are expanding into overseas markets in Germany, Indonesia, and Korea. The program is designed to help prepare managers, and their families, for the challenges they will face on overseas assignments.

**Don:** That sounds very forward looking. A lot of American companies don't pay much attention to intercultural training, thinking that doing business overseas is pretty much the same as doing it in the United States.

**Jack:** Well, I think so too. I was a participant in a Junior Year Abroad program in college, and I studied in France for a year. There was no special preparation for the students who participated, and let me tell you it was hard to adjust to the differences between studying in the United States and studying in France. So when I was in a position to introduce cross-cultural training at Barnsworth, I jumped at the chance.

**Don:** I cover cross-cultural training in my course on human resources management. How can I help you?

**Jack:** You can help me with evaluating the program. I don't know much about evaluation. The top brass here wants to know if the cross-cultural training program is any damn good and whether or not it's worth the cost.

**Don:** High-level executives often talk that way. They will also probably want to know whether or not people's performance, when stationed overseas, is better as a result of having been trained.

**Jack:** I have to write the executives a memo, within about 2 weeks, persuading them that the training program is worth continuing.

**Don:** I'll see if I can help with the contents of the memo you have to write. Can I ask you some basic questions now?

**Jack:** Sure, go ahead!

**Don:** Where are the people who will be trained?

**Jack:** We had the training program about 6 months ago, and the people are now in the three countries: Germany, Indonesia, and Korea.

**Don:** Was any information gathered from them before the training, and was this compared with information gathered after the training?

**Jack:** Not really. A lot of the program participants, though, told me that they liked the training and appreciated the opportunity to think about their upcoming overseas assignments.

**[p. 146 ↓ ] Don:** Was there a control or a comparison group, that is, were the people who received training compared with people who didn't receive training?

**Jack:** No, the company gave people time to participate, and all but two or three people who were to go overseas received training.

**Don:** Do you have figures on how much it costs to send someone overseas, and how much it costs to bring a family back to the United States if the assignment doesn't work out for some reason?

**Jack:** I've never seen figures like this. I suppose I could ask someone in the accounting office whether or not we could scrape something together.

**Don:** Do you have information on how many people, in past years, have not been successful on overseas assignments? For example, how many had to "be pulled out" before their assignments were completed?

**Jack:** Sorry, I've never seen information like that. It might not exist because Barnsworth hasn't been assigning people overseas for too many years.

**Don:** Can we get access to information about how the trained people are doing on their overseas assignments. For example, do we have access to ratings by supervisors in Germany, Indonesia, and Korea?

**Jack:** I don't think so. That information is considered pretty confidential at Barnsworth, and we'd have to get each trainee's permission to obtain

access. Listen, Don, I know I haven't been able to answer "yes" very much, but I hope you can help out. I really want to see the training program continued, but I don't think it will be unless the memo I write on evaluation gets a good reception.

Don has probably participated in similar conversations with other people who are interested in starting or continuing inter-cultural training programs. People are rarely interested in evaluation until some outside force, usually a force with power, demands information on how good, how effective, and how helpful the program has been. When people receive demands for information on evaluation, they often turn to specialists and ask them for help. The requests for help, however, often seem to be appeals to magicians who can wave magic wands and produce proof about program quality and effectiveness. In the incident, is there anything that Don can do? The answer is, "No, [p. 147 ↓ ] not very much." One of the most important points about the development of good cross-cultural training programs is that evaluation specialists should be integrated into program planning from the very earliest stages. When integrated early, evaluation specialists can point to opportunities to gather information about (a) improving programs and (b) eventually developing persuasive arguments about their effectiveness.

## Conversations between Intercultural Communication and Evaluation Specialists

When they participate in efforts to evaluate programs, intercultural communication specialists try to answer such questions as:

These are not easy questions to answer, and it takes a great deal of education and experience to deal with them effectively. Program evaluation is a specialized topic that is covered in a year-long series of graduate-level seminars at some universities. Excellent texts, literature reviews, and handbooks have been published (e.g., Berk & Rossi, 1990; Cook, Campbell, & Peracchio, 1991; Guttentag & Struening, 1975; Rossi & Freeman, 1989). The year-long seminars can involve upwards of 4,000 [p. 148 ↓ ] pages of reading as well as an internship in an organization where graduate students gain practical experience. Often, they evaluate a program within the organization as part

of their internship. Given this amount of reading and needed experience, only limited goals can be achieved in this short treatment. The purposes of this chapter are to introduce some key ideas about program evaluation so that intercultural communication specialists can (a) have good discussions with and (b) integrate the talents of evaluation experts. Three topics will be discussed.

In the discussion of key points, research studies that have focused on the evaluation of actual intercultural communication training programs will be reviewed.

## Important Concepts for Discussions about Program Evaluation

When evaluation specialists sit down with administrators who are planning intercultural training programs, the following concepts will surely become central to their discussions.

### Formative and Summative Evaluation

Now a familiar distinction (Birnbauer, 1987; Guttentag & Struening, 1975; Mathison, 1991), formative and summative evaluation refers to two different types of information that program planners can use. When efforts are focused on formative evaluation, information is gathered that will lead to program [p. 149 ↓ ] improvement. This improvement can occur within the time frame of one program (e.g., a “mid-course correction” taken at the end of the first day of a 2-day program), or the improvement can occur after a given program so that the next offering is more favorably received. The phrase “more favorably received” is carefully chosen because many formative evaluation methods focus on the responses of program participants. The most basic question is, “Do they find the material presented useful and interesting or do they view the program as a waste of time?” There are few published studies that focus solely on formative evaluation of intercultural programs, and the reason reminds us that there are different goals that evaluation specialists strive to achieve. In formative evaluation, information is presented that assists efforts to improve programs within an organization. Such information focuses on what is perceived as important, what material should

be dropped, what material should be added, what information is being learned, and so forth. Answers to these questions may be of very little interest to people in other organizations, and consequently the expense of publishing widely available journal articles based solely on formative evaluation is often a poor use of resources. Program participants often respond to questions about the following topics (Birnbrauer, 1987):

In summative evaluation, information is gathered to answer the question, “Is the program effective.” In everyday language, the question becomes, “Is the program any good?” Answers to these questions then lead to the very practical demand to present information that allows decision makers to either continue supporting a training program or dropping it because of its demonstrated ineffectiveness. Because effectiveness is conceptualized as helping people become more skillful communicators in interactions that transcend cultural barriers, the most impressive summative information is gathered after training ends. Good summative information can be gathered on how effective people are in communicating with culturally diverse individuals (as judged by those individuals), how long it takes them to perform effectively on their job in the other culture, how effective people are in solving problems, how they deal with the stress of intercultural living, whether or not they complete the full terms of their overseas assignments, and so forth.

Occasionally in all kinds of training programs, but perhaps more frequently for intercultural training (see [Chapters 4](#) and [7](#)), the results of formative and summative evaluations will be different. People may enjoy the training program (Trost, 1985) and report that it was worthwhile (formative), but they may not change anything about their work, their beliefs, their attitudes, or their behavior (issues addressed in a summative evaluation). Or, people may report that they disliked the training but later integrate much of what the trainers recommended into their actual behavior during their overseas assignments. Trifonovitch (1977) reported this pattern. As discussed in [Chapter 5](#), Trifonovitch was in charge of programs that prepared people for assignments on remote Pacific islands. During training, they had to live in a manner similar to Pacific Islanders. Many American trainees did not like this one bit as they could not believe (during [\[p. 151 ↓ \]](#) training) that they would have to gather their own food, make their own entertainment, and work without familiar technological aids. Consequently, they complained bitterly during training, “burning out” some of the staff (as discussed in [Chapter 3](#)). Years later, however, Trifonovitch received unsolicited letters from people,



the same ones who complained bitterly, expressing how well prepared they were for their intercultural assignments. More information about summative evaluation will be presented in a later section of this chapter that deals with the measurement of program outcomes.

## Control and Comparison Groups

Any discussion of summative evaluation leads to the analysis of control and comparison groups. The reason is that the most compelling evidence for the effectiveness of a program is often gathered by showing that people who have received training behave in a more effective manner than people who have not received the training. People who have not received training are referred to as members of the *control group* or the *comparison group*. The distinction between these two terms involves the amount of control the evaluator has in determining whether people receive training or not. The most straightforward and persuasive evaluation studies involve random assignment to either the training or no-training group (Campbell & Stanley, 1966). In random assignment, every person who might be a part of the evaluation study has an equal chance of being picked for the training group. Because the assumption behind randomization is that the resulting training and no-training groups are equivalent, the only differences people would eventually demonstrate on various measuring instruments must be due to training. When the evaluator can employ randomization in the selection of training program participants, the people who do not receive training are known as members of the control group. When evaluators do not have control over the decision whether people receive training or not, they often search for a similar group of people who have not received training. This group of [p. 152 ↓ ] people who bear similarity to the training group, but who were not randomly assigned to a no-training group, should be called a comparison group.

As an example of the search for a comparison group, Befus (1988) offered intercultural training to 64 North Americans who were studying Spanish at the *Institute de Lingua Española* in San Jose, Costa Rica. For a number of reasons, it was difficult if not impossible to randomly assign 32 of the students to a training group and 32 to a control group. One reason was that the 64 students interacted with each other frequently, and any helpful information that might have been presented to the training group participants

would have been shared with friends in the control group. Given that she could not form a control group, Befus searched for the possibility of a comparison group. She developed one by working with North Americans who had been students at the *Institute de Lingua Española* during the previous school term. These students had presumably been faced with cultural differences in Costa Rica that demanded adjustment and that perhaps caused culture-shock reactions, but they had not participated in a training program. Befus compared reports of stress (see [Chapter 3](#)) among members of the training group with the reports among members of this comparison group. Given that members of the training group reported less stress subsequent to the program, Befus argued that her program was effective.

## Using Control Groups When Possible

While the formation of comparison groups can seem at first glance to be reasonable, and while at times they are essential to carrying out the best evaluation efforts possible, there are difficulties that they bring. The major difficulty is that people in a comparison group may be different to start with, and consequently positive effects that might be attributed to training may be due to qualities of the people in the training group. This problem can exist even when efforts are made to make the training and comparison groups as similar as possible. For [\[p. 153 ↓ \]](#) example, Befus (1988) took steps to ensure that her groups were similar. People in the two groups were similar in terms of the number of males and females, “marital status, amount of previous cross-cultural exposure, level of education, age, religious affiliation, previous language experience, and amount of exposure to subcultures within North America” (p. 391). Although this is an impressive list, differences in stress levels could be due to other factors not addressed by Befus. For example, people in the comparison group could have been different in terms of aspects of their personalities, such as general anxiety level. This general anxiety level may have led to more self-reported stress among comparison group members than among training group members. This type of analysis admittedly sounds like carping and nit-picking, but such factors must be taken into account during serious evaluation efforts. Especially when there is competition for funds to support various programs (e.g., more Spanish language study compared to support of the

cross-cultural training program), administrators must present solid evaluative evidence to bolster their recommendations for how funds should be used.

Why aren't control groups employed more frequently than they are? One reason is that many program administrators feel that there are ethical problems with denying training to people who want it. An ideal study, from the viewpoint of sophisticated summative evaluation, would begin by identifying people who want training. Assume that 60 people sign up. Then the people would be randomly assigned to a training and to a control group, perhaps by putting all names in a box, mixing them up, and drawing them out. Thirty people would be assigned to the training group and 30 to the control group. After the program and after people have actually assumed their intercultural assignments, summative information would be gathered. The evaluator would determine whether or not the program led to benefits for the training group compared to the control group. For instance, symptoms of stress could be measured, and information on job productivity could be gathered.

**[p. 154 ↓ ]** The ethical dilemma for program administrators and evaluation specialists is that they are denying the benefits of training to some people. The basic conclusion that good training has positive benefits is well established (Black & Medenthall, 1990; Brislin, Landis, & Brandt, 1983; Cushner, 1989). Some of these benefits will be reviewed later in this chapter as part of the discussion of outcome measures in evaluation. The fact that good training leads to benefits, of course, does not mean that any one training program being currently evaluated will result in benefits. It does mean, however, that the possibility of benefits exists and that people without training may experience cross-cultural difficulties for which they could have been better prepared. Professionals involved in cross-cultural training face an ethical dilemma. They know that the use of control groups provides the most persuasive information about training's effects, but people in the control group may experience problems that could have been prevented. Are there ways of resolving this dilemma?

## The Wait-Control Design

One evaluation design, when successfully employed, involves a compromise between the need to gather good information on program effectiveness and the need to give

training to all who want it. At times, it even works into the busy schedules of program administrators and allows them to schedule their time most productively. The design is called “wait-control,” and variants of it have been used in a number of research studies (e.g., Ganster, Mayes, Sime, & Tharp, 1982; Ilola, 1989). Its features are clearest when used in an example. Assume that 60 people in an organization want to participate in an intercultural communication training program. The randomization process takes place, and 30 are assigned to a “training: time 1” group, and 30 are assigned to a “training: time 2” group. This second group is also called the “wait-control.” In a pictorial form, the design can be depicted as follows (see [Table 6.1](#)). Using the shorthand developed by Campbell and Stanley (1966), “X” refers to the [p. 155 ↓ ] training program and “O” refers to observations or measurements of program effects. In our example, let's assume that the “Xs” are the responsibility of the program administrator, that the “Os” are the responsibility of the evaluation specialist, and that Time 1 and Time 2 refer to a 2-week training program.

*TABLE 6.1 The Wait-Control Design*

Randomization	Observations: 1	Time 1	Observations: 2	Time 2	Observations: 3
Group 1	O	X	O		O
Group 2	O		O	X	O

Information is gathered from all people at the point called “Observations: 1.” For example, their ability to solve the sorts of critical incidents presented throughout this book could be assessed [an approach taken by Cushner (1989) and Ilola (1989)]. Then people in Group 1 receive the 2-week training program, while the people in Group 2 wait for training. At the point called “Observations: 2,” information is again gathered from all 60 people. At this point, there is a true training group and a true control group: If the training is effective, people in that group should score higher on various measuring instruments. For example, people in the trained group should be able to solve more critical incidents (that they haven't seen before) than people in the wait-control group (Cushner, 1989; Ilola, 1989). Then people in Group 2 receive the 2-week training program while people in Group 1 go back to their jobs within their organization. At the point called “Observations: 3,” measures are again gathered from all people. People in the second group should demonstrate increases compared to their performance at observation points 1 and 2. In addition, people in the first training group should

demonstrate that they have maintained the benefits of training as measured earlier at observation point 2.

From the organization's standpoint, the major advantage to this design is that all people desirous of training participate in a program. From the evaluator's standpoint, the advantage is [p. 156 ↓ ] that information is available about program benefits because there is a point (Observations: 2) at which there is a true training and a true control group. Further, information is available (Observations: 3) on the retention of program benefits. For many intercultural communication training programs, however, this is not a perfect design. If evaluators want to measure long-term benefits, such as job performance after overseas assignment and amount of stress experienced a year after training, the use of a wait-control can be less than ideal. If the program lasts 2 weeks, as in our example, then long-term benefits can be assessed only after all participants have received training. Long-term benefits would be assessed at a point that would be marked "Observations: 4" in the diagram, and such benefits would be assessed long after measurements gathered at the earlier "Observations: 3" point. Note that by this later time, all people have received training and there is no longer a control group.

The wait-control design is very useful and brings up issues that program planners and evaluators should discuss. The design is most applicable when measures of effectiveness can be gathered shortly after training and when there is still a control group (Observations: 2 in the diagram). Examples of such measures (more will be reviewed later in the chapter) include facts learned about life in other cultures, ability to analyze critical incidents involving intercultural contact in one's own life, performance on measuring instruments designed to assess intercultural sensitivity, and rated performance in actual encounters or role-playing involving interactions between program participants and people from very different cultural backgrounds (see review of study by Collett, 1971, in Chapter 4).

# Control Groups Whose Members Receive other Training

Are there other designs that maintain a control group, allow for the assessment of long-term impacts, and fulfill ethical obligations to offer training to all who desire it? This is a large [p. 157 ↓] set of requirements to fulfill, but the answer is “yes.” However, there is a set of conditions that places limits on the number of research studies that allow a “yes” answer.

The research design to be discussed here calls for a training group that receives a newly developed program that the evaluator wants to assess and a control group whose members receive another type of training program. This other training program should be well established and have a reputation for effectiveness since it would be unethical to assign people randomly to a poorly conceived program. This concept of a comparison between a newly developed program and an established program is similar to comparisons in the development of new drugs. When pharmaceutical companies propose a new drug, it is not enough to demonstrate that it is effective against a disease. Federal regulations require that the new drug is demonstrably more beneficial (or as beneficial with fewer side effects) than other drugs already on the market. The diagram for this design is shown in [Table 6.2](#).

TABLE 6.2 Comparison Between Programs

Randomization	Observations: 1	Time 1	Observations: 2	Observations: 3
Group 1	○	X	○	long-term follow-up
Group 2	○	other training	○	long-term follow-up

People who want training are randomly assigned to one of two groups. Given that the assignment is random, the use of the term *control group* is appropriate. Measures are taken, and then all people receive training. “X” in the diagram refers to the newly developed program and “other training” refers to another program, perhaps one that the organization has been using for a number of years. Observations are taken shortly after training and a significant amount of time after training. In cross-cultural [p. 158 ↓]

studies, the long-term follow-up is sometimes carried out after people have assumed their overseas assignments or assignments within their own countries that demand extensive intercultural interactions. The evaluator looks for evidence, at Observations 2 and 3, that people in the newly developed training program demonstrate more intercultural effectiveness than people in the control group.

The use of this design becomes clearer if an actual study is reviewed. Cushner (1989) worked with 50 international exchange students in New Zealand who were sponsored by AFS International, one of the most active organizations dedicated to sponsoring intercultural experiences for adolescents. The 50 exchange students, all of whom were to attend high school in New Zealand, came from 14 different countries. Cushner was interested in evaluating a new training program based on the culture general assimilator (Brislin et al., 1986), a set of intercultural materials that was discussed in [Chapter 2](#). By a flip of the coin, 28 of the students were randomly assigned to receive training based on the culture general assimilator. Twenty-two of the students were assigned to another training program that had been long used by AFS International in New Zealand, and that consisted of “small group activities and discussions around such issues as self-identity, hopes and fears, local home and family life, and adjustment” (Cushner, 1989, p. 132). These 22 students used materials developed and distributed by AFS International, and consequently is called the “AFS materials” group in this discussion. Both training programs consisted of four sessions lasting 1 ½ hours each, and that took place over a 2-day period. This period corresponded to the 2 days between their arrival in New Zealand and the day they left the training site to join host families in various New Zealand towns and cities. All trainees completed the measuring instruments immediately after training, and they provided other information related to their intercultural adjustment 3 and 6 months into their homestay experiences.

**[p. 159 ↓ ]** Cushner found that students in the culture assimilator training group showed a number of benefits compared to those in the AFS materials group. Members of the culture assimilator group were able to solve more difficult critical incidents and were better able to analyze problematic intercultural incidents (in which they were participants) in their own lives. Three months after training, members of the culture assimilator group expressed feelings that they had more control over problems that might face them during their everyday lives in New Zealand. Feelings that one lacks control over everyday problems is a symptom of culture shock. Six months after



training, these same students were better able to suggest ways of solving problems (known as the *means* of problem solving; Spivak, Platt, & Shire, 1976) that faced them. In addition, another long-term effect (although not statistically significant) was that there were fewer major problems in the culture assimilator group during their stay in New Zealand. “Major problems” is a carefully chosen term, because it refers to problems so severe that adolescents (a) had to leave one host family and find another or (b) leave New Zealand prematurely and return to their own countries.

Why isn't this the perfect design? One reason is that program administrators and evaluators are almost working against themselves when they use this design. When the alternative training program is reasonable and when trainers have experience using it (both true of the AFS materials), then it is very hard to develop a new program that is more effective. In addition, when both programs are equally effective as shown by the evaluation data, there is no assurance that training of any kind is more effective than no training. Consider [Table 6.2](#) again: If both programs are equally effective, there is no control group that allows statements to be made about whether trained people have gained benefits that control group participants have not.

There is one other type of study where this design (sometimes called “control group also receives training”) is appropriate. In some organizations, the concept that people must receive cross-cultural training is well established. When this is true, **[p. 160 ↓ ]** then program administrators and evaluators can concentrate on developing training modules, or sets of materials, that are useful for very different purposes. One module might deal with gender differences, another with working well across cultural boundaries, another on understanding nonverbal behaviors, another on working through language interpreters, and so forth (Brislin & Yoshida, 1994). When these modules are used, evaluators can predict very specific benefits and then assess people to determine whether the benefits exist or not. For example, if a gender differences module is used, evaluators can predict that female co-workers will report, months after training, that people who worked with this module show respect for women and their contributions. Various modules can be compared with one another using the “control group also receives training” design. The purpose is not to test the general question of whether training is effective or not, but rather to test the more specific questions of whether specific modules lead to specific benefits. Such information is very helpful for the development of intercultural communication training programs because various



administrators can formulate program goals and then select from among a number of approaches, methods, and modules whose effectiveness has been established.

## Avoiding Biases in Evaluation Studies

In addition to providing persuasive information about the benefits of a training program, careful evaluation designs allow arguments to be made that the measured benefits are not due to other factors. The term *biases* is used frequently to refer to other factors, such as preexisting traits of the program participants or the pleasant personalities of the training personnel (in contrast to the training itself). More specifically (Campbell & Stanley, 1966), biases refer to factors that threaten the conclusion that a training program has led to various benefits. Six of the biases discussed most frequently among program administrators and evaluators (more are discussed in Campbell & Stanley, [p. 161 ↓ ] 1966; Cook et al., 1991; Rossi & Freeman, 1989) will be examined here.

### The Selection Bias

When the effects of training may be due to features of the program participants rather than to the program content, then the selection bias must be discussed. The possibility of a selection bias is especially threatening when the evaluation design does not have a control group. When the study by Befus (1988) was discussed earlier in this chapter, the possibility that people in her comparison group had higher levels of anxiety was mentioned. If certain people who are selected for either the training or comparison groups possess qualities that may affect the results of the evaluation study, then confidence in the results has to weaken. The best way to overcome a selection bias is to assign people randomly to the training or the control group. In this way, any preexisting qualities of people (such as anxiety level) are equalized prior to the start of training.

Another aspect of evaluation that invites discussions of a selection bias stems from the use of questionnaires (Cogswell & Stubblefield, 1988; Dixon, 1987; Trost, 1985). Many evaluation specialists fear that an overreliance on questionnaires leads to a bias given the reluctance of many trainees to give their honest reactions. So as not

to offend members of the training staff, many trainees may simply report that they enjoyed training even if they had reservations. This reaction is especially possible if the training staff consists of very pleasant, congenial people who are clearly trying to make a living offering programs to various companies. Rather than interfere with the trainers' livelihood, program participants make a point of saying nice things. To avoid this bias, evaluators make other types of measurements, such as ratings of respect shown when trainees interact with people from other cultures, observations of on-the-job performance, the ability to solve difficult critical incidents, the ability to suggest exact ways of solving problems, and so forth.

[p. 162 ↓ ]

## The History Bias

When experiences in people's lives can affect the results of training, then the history bias must be discussed. The term refers either to events in people's personal histories, such as previous intercultural experience, or events that occur at the same time as training. For example, consider a training program that is organized to prepare executives for sojourns in Germany. If the program was held during the week that the Berlin Wall was torn down, then evaluators would have to worry whether the results gathered after training were due to program content or to this major symbolic event. The history bias can be problematic in the wait-control design if there is a long period of time between assignment to the control group and the start of training (at Time 2, as previously depicted in [Table 6.1](#)). If the time period is long, events may happen that can be as impactful as training. Consider international students who have recently arrived in the United States. If they are asked to wait 3 days for their training, they may happen to meet some American students who are attending the same school. If those American students act in a friendly manner, tell the sojourners about key points such as how to register for the best courses, and take them on tours around the campus and nearby town, these events can easily be as impactful as a good training program.

## The Mortality Bias

Especially when an evaluation study takes place over a long period of time, participants may either drop out or become unavailable due to various changes in their lives.

*Mortality* refers to the loss of people in either the training group(s), control group, or comparison group. For example, Befus (1988) offered intercultural training to 64 North Americans who were studying in Costa Rica. Fifty-two people accepted her invitation and signed up for training, and in our experience this is a good response because the training was voluntary and the participants' free time was involved. However, 20 participants dropped [p. 163 ↓ ] out over the course of the 6-week program and consequently only 32 were available for final observations that were essential for a verdict concerning program effectiveness. Evaluators have to wonder about the representativeness of the final 32 people. Were they typical of the original 52 that signed up for the program? Perhaps the final 32 were the people most interested in intercultural interaction and most concerned about understanding culture and cultural differences. This initial level of interest could influence observations made at the conclusion of training.

## Biases Due to Reactive Effects

At times, people who find themselves selected for training may view themselves as privileged and as singled out for special attention. Further, especially if the training appears well-thought-out and potentially beneficial, participants may develop favorable views of their organization for choosing them. These possibilities are reactions to *participation in training*, not to the content of the training. These reactions are sometimes called “Hawthorne effects” because they were documented in a classic study that took place at the Hawthorne plant of the Western Electric Company in Cicero, IL (see Landy, 1989, for a discussion). One of the surprising findings of that research was that many kinds of changes in the organization led to greater worker productivity. One of the few common elements underlying the changes was that workers felt special given the attention they received from management and prestigious scholars.

When evaluators talk about “controlling for Hawthorne effects,” they are referring to biases brought on by the special attention that accompanies training. Cushner (1989) thought carefully about this possibility, and his analysis was one reason for the choice of a design in which all participants received training (as reviewed previously). He reported, “There is no reason to believe that any [participant] would perceive oneself as receiving anything special during the weekend as all subjects [p. 164 ↓ ] received the same number of hours of good training by concerned and experienced representatives from local AFS chapters” (p. 132).

## Compensatory Treatment of the Control Group

The ethical qualms that administrators experience when designating a training group and a no-training control group have been referred to several times. Administrators do not like to deny training to people who desire it. Assume that the training program is to last 4 weeks, and that participants are to be taken off their jobs for 2 hours a day to attend training sessions. If they add the “compensatory treatment of the control group bias,” the administrators may unconsciously treat control group members in special ways to make up for the decision that delegated these people to the no-training group. They may approve requests for leave, overtime with double pay, and attendance at lectures in the community that bear a relationship to the organization's international goals. The treatment the control group members receive may lead to benefits that are similar to those demonstrated by members of the training group. Even if the training was beneficial, it would be difficult to document this fact given the compensatory treatment received by control group members.

## Compensatory Activities among Control Group Members

At times, control group members learn that there is a training group and become resentful. As a result, they organize and engage in activities that develop group

cohesiveness and morale. They may talk among themselves and say, “I don't know why those other people got to be in the training group and we didn't. But we're just as good as those other people. Let's do important things around here and show management that we're right!”

The alternative name for this bias is the “John Henry” effect, and this is our favorite name for a term used by evaluators given [p. 165 ↓ ] our interest in music. John Henry was a steel driver who could hold a heavy hammer in each hand and drive steel into rocks to cut tunnels for railroads. One day, an inventor with a steam drill came to the work sight and argued that he could cut a longer tunnel than any worker. So John Henry competed against the steam drill, perhaps resentful of the positive attention that management lavished on the inventor. John Henry won the contest because he drove a 14-foot tunnel while the steam drill had to stop at 9 feet. The tragedy of the story is that John Henry broke a blood vessel during the contest and died. In one of American folk music's best known tributes to feminism, his wife Polly took over his railroad job to earn money for the support of their children.

Whenever they discuss the best ways to document the benefits of training programs, administrators and evaluators should examine the various biases that can threaten the conclusion of “positive effects.” This list of six discussed here is a good start, and others can be found in the scholarly literature on program evaluation (e.g., Campbell & Stanley, 1966; Cook et al., 1991; Rossi & Freeman, 1989). The purpose of this exercise is to rule out explanations that could be due to biases, and to build a strong set of arguments that the program leads to genuine benefits. Researchers interested in intercultural training have involved themselves as evaluators of many different types of programs, and they have documented a wide variety of benefits. Knowledge of these benefits is very useful to program administrators and evaluators.

## The Benefits of Intercultural Training

Over the past 25 years, a large number of researchers have become involved in the careful evaluation of various intercultural training programs. The many benefits they have documented can provide very helpful guidance to administrators contemplating the development of new programs. Administrators can realistically adopt the goals

of training represented in [p. 166 ↓ ] this list because they have been achieved by others. Evaluation specialists can benefit because the studies listed give information on important topics such as study design, measuring instruments, and the development of theory. In all these studies, people receiving training were compared with either a control group or a comparison group. When there were potential biases, as in the study by Befus (1988; as discussed above, especially the mortality bias), researchers presented a reasonable set of arguments that the potential threats to validity did not overwhelm the conclusion of positive benefits. The benefits can be organized in a manner similar to the coverage of training content in this book: effects on people's thinking, on their attitudes and emotions, and on their behavior.

## Positive Effects Involving Peoples Thinking and Knowledge Development

The benefits due to knowledge development are probably the easiest to measure because people can demonstrate changes in thinking and knowledge acquisition on paper-and-pencil tests and questionnaires.

## Positive Effects Involving Peoples Affective Reactions

Affective reactions include people's attitudes, self-concepts, emotions, and feelings of comfort in another culture rather than stressful reactions. Although most studies to date are based on people's self-reports, we predict that future research will see the addition of more objective indicators such as measures of blood pressure (Charlesworth, Williams, & Baer, 1984) reports by others who know program participants well, careful records of visits to physicians, and so forth.

## Positive Effects Involving Peoples Behavior

Program administrators and evaluators are correct when they argue that changes in people's knowledge and feelings are extremely important. However, evidence about changes in people's visible behaviors will always be more persuasive to outsiders who make decisions concerning the amount of support that training will receive.

[p. 171 ↓ ] As discussed in [Chapter 4](#), our strong recommendation is that training give more attention to identifiable behaviors. In addition to excellent program content, a focus on behaviors makes evaluation studies more impactful when evidence can be presented that people have changed their behavior (e.g., visible job performance) in desirable directions.

## Evaluation: Its Role in Continuing the Support of Intercultural Training

In addition to providing important information central to the improvement of programs, evaluators can make other contributions. One, as seen in the critical incident that introduced this chapter, is that they assist program administrators defend the continued funding of intercultural training programs. In almost all the organizations that we know, the budget for training programs is not large and there is a great deal of competition for available funds. If program administrators cannot convince executives that intercultural training is effective, then the funds will go to other types of training. We recommended attention to measures of actual behavior, not only because they provide important information that can be integrated into books like this one, but also because behavioral measures provide the most convincing evidence of program effectiveness. When a concept such as “convincing funders” is treated, we are moving away from discussions of intercultural training itself and are beginning to examine the larger social context in which training exists. Intercultural training does not exist in a vacuum: It takes place in societies where many administrators are competing for program funds, where young professionals want to make a living by offering training, and where many potential clients do not know how to deal with the issues stemming from the increased

intercultural interactions they are experiencing. In the final chapter, we discuss some of these issues that will affect the future of intercultural communication training.

<http://dx.doi.org/10.4135/9781452243603.n6>