Leadership training has led to a large amount of research due to the belief that such training can lead to (or more precisely cause) positive changes in followers’ behavior and work performance. This chapter describes some of the conditions necessary for research to show a causal relationship between leadership training and outcomes. It then describes different research designs, employed in leadership training research, and considers the types of problems that can affect inferences about causality. The chapter focuses on the role of randomization of leaders (e.g., into training vs. non-training conditions) as a key methodological procedure and alludes to problems of achieving this in field settings.

Keywords: Leadership training; criteria for causality; threats to validity; research designs; leader randomization

Introduction

How to develop and train managers to be effective leaders has led to a large amount of research (Day, 2001; Day & Dragoni, 2015; Day, Fleenor, Atwater, Sturm, & McKee, 2014; Feldman & Lankau, 2005; Riggio, 2008). Leadership training and development initiatives are highly popular in organizations due to the belief that training managers can lead to (or more precisely cause) positive changes in followers’ behavior and work performance. Research shows many positive benefits of leadership training on a wide range of outcomes, but these vary considerably between different types of intervention, theoretical foci, and outcomes (Aguinis & Kraiger, 2009; Avolio, Reichard, Hannah, Walumbwa, & Chan, 2009; Collins & Holton, 2004). The extent to which these studies show a causal relation between training and outcomes is much disputed due to the wide range of problems associated with various research designs. However, given that leadership training is one of the most common forms of managerial training, it seems imperative that organizations are able to properly assess whether such training is likely to lead to positive outcomes. The aim of this chapter is to focus on the research designs (with associated methodological problems) employed in evaluating leadership training and consider the types of problems inherent in them that can affect inferences about causality between training and outcomes. We also incorporate insights from implicit social cognition to demonstrate how leadership training interventions can indeed help “harness managers’ reflexive and reflective abilities” (Hodgkinson & Healey, 2011, p. 1510). Whereas most current leadership programs focus on the lower right-hand quadrant of Hodgkinson and Healey’s (2011) managerial cognition circumplex (conscious/deliberative – cold cognition), we will demonstrate examples of interventions that address
The chapter is organized into six sections. First, we summarize research showing the benefits of leadership training on a range of work outcomes. Second, we examine the conditions necessary in order to infer causality in research and some practical problems of achieving these conditions in studies examining leadership training. Third, we examine a number of research designs that are best suited to deal with a range of threats to determine causality and give examples from leadership research. Fourth, we focus on the role of randomization of participants (e.g., into intervention vs. control conditions) as a key methodological research procedure and allude to problems of achieving this in field settings. Fifth, we address the implications of implicit social cognition research on leadership training and offer examples of related interventions. Finally, we summarize the main aspects of the chapter and provide some guidance for future research.

Evaluation of Leadership Training Research

In terms of nomenclature for this chapter, we define leadership interventions as “...those studies where the researcher overtly manipulated leadership as an independent variable through training, assignment, scenario or other means” (Avolio et al., 2009, p. 764). These can consist of both laboratory and field studies conducted in organizations. In this review we are explicitly focusing on leadership training where people are directly taught specific leadership skills and not on other leadership development techniques (e.g., mentoring, coaching, job assignments, see Day, 2001, for a review). Leadership training is the most direct way to develop manager’s leadership competencies, can cover many topics related to leadership, and typically occurs in small group classroom-type settings. Through various techniques (e.g., multi-source feedback, reflection, role-playing) managers develop various leadership skills and competencies. The aim of the training is to change the way the leader behaves toward their followers, which can lead to improved follower well-being (such as enhanced job satisfaction and organizational commitment, or lower stress), team well-being (such as cohesiveness), and work performance. Often the training program is referred to as the “intervention” and those managers who do not receive training are in the “control” or “non-intervention” condition. A range of different types of data can be collected to evaluate the training, and this can occur before (pretest) and/or after (posttest) the training. In some cases multiple data collection points are obtained to enable closer evaluation of the impact of the training. The types of data can vary considerably and can be obtained from many sources (e.g., from leaders, followers, HR, clients, etc.).

There have been a number of narrative reviews and meta-analytic reviews examining leadership training, as cited above; however, the one by Collins and Holton (2004) is particularly relevant for this chapter as it distinguished between different research designs. They identified 346 studies that involved a leadership intervention of which only 132 were empirical and 103 met various selection criteria for inclusion (many studies yielded multiple effect sizes). One of the main selection criteria was that it employed one of three research designs that are of interest: posttest only with control group (36 studies), pretest–posttest with control group (26 studies), and single group pretest–posttest (25 studies). Of the studies, 83 studies were formal training sessions and the remainder predominantly feedback and coaching sessions. The content of the training varied considerably and included: problem solving and decision-making, strategic stewardship, employee performance, human relations, job and work redesign, and general management. We will comment on the content of training later. The main unit of analysis was effect sizes (gain score for posttest–pretest designs) or the difference between training and control conditions (posttest only design). They found “...the effectiveness of managerial leadership development programs varied widely; some programs were tremendously effective, and others failed miserably” (p. 232). Overall, the effect sizes for all research designs, across different outcomes, were positive showing either improvements over time or compared to a control condition. While not a focus of the meta-analysis, there seemed to be more variability, and negative effect sizes, in the posttest only with control condition. This might be expected as the comparison is not within-person (i.e., pretest vs. posttest) but across individuals (training vs. control conditions) with more opportunity for confounding variables to have an effect.

One observation of the Collins and Holton (2004) meta-analysis was the wide range of topics that were included in the training programs. Indeed, they appear more akin to management training in general rather than to the leadership process. More so, they do not clearly appear to be related to well-known leadership theories. This, we suspect, is likely to be due to the need for trainers to meet client needs, and therefore tailor training content, rather than testing specific leadership theories. However, the large variation in content in training interventions further makes it difficult to evaluate between different training programs.

The above issue concerning content of training programs was examined by a meta-analysis by Avolio et al. (2009) that focused both on laboratory and field studies (and hence a different sample to the Collins & Holton, 2004 meta-analysis). More specifically, they categorized studies into two groups based on the theoretical models the training was based upon. Traditional theories (up to 1970s, 41 effect sizes) included behavioral and contingency models of leadership and a focus on non-conscious/automatic and hot cognition processes.

The chapter is organized into six sections. First, we summarize research showing the benefits of leadership training on a range of work outcomes. Second, we examine the conditions necessary in order to infer causality in research and some practical problems of achieving these conditions in studies examining leadership training. Third, we examine a number of research designs that are best suited to deal with a range of threats to determine causality and give examples from leadership research. Fourth, we focus on the role of randomization of participants (e.g., into intervention vs. control conditions) as a key methodological research procedure and allude to problems of achieving this in field settings. Fifth, we address the implications of implicit social cognition research on leadership training and offer examples of related interventions. Finally, we summarize the main aspects of the chapter and provide some guidance for future research.

Evaluation of Leadership Training Research

In terms of nomenclature for this chapter, we define leadership interventions as “…those studies where the researcher overtly manipulated leadership as an independent variable through training, assignment, scenario or other means” (Avolio et al., 2009, p. 764). These can consist of both laboratory and field studies conducted in organizations. In this review we are explicitly focusing on leadership training where people are directly taught specific leadership skills and not on other leadership development techniques (e.g., mentoring, coaching, job assignments, see Day, 2001, for a review). Leadership training is the most direct way to develop manager’s leadership competencies, can cover many topics related to leadership, and typically occurs in small group classroom-type settings. Through various techniques (e.g., multi-source feedback, reflection, role-playing) managers develop various leadership skills and competencies. The aim of the training is to change the way the leader behaves toward their followers, which can lead to improved follower well-being (such as enhanced job satisfaction and organizational commitment, or lower stress), team well-being (such as cohesiveness), and work performance. Often the training program is referred to as the “intervention” and those managers who do not receive training are in the “control” or “non-intervention” condition. A range of different types of data can be collected to evaluate the training, and this can occur before (pretest) and/or after (posttest) the training. In some cases multiple data collection points are obtained to enable closer evaluation of the impact of the training. The types of data can vary considerably and can be obtained from many sources (e.g., from leaders, followers, HR, clients, etc.).

There have been a number of narrative reviews and meta-analytic reviews examining leadership training, as cited above; however, the one by Collins and Holton (2004) is particularly relevant for this chapter as it distinguished between different research designs. They identified 346 studies that involved a leadership intervention of which only 132 were empirical and 103 met various selection criteria for inclusion (many studies yielded multiple effect sizes). One of the main selection criteria was that it employed one of three research designs that are of interest: posttest only with control group (36 studies), pretest–posttest with control group (26 studies), and single group pretest–posttest (25 studies). Of the studies, 83 studies were formal training sessions and the remainder predominantly feedback and coaching sessions. The content of the training varied considerably and included: problem solving and decision-making, strategic stewardship, employee performance, human relations, job and work redesign, and general management. We will comment on the content of training later. The main unit of analysis was effect sizes (gain score for posttest–pretest designs) or the difference between training and control conditions (posttest only design). They found “...the effectiveness of managerial leadership development programs varied widely; some programs were tremendously effective, and others failed miserably” (p. 232). Overall, the effect sizes for all research designs, across different outcomes, were positive showing either improvements over time or compared to a control condition. While not a focus of the meta-analysis, there seemed to be more variability, and negative effect sizes, in the posttest only with control condition. This might be expected as the comparison is not within-person (i.e., pretest vs. posttest) but across individuals (training vs. control conditions) with more opportunity for confounding variables to have an effect.

One observation of the Collins and Holton (2004) meta-analysis was the wide range of topics that were included in the training programs. Indeed, they appear more akin to management training in general rather than to the leadership process. More so, they do not clearly appear to be related to well-known leadership theories. This, we suspect, is likely to be due to the need for trainers to meet client needs, and therefore tailor training content, rather than testing specific leadership theories. However, the large variation in content in training interventions further makes it difficult to evaluate between different training programs.

The above issue concerning content of training programs was examined by a meta-analysis by Avolio et al. (2009) that focused both on laboratory and field studies (and hence a different sample to the Collins & Holton, 2004 meta-analysis). More specifically, they categorized studies into two groups based on the theoretical models the training was based upon. Traditional theories (up to 1970s, 41 effect sizes) included behavioral and contingency models of leadership and a focus on
identifying what leader behaviors are effective in different situations. Newer theories (1980s to present, 40 effect sizes) that included charismatic, inspirational, transformational, and visionary leadership, focus more on changing leaders’ beliefs and cognitions about the leadership process. The main outcomes were affective, behavioral, cognitive, and organizational performance criteria. Overall, leadership training had a positive impact on a range of outcomes in that those in the training condition had a 66% chance of improvement while those in control conditions had only a 34% chance of improvement. While there was no overall difference in effect sizes between the traditional and newer theories, there were some interesting differences when taking outcomes into consideration. The traditional theories had a greater impact on behavioral outcomes while the newer theories had a bigger impact on affective and cognitive outcomes. This might not be surprising as traditional theories tend to specify the behaviors leaders should engage, while newer theories (such as transformational leadership) are designed to shape the way followers think and feel about their work.

While there is research showing the benefits of leadership training, the benefits vary considerably between different types of interventions, the content of the intervention, and the outcomes assessed. However, while there seems to be an association between leader training and outcomes, demonstrating that training causes positive change in followers is not considered in these reviews. In the next section, we examine the concept of causation and how this applies to understanding the link between leadership training and outcomes.

The Causal Link Between Leadership Training and Outcomes

The aim of research into leadership training is to show how the application of research in this field can benefit the leadership process and has positive benefits on a number of important work-related outcomes. A major methodological challenge for research into leadership training, as indeed in any research, concerns demonstrating causality (i.e., that training causes or leads to benefits in a range of outcomes). Determining causality requires three conditions to exist between a presumed cause x (i.e., leadership training) and an effect y (i.e., posttest/intervention outcomes such as, job satisfaction, performance): (i) x must precede y temporally (i.e., the intervention needs to occur before posttest measures are taken), (ii) x must be reliably correlated with y (beyond chance; i.e., the intervention needs to reliably predict posttest measures), and (iii) the relation between x (intervention) and y (outcomes) must not be explained by other causes (see Antonakis, Bendahan, Jacquart, & Lalive, 2010, for a discussion of these in relation to leadership research).

The “gold” standard to examine causality, where all three above conditions are met, is randomized experiments where leaders are randomly allocated to an intervention vs. non-intervention (control) conditions. When these conditions occur then changes in the dependent variable (e.g., outcomes) can be inferred to be caused by the independent variable (e.g., training condition) and not due to confounding variables. While such studies can be conducted in laboratory situations (where the researcher has control over the situation), they are extremely difficult to conduct in the field, and therefore it is useful to reflect on the types of problems that each of the three conditions for causality pose for evaluating leadership training.

The first condition that the training precedes the outcomes is relatively straightforward in research designs. However, the time period between training and posttest evaluation of outcomes is a much more difficult issue to determine. This should reflect the theoretical content of the training and correspond to how long it should take for (a) training to be manifested in the predicted changes to leaders’ behaviors, (b) leaders’ behaviors to have a reliable effect on team members’ attitudes and behaviors, and (c) changes in team members’ behaviors to affect performance. These decisions will be based on the theoretical content of the training and the context of the trainees’ jobs. However, in our experience, very little consideration is given to understanding these steps, and often the time period between training and measuring outcomes reflects intuitive ideas at best and client needs at worst. Consequently the time periods vary considerably between studies and are not factored into reviews of the area.

The second condition that the independent variable affects the dependent variables (i.e., that the training should reliably predict the outcomes) is determined by statistical procedures (e.g., effect sizes). "Affects," in this research context, implies some statistical coefficient of change and, of course, there are a number of statistical issues that can affect the interpretation of such results (such as assumptions over normality of data, heteroscedasticity, etc.). In addition, training is not a continuous variable, but binary (a leader either does or does not get trained, although one might examine the amount of training). Therefore, the effect of training needs to be assessed either over time (pretest vs. posttest) and/or compared to a non-training condition. Such research designs can build in a number of threats to the validity of the findings which we discuss in the next section.

The third condition that the effect of training on outcomes is not explained by other causes is the most difficult one to determine in research generally and specifically in relation to leadership training. In order to achieve this, studies need to include appropriate control (non-training) conditions or be quasi-experimental in nature with multiple measurement times. Referring back to the experimental design as one way to determine causality leads to research designs that involve training and non-training conditions where leaders are randomly allocated into different conditions (to control for confounding variables, such as selection bias) and with different combinations of pretest and posttest measurement conditions (in order to account for different threats to establish the cause of the effects). In the next section, we describe a
Evaluating Intervention Research Designs

We draw upon the classic analysis of how to deal with internal and external threats to validity (or confounding explanations) in different research designs by Campbell and Stanley (1963). This seminal publication eloquently explains eight internal and four external threats to test the effects of a treatment variable. We focus mainly on the internal threats and while it is not our aim to describe these in detail, we explain how some of these are important in different research designs that can be employed in leadership training research. Table 1 describes internal threats to validity and gives examples of how they might apply to leadership training.

In this section we focus on a number of research designs that have been employed to evaluate leadership training and describe some of the strengths and weaknesses of each. Note in this section we refer to the training intervention as x.

One Shot Case Study (x–Posttest) is a situation where only those undergoing the training are examined and measures are only taken posttest (e.g., Gilpin-Jackson & Bushe, 2007). In many cases, such a design might be a consequence of restrictions placed upon the researcher by the client organization in terms of data collection or a lack of time to collect pretest data. However, this design is the one that is most open to many threats to internal and external validity. The most notable ones are the inability to determine whether posttest scores are due to training or background effects (history, e.g., performance may have been increasing over time independent of the training), changes in the trainees over time (maturation, e.g., leaders and/or team members’ competencies improve over time with increased experience), particular skills of those chosen to take part (selection, e.g., “better” managers are chosen to receive the training), or loss of participants (mortality, e.g., poorer leaders or team members leaving the company during the study period). A derivation of this design is when posttest data is also collected from a non-trained group, and this design can have some research merits (Collins & Holton, 2004), especially if allocation to conditions is random (e.g., Urban, Ferris, Crowe, & Miller, 1985).

One-group Pretest–Posttest Design (Pretest–Posttest) is similar to the previous design but also includes a pretest measure (e.g., Arthur & Hardy, 2014; Biggs, Brough, & Barbour, 2014). This is an improvement on the one shot case study design as it is able to control for potential threats such as mortality (one could examine differences in pretest for those that complete posttest and those that do not and if no difference appears this indicates that mortality is not a threat). In addition, this design allows gain scores to be calculated (post- minus pre-test) to gauge the size of the effect of the training. However, the design does introduce potentially new threats to the validity to causality such as testing (the effect of measures being affected by previous measurements). In the context of leadership training, it might be that the completion of pretest measures may impact upon posttest measures. This can occur in many ways including the pretest measure unintentionally alerting team members to aspects of the leader’s behavior that they had not hitherto considered, and this affects their subsequent reactions to them.

Pretest–Posttest Control Group Design (Pretest1–Posttest2; Pretest2–Posttest3) is a development of the previous design and also includes a condition where leaders do not receive the training and, in many respects, represents a typical experimental design (e.g., Barling, Weber, & Kelloway, 1996; Dvir, Eden, Avolio, & Shamir, 2002; Graen, Novak, & Sommerskamp, 1982; Yeow & Martin, 2013). In ideal situations, the allocation of leaders into the two conditions is random and pretest and posttest measures are the same in each condition and collected at similar time points. There are many practical problems of randomizing leaders between conditions in field settings and we discuss these below. This is a very powerful design and it can account for all eight internal threats to causality (see Table 1). If tests show that the internal threats to causality have not occurred, then the key comparison is the change over time in the training condition compared to that in the control condition. If the change is more positive in the training condition, compared to the control condition, then it can be deduced that training has caused the change. The rate of change ironically might be negative in both conditions (i.e., lower at posttest) but if the deterioration is less in the training condition than the control, then the “positive” benefits of training can be established. One important potential external threat does exist and that is the interaction between testing and the independent variable. In the context of leadership training it would occur where the pretest measure differentially alerts respondents in the training condition to key causal variables, compared to those in the control condition. For example, consider a situation where the aim of the intervention is to develop leaders’ relationship skills and the pretest assessed team members’ judgment of leaders’ relationship building ability. While team members in both conditions complete the same pretest measures, and therefore each might be more aware of their leader’s relationship skills, this might be most salient for those in the training condition when they subsequently observe their leader exhibiting better relationship skills following the training. The next design deals with this particular threat to validity.

Solomon Four Group Design (Solomon, 1949: Pretest1–Posttest1; Pretest2–Posttest2; x–Posttest3; Posttest4) is a design that repeats the previous one but also includes two conditions without the pretest measures (e.g., for management training, Holdnak, Clemons, & Bushardt, 1990). As with the previous design, it is important that leaders are randomly allo-
cated into each of the four conditions and that measurement periods coincide. If the differential pattern in posttest outcomes between the training and control conditions is the same irrespective of whether pretest measures are taken, then one can conclude that the threat associated with the interaction between testing and independent variable has not occurred. While this might be considered one of the best research designs to examine the impact of leadership training, it occurs very infrequently due to practical difficulties of collecting such data.

The final design we consider is quasi-experimental designs (sometimes referred to as time-series designs, e.g., Pretest; Pretest-x-Posttest; Posttest) which involves multiple testing before and after the intervention (e.g., DeRue, Nahrgang, Hollenbeck, & Workman, 2012). In other words, outcomes are obtained on multiple occasions before and after the training. The time period between data collection is crucial, especially around the time of the training intervention and should follow theoretical considerations related to the content of the training (see discussion above on establishing the first criteria for causality). There are many variants of this type of design but the aim is to show that the pattern of outcomes significantly changes as a result of the training. Various patterns can occur but the “ideal” one is to show (a) a consistent level in outcomes on all pretest time intervals, (b) an increase in outcomes immediately following the training (or very soon afterwards), and (c) a consistent increase on all posttest time intervals. The data rarely fits such an ideal profile (see, e.g., Graen et al., 1982) and so trends in the data are examined. Many other patterns across the time intervals can indicate benefits of the training. Since this approach does not necessarily need a control condition (although the inclusion of one would be highly desirable) it is ideally suited to evaluate leadership training research.

Problems of Participant Selection and Random Assignment in Leadership Training

In the previous section we described a number of research designs that can be used to evaluate leadership training. We made the observation that research designs that are better able to deal with a range of threats to validity are ones with randomization of leaders into an intervention and non-intervention (control) conditions. Random assignment helps to ensure that any differences in leaders between conditions are not systematic at the outset of the study and therefore cannot explain differences in outcomes. This assumption is based on the premise that sufficiently large samples are obtained to be confident that there has been random allocation of all potential confounding variables between conditions. In our experience, this assumption tends to be assumed rather than explicitly examined. If true randomization has been achieved, then one can assume that differences in outcomes can be more confidently attributed to the treatment (i.e., training). Of course, randomization does not guarantee that the conditions are equivalent as there may still be differences on some pre-existing attributes due to chance, but it greatly reduces it.

Ideally, allocation into an intervention and control condition should be random (e.g., determined by random tables or flip of a coin) such that there is an equal chance that each manager is placed into either condition. However, random allocation of managers is extremely difficult to do practically in a field setting (for an exception see Bellé, 2014). There are many reasons for this and we describe some of these below.

A pragmatic problem in obtaining randomization is that the client, and not the researcher, typically has overall control of who is selected for training. Training comes with considerable costs for the organization (both financially and time-wise) and this can significantly affect the selection process. Perhaps counter-intuitively, in our experience, organizations often select their “better” managers to be trained in the belief that they would benefit the most. This might seem somewhat ironic given that one might argue that poorer managers have more to gain, as they have a greater scope for improvement. However, such a selection strategy increases the chances of there being many differences between those selected and those not selected for training that can act as confounding variables (the selection bias).

Related to the last issue are potential ethical issues concerning denying some managers (and their team members) training and development that might aid their wellbeing and performance. We observed this issue in a study conducted by two of the authors that employed student project teams on a university module that required them to work as a virtual organization, with one member of the team acting as the leader. In a previous study we proposed that we would randomly select team leaders to undergo a self-regulation training intervention, and we were surprised when our initial application to the ethics committee was rejected. The logic, the committee argued, is that if you have a technique that you know helps some leaders and their teams, it would be unfair to deny this to all teams. We were able to overcome this issue, and hence run the study, by arguing that this was the first time we have delivered this intervention, and while we knew it would not harm participants’ performance, we did not know if it would be successful. In addition, we offered to share the training materials with leaders not trained after the study. The intervention proved to be highly successful (Yeow & Martin, 2013) which meant that the next time we conducted a study with a subsequent cohort we were unable to achieve leader training randomization but had to offer the training to all leaders (Martin & O’Brien, under review). In general, issues over the selection of leaders to training and non-training conditions and the context in which randomization is achieved are rarely, in our experience, considered.

Even if random assignment does occur, there are a number of additional problems that might arise. One such problem
concerns the independence of managers from others and the potential of carry-over effects between different managers. Managers do not work in a vacuum but in a complex social system involving other managers, team members, and support personnel. If managers are drawn from the same organizational site it is highly likely that trained managers (and their team members) interact with, and therefore potentially have an impact upon, control managers (and their team members). One way to potentially avoid this is to select managers from each condition from different regional sites (see, e.g., the wonderful study by Paul, Robertson, & Herzberg, 1969). This might achieve independence between the managers in the different conditions but equally raise issues of regional selection biases.

Another problem with randomization of managers is that this does not mean randomization of team members. Training interventions are designed to help leaders change the behaviors of their team members to meet organizational goals. While it is the manager who is trained, the target of the training is typically the team members. The motivation, skills, and abilities of the team members will clearly have an impact on the success of the training. Ideally, one would wish to randomly allocate team members to different managers—a process that would be impossible in a field experiment (though it would be standard in a laboratory experiment). How much of a problem this is for studies depends on the role of the manager in selecting and recruiting team members. If the manager plays a significant role in recruiting and maintaining team members, as we suspect is common, then selection biases at the leader level would also apply to the team member’s level. We believe this is an unexplored but extremely important issue and highlights the fact that leader training is directed not just at the leader but at their team members.

Implicit Social Cognition and Implications for Leadership Training

As mentioned in the Introduction, a shortcoming of common leadership interventions is that they do not directly address the role of automatic/implicit processes in the exercise of leadership. Training programs tap mainly upon the reflective (conscious/deliberative) system, whereas the reflexive (non-conscious/automatic) system gets downplayed (Hodgkinson & Healey, 2011). Implicit social cognition theories (e.g., Galambos, Abelson, & Black, 1986; Lord & Maher, 1991; Payne & Gawronski, 2010) suggest that people rely on cognitive simplification mechanisms, mental representations, or schemas to cope with information complexity (e.g., Fiske & Taylor, 1991; Hodgkinson & Healey, 2008). In the context of leadership, past research has shown that people are categorized as leaders (or followers), based on an implicit matching process between the behavior or character of the target person and the attributes of a pre-existing leader (Implicit Leadership Theories, ILTs), or follower category (Implicit Followership Theories, IFTs) held in memory (Epitropaki & Martin, 2005; Epitropaki, Sy, Martin, Tram-Quon, & Topakas, 2013; Lord & Maher, 1991; Sy, 2010). Although ILTs and IFTs tend to operate in implicit (automatic) fashion, they may also be processed explicitly (consciously; Epitropaki et al., 2013; Lord & Maher, 1991). Whereas early theories of implicit social cognition imposed a strict dichotomy of explicit vs. implicit processing, the current consensus is that most schemas may be processed explicitly and implicitly, with implicit processing as the default mode and explicit processing operating only in situations when sufficient motivation and opportunities (e.g., time and cognitive capacity permitting) are present (Bargh, 2006; Gawronski & Payne, 2010; Smith & DeCoster, 2000; Strack & Deutsch, 2004).

A question that naturally arises in the context of leadership training is whether implicit theories of leadership can change and thus be learnt. Recent implicit social cognition research challenges the dominant assumption that schemas are resistant to change and offers support for malleability (e.g., Payne & Gawronski, 2010). Connectionist perspectives (Hanges, Lord, & Dickson, 2000; Lord, Brown, & Harvey, 2001) in particular highlight the dynamic nature of leadership schemas. The core argument is that leadership categories are sensitive to context and they vary both within and between individuals and are thus generated in real time as a response to contextual, task-related, and person- and organization-influenced factors. Despite their dynamic nature, connectionist models also allow for ILTs and IFTs generalizability and stability over time (Epitropaki & Martin, 2004). As Lord et al. (2001) state “this model can be used to understand both the stability and flexibility that is witnessed in the application of leadership prototypes” (p. 311). A generic leadership schema that is generalizable in different contexts can exist but there is also the possibility of different node activation of the generic schema in different contexts.

In the context of leadership training, leaders and followers need to develop awareness about the expectations for their leadership context and how implicit theories shape decision-making and action tendencies. Given that many of the daily behaviors operate without full conscious awareness (Wood, Quinn, & Kashy, 2002), individuals may not be fully aware of how their implicit theories shape action tendencies without such training. An example of a training intervention on ILTs is the drawing exercise developed by Schyns, Kiefer, Kerschreiter, and Tymon (2011) that can be used to help managers understand their implicit assumptions about leadership and how these are often in conflict with their espoused theory of action (Argyris & Schön, 1978). For example, managers may explicitly state that all necessary conditions for women’s advancement in leadership positions exist in their organization, but they predominantly draw male leaders (implicit schemas). Thus, raising awareness about implicit theories and addressing discrepancies between implicit and explicit
processes are important.

Two other interventions show promise for changing implicit processes: Conditioning and Selective Prototype Activation (Dasgupta & Asgari, 2004; Epitropaki et al., 2013; Hofmann, De Houwer, Perugini, Baeyens, & Crombez, 2010; Gawronski & Bodenhausen, 2006; Gawronski & Sritharan, 2010). Conditioning involves pairing a concept (e.g., leader) with a new association (e.g., woman). Prior studies have provided support for the possibility of changing implicit assumptions. For example, spending a few minutes imagining a “strong woman” reduced implicit gender stereotyping (Blair, Ma, & Lenton, 2001) and viewing photos of admired African-American leaders along with loathed white Americans such as serial killers reduced implicit race bias (Dasgupta & Greenwald, 2001). Such conditioning may require several hundred trials (Olson & Fazio, 2001) and it may not have been feasible in the past, but recent technological advancements and the availability of software training programs that have shown positive results for changing implicit associations (e.g. IAT; Kemps, Tiggemann, Martin, & Elliott, 2013) make it possible.

Selective Prototype Activation is another possible intervention in the context of leadership training that further blends cold and hot cognitions. It assumes that individuals hold multiple (positive, negative, and neutral) implicit theories, and any given prototype may be activated in a given instance (Hanges et al., 2000; Lord & Shondrick, 2011; Sy, 2010; Sy et al., 2010). Prior research has argued that affect and ILTs and IFTs possess structural similarities in that both are positively and negatively valenced. Kruse and Sy (2011) across four experiments incorporating multiple affect (sadness, anger, and happiness) and samples (students, working adults, and leaders), demonstrated that affect activated corresponding IFTs such that sadness and anger activated negative IFTs (but not positive IFTs), and happiness activated positive IFTs (but not negative IFTs). Emerging research in ILTs shows similar results. For example, Johnson, Walczak, and Sy (2013) using neural networking found stress to lead to an increased activation of negative prototypes such as Tyranny and Masculinity, and deactivation of the positive prototype of Sensitivity.

Accordingly, a Selective Prototype Activation intervention would involve repeated cueing of positive prototypes such that they become chronically accessible and can be implemented via games and online training methods (Kruse & Sy, 2011). Such an intervention may influence managerial action tendencies, as individuals tend to use schemas that are most readily accessible in responding to others (Chen & Bargh, 1997; Devine, 1989; Srull & Wyer, 1979). It is also possible that over time, it may shift individuals’ implicit theories toward a permanent positive change.

Summary

Leadership training is one of the most popular aspects of managerial training due to the belief that such training adds benefits to a range of important work-related outcomes. Overall, while research shows a strong association between leadership training and positive work outcomes, this varies considerably between different types of interventions, outcomes assessed, and research methods employed. While there seems to be an association between training and outcomes, it is unknown whether the former causes the later. We argue that it is important for research to show a causal relation between training and outcomes both theoretically, to support various aspects of leadership theory; and practically, to justify organizations’ costs in engaging in leadership training. The aim of this chapter has been to consider the issue of causality in leadership training research and how different research designs can mitigate against potential threats to the validity of research findings.

To examine these issues, we first describe three main criteria that studies need to establish to infer causality between an independent (in this case, training vs. non-training) and dependent (in this case, follower, team, leader, organizational outcomes) variables and some of the major internal and external threats to the validity to findings. Conditions for infer causality are typically met in laboratory experimental situations where the researcher has complete control over the situation. However, such control is rare in applied settings examining leadership training, resulting in there being many potential threats to the validity of the findings. We then describe a number of different research designs that vary conditions in terms of training (training vs. non-training) and measurement timing (pretest vs. posttest) that can test for various threats to the validity of the findings. Research designs that have adequate controls for training and measurement before and after training offer the best opportunity to deal with a range of threats to validity and to be able to infer causality. The Solomon design is clearly the most able to deal with threats to validity but we note, in our literature search, very few such studies in relation to leadership training. We can understand the enormous practical problems of achieving such a design but urge researchers to utilize such a design where possible. In addition, while we are aware of many well-designed studies that can cope with various threats to validity, we found it interesting that these studies often do not test for and check for these threats. There seems to be a logic that if the design controls for it, it must not be there! This we disagree with, and many threats (such as instrumentation and mortality) should be tested prior to conducting hypothesis testing.

A key factor in determining causality is the randomization of leaders into training vs. non-training conditions. Given sufficient sample size, randomization controls for many known and unknown factors that might vary between conditions and explain findings. We further note how difficult it is to achieve such randomization in field studies but urge researchers to try to achieve this when possible (e.g., Barling et al., 1996; Yeow & Martin, 2013). One potential solution to this prob-
lem might be to stage training sessions over time so that all managers receive training at different time points. Such a design, coupled with repeated measurements of outcomes, could be very powerful at handling a number of threats to validity and also resolve the ethical implications of only training a subsection of managers.

In this chapter we have specifically not commented on the content of leadership training (with the exception of ILTs and IFTs) as our focus has been on how one might infer causality from such research. We wish to make some comments on this. First, as noted in the Collins and Holton (2004) meta-analysis, the majority of leadership interventions cover general managerial skills rather than specific leadership ones. This may reflect the need to meet client needs to have content bespoke to the organization but makes comparison across studies very difficult. In addition, there is a theoretical need to test leadership theories in applied settings to show that changes in key leader behaviors cause predicted changes in work outcomes (e.g., Dvir et al., 2002; Graen et al., 1982). Furthermore, in our review we noted that leadership training does not adequately consider the role of automatic/implicit processes and training, generally focusing on the reflective (conscious/deliberative) system, whereas the reflexive (non-conscious/automatic) system is underplayed (Hodkinson & Healey, 2011). We describe a number of ways to improve this through such techniques as Conditioning and Selective Prototype Activation.

In conclusion, this chapter recognizes that there is considerable research evidence to state there is an association between leadership training and important work outcomes. However, the size of this effect varies due to many factors and we note that studies vary considerably in terms of their content (and often it is a-theoretical). We describe some of the key conditions to be able to infer causality and describe some research designs that are able to control threats to the validity of findings. We identify the process of random allocation of managers into training and non-training conditions as the one that is key in inferring causality and most problematic in the field. However, we do note several well-designed studies that embrace many of these ideas. We urge researchers to be more mindful of these issues in designing research in this area.

References


Martin, R., & O’Brion, L. (under review). Developing relationship-based leadership: A field experimental test of leader-member exchange (LMX) theory.


**Table 1: Internal Threats to Validity.**

<table>
<thead>
<tr>
<th>Threat</th>
<th>Description</th>
<th>Example Application to Leadership Training</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>History</strong></td>
<td>Events, other than the experimental treatments, influence the results.</td>
<td>Ongoing changes in organization, other than the training, affect changes between pretest and posttest (such as changes to product, processes, market conditions, and working conditions).</td>
</tr>
<tr>
<td><strong>Maturation</strong></td>
<td>During the study natural changes occur within the participants that affect outcomes.</td>
<td>If the evaluation period (e.g., pretest vs. posttest) is very long, there may be developmental improvements within the managers and/or team members that are independent of the training intervention.</td>
</tr>
<tr>
<td><strong>Testing</strong></td>
<td>Exposure to a pretest or intervening assessment influences performance on a posttest.</td>
<td>Pretest measures (e.g., leader’s relationship skills) make salient factors that the followers had hitherto not considered which renders them more sensitive to these factors post-training (e.g., if manager shows improved relationship skills).</td>
</tr>
<tr>
<td><strong>Instrumentation</strong></td>
<td>Testing instruments are not consistent; or pretest and posttest are not equivalent.</td>
<td>Different types of measures or measurement techniques are employed pretest and posttest to capture similar constructs.</td>
</tr>
<tr>
<td><strong>Statistical–Regression</strong></td>
<td>Scores of participants are initially very high or very low and tend to regress toward the mean during retesting.</td>
<td>If managers have very high scores on leadership capabilities involved in training, then these scores are unable to improve over time and can only decrease giving false estimates of change.</td>
</tr>
<tr>
<td><strong>Selection</strong></td>
<td>Systematic differences exist in subjects’ characteristics between treatment groups.</td>
<td>If allocation of leaders between training and non-training conditions is not random, then pre-existing differences between the conditions might explain difference in outcomes.</td>
</tr>
<tr>
<td><strong>Experimental Mortality</strong></td>
<td>Participants’ attrition may bias results.</td>
<td>Some team members (e.g., poorer performers) leave the organization during the evaluation period such that the sample populations are different pretest and posttest.</td>
</tr>
<tr>
<td><strong>Selection–Maturation Interaction</strong></td>
<td>The selection of comparison groups and maturation interact so that some groups change differently to others.</td>
<td>Team members in a leader non-training condition mature (i.e., change naturally over time) at a different rate than those in the leader training condition.</td>
</tr>
</tbody>
</table>