Beyond Exit Surveys: A Critical Examination of Current Evaluation Practices for Diversity Certificate Training Programs in Higher Education

A DISSERTATION
SUBMITTED TO THE FACULTY OF
THE UNIVERSITY OF MINNESOTA
BY

Molly M. Illes

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

Jean A. King
Advisor

February 2017

© Molly M. Illes, 2017

Acknowledgements

It has been an incredible privilege for me to attend graduate school and receive my doctorate. I am sincerely grateful to the many people who offered support during my odyssey. During graduate school, many professors, especially Samuel Myers, Enid Logan, Jarrett Gupton and Kathy Fennelly inspired me to develop my identity awareness, including acknowledging my own power and privilege as a white woman.

Without the ten participants, this study would not have happened. I am grateful for their honesty and willingness to share information about their programming. I am truly inspired by their passion and unflinching commitment as the catalyst leading the change at their intuitions.

I am beholden to my committee members, Gerald Fry, Melissa Chapman Haynes, and Karen Storm. Moreover, I extend heartfelt thanks to my advisor, Jean King. She welcomed me into the doctoral program, taught, nurtured, and challenged me. She gave me the foundation from which I have blossomed.

Many friends and family offered words of encouragement and support throughout this long journey. I want to thank members of my humanist community, the First Unitarian Society of Minneapolis for their support and especially Laura Butterbaugh for her friendship and support.

My dissertation writing support group was invaluable. The six other women who shared this extraordinary experience with me showed me that some aspects of the process are universal, regardless of the field of study or topic. I am so very grateful to you all for pushing me to write every day and helping me to see that I was not alone on this journey.

My children George and Margit were just three and four years old when I started graduate school. They have been incredibly supportive, especially when attending meetings with me or carrying large stacks of books back to the library. I cherish you both and I look forward to watching you grow through your own academic journeys.

Thank you to my siblings and extended family for their support. Without the love, kindness and support from my mother, Mimi Cerny, and my husband Mike, I might not have finished the program. I love you always and will never take either of you for granted.

The one way I can repay the many kindnesses and support I have received is to use what I learned in my graduate studies as well as through this dissertation to focus on evaluating systemic transformation toward an inclusive, respectful, equitable culture in which all people are supported and thrive.

Abstract

In contemporary American society, evidence of bias within educational institutions abounds (Douglas & Halas, 2013; Hardie & Tyson, 2013; Martin, 2009; Milner, 2013; Ropers-Huilman, Winters, & Enke, 2013; Rothstein, 2013; Wildhagen, 2012). Educational institutions and their leaders articulate a commitment to fostering an inclusive, equitable environment. As a result, diversity and training departments develop programs to help individuals gain the skills and knowledge to change their behavior in the workplace and beyond.

This study highlights a problem facing diversity training staff and evaluators. Current practices based on practitioner reports indicated that diversity program evaluations were not measuring the outcomes of their program on individuals or institutions. They reported changes such as enactment of new policies or the numbers of participants in the program, implicitly suggesting that these measures translate to changes in the organization's culture, without offering any evidence to support such a claim.

Scholarly human resource literature posited generic training evaluation models, such as Kirkpatrick's (2006) four-level model created more than fifty years ago. After analyzing and contrasting additional training evaluation models (Guskey, 2014a; Hamblin, 1974; Phillips & Stone, 2002), it was noted that little scholarly research exists addressing how, if at all, these models may fit for a diversity certificate training program evaluation considering the nuances and politics related to diversity programming in higher education.

The purpose of this study was to understand how managers evaluated the outcomes of their diversity certificate training programs in higher education. It is important to note that the purpose of the study was not to assess the outcomes of the programs themselves or delve into which practices in the various programs may or may not be promising. The study employed a grounded theory approach to develop an evaluation framework. Nine higher educational institutions throughout the United States participated in this study.

DCT program managers focused primarily on gathering formative feedback for program improvement. Study participants relied on informal ways of knowing their programs were having an impact. They used their intuition and made implicit assumptions, such as when a program is full, that equates to the quality and reputation of the program. Furthermore, they described informal conversations with past and current program participants that reassured the training staff that their program was having an impact. Study participants described lack of knowledge and skills as well lack of staff time as key perceived barriers to conducting evaluations.

The result of this study is a framework depicting its findings including context, supports, barriers and opportunities for DCT evaluation. Opportunities to incorporate varied DCT evaluation methods are discussed. As little has been published about this model of diversity training, suggestions for future research include conducing and publishing results from DCT program evaluations. Using multiple methods to conduct the evaluation and publishing the results would contribute to the evaluation knowledge base as well.

Table of Contents

Acknowledgements ... i
Abstract ... ii
List of Tables ... v
List of Figures ... vi
Chapter I: Introduction ... 1
Study Overview ... 2
Problem Statement ... 3
Significance of the study ... 4
Research Purpose and Questions ... 19
Chapter Two: Literature Review ... 24
Chapter Three: Methodology ... 49
Grounded Theory ... 50
Data Collection ... 61
Data Analysis ... 62
Study Limitations ... 62
Chapter Four: Study Findings ... 65
About the Participating Institutions ... 66
Diversity Certificate Training Program Description ... 70
DCT Program Evolution and Goals ... 74
Diversity Certification Training Program Components ... 76
Diversity Certificate Training Evaluation Practices ... 81
Chapter 5: Discussion ... 100
Discussion of the Findings ... 100
DCT Evaluation Practices Framework ... 104
Diversity Certificate Training Evaluation Design Opportunities ... 107
Suggestions for Future Research ... 117
Conclusion ... 119
Epilogue ... 121
Bibliography ... 123
Appendices ... 136
Study Participation Consent Form ... 136
Diversity Certificate Training (DCT) Practitioner Interview Instrument ... 139

List of Tables

Table 1. Institution Demographics 67

Table 2. Institutional Commitment to Diversity and Inclusion 69

Table 3. Diversity Programming Comparison 73

Table 4. Diversity Certificate Training Program Attributes 77

Table 5. Comparison of Unique Program Attributes 78

Table 6. Diversity Certificate Training Staff Job Focus 80

Table 7. Evaluation Practices 83

Table 8. Study Findings 100

Table 9. DCT Evaluation Design 108

List of Figures

Figure 1. Diversity Certificate Training Evaluation Practices Theoretical Framework 106

Figure 2. DCT Program Evaluation: Individual Participant's Narrative Approach 112

Chapter I: Introduction

On February 8, 2017, an undergraduate student at the University of Minnesota Twin Cities campus (UMN.edu, February 9, 2017) returned to his dorm to find a picture depicting a swastika and a concentration camp with the words "Nazis Rule" sprawled across the student's whiteboard. The student is Jewish.

The University posted a statement about the incident stating that it was "not an isolated incident," but one of seven reports "involving swastikas, neo-Nazi propaganda, and other anti-Semitic graffiti since the beginning of December [2016]." The statement also recognized that, "…this kind of graffiti and expression is also taking place on other campuses across the nation" (UMN.edu, February 9, 2017). There has been an increase in overt bigotry since the election of the new president in November 2016 (Miller & Werner-Winslow, 2016).

"My 12-year-old daughter is African American. A boy approached her and said, 'Now that Trump is president, I'm going to shoot you and all the blacks I can find.' This account was given by a mother in a report published by the Southern Poverty Law Center (Miller & Werner-Winslow, 2016). This incident was one of 867 hate incidents that occurred between the United States Election Day on November 8th and November 28, 2016. K-12 schools as well as colleges were the "most common venues for hate incidents" (Miller & Werner-Winslow, 2016). Of the total hate incidents reported to SPLC, 280 were anti-immigrant; 187 were anti-black; 100 were anti-Semitic; 95 were anti-LGBT; 49 were anti-Muslim; and 40 were anti-woman. The report documented white nationalist incidences as well. "White nationalists have openly embraced Donald

Trump and following his election victory, the language, literature and symbols of white nationalism have cropped up throughout the country" (Miller & Werner-Winslow, 2016).

Hate crimes are not a new phenomenon in the United States (Hall, 2013; Whitley & Kite, 2010). Differential treatment and systemic bias have been deeply engrained in U.S. culture as well (Alexander, 2012). In the last decade, educational outcomes showed widening gaps between students of color (Snyder, de Brey & Dillow, 2016) in relation to white peers. The work of racial justice advocates, educators, and educational institution administrators may be even more challenging under the current United States president who has "normalized hatred" (Graham, 2016) and "bigotry" (Washington Post Editorial Board, 2016).

Study Overview

This study presents a critical examination of diversity training evaluation practices, specifically diversity certificate training programs. The diversity certificate training model was chosen due to its similarities across institutions compared to the many different types of diversity training that otherwise exists. The informants who participated in the study were all in higher education settings, working with a diversity certificate training program. Higher educational institutions with DCT programs were intentionally chosen to enable comparisons across institutions.

Prior to conducting the study, I reviewed practitioner and scholarly reports evaluating diversity training. During this critical review of existing program evaluations, I noted a trend that diversity training evaluation reports reflected the number of participants and relied on implicit assumptions that the participants gained knowledge

and skills that they then transferred from the training to the workplace or to the participant's personal life. The culmination of the initial review inspired the study, thus offering a critical examination of diversity certificate training program evaluation practices in higher education.

Problem Statement

Systemic bias and racial discrimination exist in higher education (Cabrera, 2014; Hikido & Murray, 2015; Harper, 2012; Haider, Sexton, Sriram, Cooper, Efron, Swoboda, Villegas, Haut, Bonds, Pronovost, Lipsett, Freischlag, & Cornwell, 2011; Posselt, Jaquette, Bielby, & Bastedo, 2012; Rubineau & Kang, 2012). Diversity programming is developed to build tolerance among diverse stakeholders (Clements & Jones, 2008) and diversity training is viewed as a significant part of diversity work (Kulik & Roberson, 2008). Some scholars argue that "too much of the popularity of diversity management training is symbolic in order to convey a message of top managerial support" (Ivancevich & Gilbert, 2000), thus leading people to view the institution as supportive of diversity without addressing underlying problems.

Appropriately evaluating diversity training outcomes is vital for two reasons. First, program staff needs to understand what improvements might be made for the programming to increase the likelihood of achieving the desired outcomes. Second, the training evaluation must measure whether or not the program's goals are met, including whether or not training transfer occurred and a change in work behavior was achieved (Roberson, Kulik, & Pepper, 2009).

Formal evaluations have not documenting measured outcomes from diversity training. Diversity practitioner reports (Cordes, 2009; Leiderman, & George, 2005; Ngo, Myers, Chang, Orville, & Illes, 2012) included evidence of adoption of new policies relating to diversity in the workplace, and they referenced the numbers of participants in training programs. Program practitioners asserted implicit assumptions about the value of the training itself, yet offered little evidence to illustrate program outcomes, whether positive or negative, on training participants. This issue matters because evaluations inform program developers, policy makers, and funding decisions that have the potential to dismantle or support programs that ultimately may affect the work climate in which a program exists. Furthermore, diversity programming exists in institutions that are political. Therefore, diversity program practitioners should gather evaluation data to demonstrate program outcomes. However, the practice of program evaluation relating to diversity training has unique challenges and barriers due to emotional and political differences in worldviews toward race and ethnicity. Roberson, Kulik, and Tan (2014) asserted that measuring such outcomes is challenging. If diversity programs are intended to transform individuals and ultimately the institutions in which they work, the lack of evidence about individual and institutional outcomes provided in diversity program evaluation reports is a problem.

Significance of the study

This study contributes to the knowledge of evaluating diversity programing, specifically diversity certificate training programs, and offers insight into current diversity certificate training evaluation practices. For evaluators, the study may offer new opportunities for designing robust evaluations that will measure both program outcomes

as well as formative feedback for program improvement. For training program managers and researchers, evaluating diversity certificate training programs utilizing methods beyond an exit survey may help demonstrate program effects, including training transfer and skill application as well as unintended consequences of DCT programs that might otherwise be unknown. The implications of this study for higher education and the broader society are to make diversity programming more effective, in turn creating positive outcomes, and ultimately becoming a more inclusive society.

Program Evaluation Defined

Michael Scriven (1967), in his foundational work on program evaluation, offered an early definition of the field, suggesting that evaluation measures the worth and merit of a program. He also contributed the concepts of formative and summative evaluation. He suggested that formative evaluation is an inquiry focused on feedback for program improvement. Summative evaluation serves the purpose of providing evidence to determine the overall merit or worth of the program. Patton (2002) defined program evaluation as the "systematic collection" of information about a program. Weiss (1998), Henry (2000), and Mertens (2009) suggested that program evaluation can be a tool to understand social transformation and improvement for people who are historically oppressed, whereas Trochim (1998) acknowledged the political nature and potential barriers to practicing program evaluation. He defined program evaluation as a profession that exists in "decision-making contexts that are inherently political and involve multiple often conflicting stakeholders, where resources are seldom sufficient, and where time pressures are salient" (p. 248). Preskill and Torres (1999) defined the field of program

evaluation in a learning context, such as an organization delivering training to its employees. They posited that evaluation is

> …an ongoing process for investigating and understanding critical organizational issues. It is an approach to learning that is fully integrated with an organization's work practices, and as such, it engenders (a) organization members' interest and ability in exploring critical issues using evaluation logic, (b) organization members' involvement in evaluation processes, and (c) the personal and professional growth of individuals within the organization. (pp. 1-2)

Thus, the practice of program evaluation exists in contexts that are political and with limited resources (Trochim, 1998) and in organizations that seek to understand critical issues, to steward professional and personal growth (Preskill & Torres, 1999), and to further social betterment (Henry, 2000; Weiss, 1998).

The Context for Diversity Programming in the United States

Equity planning and training and occurring in educational systems across the United States. In response to the changing demographics in schools, kindergarten through high school administrators have developed equity plans to better meet the needs of diverse students. School districts such as San Jose Unified School District (Murray, 2004) or Castro Valley Unified School District (2004) both in California, Portland Public Schools (2014), Montgomery County Public Schools in Maryland (2016), and Washnaw Intermediate School District in Michigan (2013) have created race equity plans to guide the work of their districts. Each plan recognizes the importance of training teachers to better meet the needs of the diverse students in their respective districts.

Equity planning that included teacher training and development require a financial commitment from the district for the initiative. For example, the Saint Paul Public

School Board passed a race equity policy (www.spps.org, 2013) heralded by the *Pioneer Press* as "groundbreaking" and a "one of its kind" initiative. The policy was supported by a "seven-figure investment in professional development to help staff members confront their racial prejudices" (Koumpilova, 2013).

Large-scale educational equity initiatives like the Saint Paul Public Schools invest significant amounts of money into teacher training. Diversity or cultural responsive training is not unique to education. Many organizations have conducted diversity training for staff to help them better meet the needs of diverse customers. According to the American Society for Training and Development (ASTD) State of the Industry Report (Miller, 2014), in 2012 the estimated expenditure on employee training was $164.2 million dollars in the United States. Employees spent an average of 33.3 hours training annually. These data indicated that training was a significant budgetary concern and function in organizations, with millions of dollars invested in it annually. The figures included diversity training efforts. How much money is allocated annual for diversity training is unknown. According to the Association of Training and Development (ATD), national statistics on the investment in diversity training versus other types of training have not reported (T. McCutchen, personal communication, May 5, 2016).

Other institutions conduct diversity training to reduce the likelihood of a lawsuit. According to the United States Equal Opportunity Commission, more than 90,000 discrimination charges were filed annually since 2008, with 93,727 charges filed in 2013. The charges were based on all forms of discrimination, including race, ethnicity, religion, sex, national origin, color, age, and disability

(http://eeoc.gov/eeoc/statistics/enforcement/charges.cfm). Legislation such as the 2009 Lilly Ledbetter Fair Pay Act, the 1990 Americans with Disabilities Act or the 1964 Civil Rights Act banned systemic discrimination.

Despite the prevalence of racial gaps in educational achievement and outcomes (Condron, Tope, Steidl, & Freeman, 2013; Morgan, Farkas, Hillemeier, & Maczuga, 2016; Snyder, de Brey, & Dillow, 2016; Wilder Research, 2011), legislation to address inequity within systems, legal action, and reports from diversity training efforts show little evidence of changes within the systems in which the gaps are perpetuated. One barrier noted above is that program evaluation by its nature is highly political (Mertens, 2009; Russ-Eft & Preskill, 2009), and evaluating diversity programming has challenges that other types of training do not have.

This dissertation presents an analysis of a sample of diversity training evaluation reports developed by practitioner diversity trainers and or evaluation professionals (for example, see Charles Mott Foundation, 2007; Ngo, Myers, Chang, Orville, & Illes, 2012; Ramsey County, 2010). The reports included details about the training, such as number of participants, and often indicated the participants' experience or the number of policies derived from the training. However, the training reports failed to address training transfer of learning and behavior change in the work environment. There was an implicit assumption that training affected behavior that led to a shift in cultural normative behavior within the workplace. However, institutions did not design and conduct evaluations that were capable of measuring such outcomes, especially the longitudinal effects of diversity certificate training over time.

Diversity training is different from other types of training in that its goals are to change attitudes, knowledge, and behaviors relating to how people interrelate with people of different races. Its ultimate goal is to shift the normative behavior within an institution toward inclusion. Social theory (Berger & Luckmann, 1966) suggests that actors are producers as well as products of society and its structures. Without an individual's shift in behavior, institutional change may not occur at all, asserted Berger and Luckmann (1966) in their landmark book, *The Social Construction of Reality: A Treatise in the Sociology of Knowledge*. The researchers theorized that if change does occur, it may not be sustained. Furthermore, researchers propounded that "[r]eality is socially defined. But the definitions are always embodied, that is, concrete individuals and groups of individuals serve as the definers of reality" (p. 107). In their discussion about cultural differences, Berger and Luckmann contributed the notion of "monopolistic situations" in which one dominant view prevails. They argued that when segregating the views of the "other," stability in the social structure of a system continues; institutions remain cohesive (p. 60). "Traditional definitions of reality inhibit social change," yet "breakdown in the taken-for-granted acceptance of the monopoly accelerates social change" (p. 113). Thus, if a diversity training evaluation provides substantive evidence of transfer of learning and behavior, the results then may support a shift in cultural normative behavior within the institution toward inclusion. How practitioners evaluate diversity training transfer presents a challenge in the field of evaluation. If training transfer does not occur or it is not sustained, the struggle between competing views will continue for a long time as the views may be deeply embedded in the group's ideology.

Thus, the problem I address in this dissertation is the challenge of accountability in the evaluation of diversity certificate training programs and the lack of measuring program outcomes. Berger and Luckmann (1966) posited, without demonstrating a shift in social normative behavior and long-term change, diversity initiatives lack the evidence to support the effect their programming has on individuals and, ultimately, institutions in which the training occurs.

Context: Diversity in the United States Today

There is a debate in United States today about the role of race as a barrier to full participation in society. Some argue that we are now post-racial and that race is no longer the leading indicator of one's success, but that success is determined more by socio-economic factors (Wilson, 2012). Others argue that race is still a significant factor in determining the extent to which a person can reach his or her ultimate potential as a productive member of society. Scholars such as William Julius Wilson (1978) studied the Black lower class in relation to the emerging Black middle class. He examined class as a distinguishing factor at the intersection of race. His recent publication (2012) concluded that some social issues like welfare dependency and "out of wedlock" births cannot be accounted for with an "easy explanation of racism" (p. ix). Other scholars determined that racism was not a barrier. One empirical study (Fernandez & Fernandez-Mateo, 2006), for example, found no evidence of job hiring preferences based on race. However, the unskilled job setting predominantly employed people of color and relied on social networking of its employees to attract new hires. Another study (Stainbeck & Irvin,

2012) found that work environments with predominantly the same race had a decreased perception of racial discrimination.

In contrast, other studies conducted in educational settings offer evidence of racial bias, racism, and racial segregation (Douglas & Halas, 2013; Hardie & Tyson, 2013; Martin, 2009; Milner, 2013; Ropers-Huilman, Winters, & Enke, 2013; Rothstein, 2013; Wildhagen, 2012). These studies indicate the continued relevance of diversity programming and its impact on participants as a means to overcome the institutional bias in everyday transactions between members of the dominant race and people of color.

Stratification due to competition for admission continues to reinforce racial inequity in post-secondary enrollment (Posselt, Jaquette, Bielby, & Bastedo, 2012). Harper (2012) conducted interviews with 219 Black male undergraduate students at 42 institutions from 20 different states. He concluded that post-secondary faculty and higher education administrators were unprepared to address and reduce racial inequities in education. In a different study, the authors examined the impact of race-neutral admissions. Garces and Cogburn (2015) conducted research to understand the impact of the Michigan State Constitutional amendment, Proposal 2. This amendment passed by voters in 2006 and prohibits racial or ethnic consideration during public higher educational institution admissions process. The study found the constitutional amendment negatively influenced other diversity work critical to the success of students of color in higher education.

Studies also indicated the prevalence of racial bias among students. One study (Haider et al., 2011) found that 69% of medical students demonstrated an implicit

preference toward white people. Another study measured differences in medical student bias between first year and second year practice in their service delivery. The longitudinal study determined that medical students increased their racially-biased service discrimination (Rubineau & Kang, 2012). Research on racial bias demonstrated that white students sought to maintain their dominance in higher education settings (Cabrera, 2014). Hikido and Murray (2015) determined that white students in one large university felt excluded and attempted to hold on to their privilege. They suggested that institutions must include pedagogical measures to facilitate white students learning how to overcome such tendencies.

Research has documented that racism shifted from overt, Jim Crow segregation to covert, colorblind racism in contemporary United States society (Alexander, 2012; Barndt, 2007; Bonilla-Silva, 2010; Delgado & Stefancic, 2001; Feagin, 2010; Gallagher, 2006; Logan, 2011). Colorblind racism is the perception of Whites that the socioeconomic playing field is now level (Gallagher, 2006), that racism is as destructive as before, but is now hidden and more sophisticated (Barndt, 2007) and espouses the belief that we should treat all people equally, regardless of race (Delgado & Stefancic, 2001). Feagin (2010) suggested that colorblind racism does not allow one to see the actual racist behavior despite the fact that it exists in everyday interactions.

Significance of the Issue

The lack of supporting evidence of individual and systemic change resulting from diversity training in institutions is an issue because racial gaps in educational achievement persist (US Department of Education, 2016). This is a particular concern

because based on 2010 census data projections, the United States will shift from Whites as majority to minority in 2043 (Cooper, 2012; Muskal, 2012; Yen, 2012). In just thirty years, Whites will no longer be the dominant racial group in the United States. “We live in a time when our social perspective is increasingly global in nature, national norms, behavior, and approaches to race, identity, and racism have not been transcended and may never be” (Wijeyesinghe & Jackson, 2012, p. 4), and race as a social construct carries significant social, political and group meaning. Furthermore, “from psychological, sociological, social psychological, and human ecological standpoints, racial identities matter in the study of human development” (p. 15).

Banks (2006) suggested that diversity provides educational institutions an opportunity to foster an environment that “reflects the reality of the nation and the world” (p. 201) and facilitates people getting along. Yet that diversity also poses challenges to schools as students and teachers bring their “stereotypes, misconceptions and negative attitudes toward outside groups” with them (p. 200).

Diversity Training as Remedy to Discrimination and Institutional Bias

To remedy perceived institutional racism or racial bias, organizations frequently implement diversity training intended to educate employees about inequality and effect change in the institutional structure toward inclusion, eliminating the subordination of people of color. The Saint Paul Foundation’s Facing Race initiative, the Higher Education Anti-Racism Team initiative (H.E.A.R.T), Ramsey County Human Services, and the Charles Mott Foundation’s Crossroads reports are examples of large-scale

programs intended to transform systems that the staff of the organizations themselves determined were racially biased. An integral part of every diversity initiative is training.

A diversity initiative sponsored by the Levi-Strauss Foundation (Leiderman & George, 2005) resulted in a report that touted Project Change as the “first national foundation/community partnership explicitly aimed at reducing institutional racism as one of its four major goals” (Leiderman & George, 2005, p. 5). Organizations in four cities implemented diversity programming: Valdosta, Georgia; Knoxville, Tennessee; Albuquerque, New Mexico; and El Paso, Texas. The report indicated that the regional organizations established fair lending centers, resulting in increased home ownership for people of color, implemented conflict resolution training in schools, created groups to educate the public about hate crimes, established financial counseling centers, and worked to raise awareness in local media on how to portray people of color. Two of the other three programs addressed training as a key outcome of their programming. In Knoxville, for example, the report states that 200 people were trained and participated in furthering diversity work (p. 22). The Project Change report provided details about actions and the evolution of structures intended to promote diversity. It did not, however, convey any details about training transfer of skills or knowledge from the training to the workplace or whether or not program outcomes were achieved.

In 1992, Delgado and Browne published a diversity program evaluation report through the Applied Research Center in Oakland, California. This report was not from a peer-reviewed journal, and it is also twenty years old, which may affect its application to

research today. However, there are several points in the evaluation report that may be considerations that relate to the problem of accountability in diversity training evaluation.

The report contained evaluation data and analysis from ten diversity programs around the United States. In the report, Delgado and Browne offered a typology of diversity work that included prejudice reduction work, focused on changing individual attitudes; diversity training/diversity management, concentrated on appreciation of cultural differences; and development of educational material that was reflective of the historical contributions of both genders from a variety of cultures. Diversity management combined prejudice reduction and organization integration; opposition to racist violence in which organizations assist communities of color in responding to organized initiatives of violent racial bigotry; antidiscrimination work that focused on issues of racial discrimination and segregation; self-determination initiatives building viable independent organizations that can exercise their own countervailing power; and anti-apartheid work relating to racial conditions in Southern Africa. The distinct typologies could be considerations for diversity program evaluation since each program may have different goals and thus varied outcomes, yet still require training accountability.

Delgado and Browne (1992) offered a hypothetical example of how a social worker may—or may not—be affected long-term by a workshop. They concluded

> The sessions do not address the relationship between a specific person and the institutional arrangements which give them power in specific situations. They do not demystify institutional racism. Not only is there an unproven assumption that an attitude change correlates with a change in behavior—we also find that power relations are obscured. In this case, personal relations are power relations only because the social worker is wielding the institutional power of the state. (p. 55)

The ”unproven assumption” as well as recognition of power relations in institutional settings exemplified the need for a robust evaluation of diversity training The authors felt that the prejudice reduction training they were evaluating was not sufficient to address issues of power. “Most prejudice reduction training, while useful in introducing people to some dimensions of racism, is not sufficiently sophisticated to address the nuances of institutional racism or to address issues of power therein” (p. 58). This notion that the training itself is insufficient also harkened to the issue of lack of clearly articulated program theory and effective assessment of training or workshop impact based on that theory, which is the concern I will address in Chapter Five.

Over 20 years ago Delgado and Browne (1992) concluded that little scholarly evidence existed and few studies offered step-by-step examinations of goals, methods and results, despite the fact that corporations and organization undertaking such training and diversity transformative programming claimed to have been successfully “diversified” (p. 60). They suggested researchers conduct case studies of diversity interventions to understand power dynamics better within an institution, as well as the organization’s motivations for change. Finally, the authors recommended that future researchers and evaluators should work to demystify institutional racism. They asserted that

> [T]his is a very complex problem requiring a combination of grounded analytical work and an understanding of the theories and practice of organizational development, policy analysis, and social psychology… However, until we have a working understanding of the dynamics of institutional racism, many of our efforts in the field of antiracist work will be, at best, misdirected, and at worst, counterproductive. (pp. 60-61)

In Saint Paul, Minnesota, the Facing Race initiative had an annual budget of approximately $800,000, largely supported by internal resources and the W. K. Kellogg Foundation (R. Shiphandler, personal communication, June 27, 2013). The evaluation report stated the program's goal was to "reduce racism at both the individual and institutional level through education, direct action and/or increased dialogue and understanding" (Cordes, 2009, p. 3). In the report, a key recommendation is to "address racism at the individual level," and the report authors recognized the need for funding diversity program evaluation. Yet the report offered little evidence of the training impact on participants. They did not include diversity training participant outcomes or attend to training transfer in their evaluation.

Three additional practitioner reports (Charles Mott Foundation, 2007; Ngo, Myers, Chang, Orville, & Illes, 2012; Ramsey County, 2010) offered similar information and conclusions to the Facing Race report. The H.E.A.R.T. report was commissioned by Bremer Foundation to evaluate a collaboration of diversity initiatives at four separate institutions, including Minnesota State University Moorhead, North Dakota State University, Saint Cloud State University, and Bethel University. The evaluation team conducted interviews with leaders in the diversity initiatives at each of the four institutions. Evidence was offered regarding the programming's effect on the institution, including policies enacted, evolution of teams to combat racism, and other grass-roots initiatives. The report addressed the process of training administrators and staff, but lacked documentation about whether or not training program outcomes were achieved.

Ramsey County Human Services adopted a diversity initiative to transform the County service organization and the way staff serve families of color. They adopted a system-wide staff training program as well as leadership development (Ramsey County, 2010). Part of their training included "authentic dialogues," offering staff an opportunity to discuss topics that challenged their worldviews. When asked how they evaluated the outcomes of the dialogues and other trainings, a Ramsey County Human Services evaluator indicated that they were unsure how to evaluate the training's effects on individual participants. Ramsey County Human Services viewed participation in diversity training as a component of their transformative efforts, yet they were unable to fully evaluate the training program's outcomes.

Crossroads executive director, Robette Ann Dias, stated that "[t]he subject of evaluation in diversity organizing is a bit of a touchy subject as it does not lend itself to traditional forms of quantitative evaluation, yet this is typically what funders and decision makers ask for" (R. Dias, personal communication, June 20, 2013). All four reports reviewed for this study aligned with Dias' comment that their focus was on the number of policies changed, the number of program participants, or the experience of the grassroots organizers working toward change. The reports did not address the extent to which participants' behavior may have changed, nor did they identify obstacles or successes in the transfer of training learning.

Herein lies the problem: accountability for training outcomes relates to the evaluation design. In the reports reviewed earlier in this chapter, the design and process omitted gathering evidence to support whether or not the training conducted had achieved

its intended outcomes. . Since attitudes toward race are strongly associated with an individual's behavior (Faegin, 2010; Moniz & Spickard, 2006) and since social theorists maintain that without providing evidence of a sustained shift in behavior toward inclusion, institutional change is unlikely to be maintained if it occurs at all (Berger & Luckmann, 1966; Bicchieri, 2006), then evaluation reports must offer evidence that either a transfer of skills and behavior did or did not occur in order to make a judgment about the program's effectiveness. Part of the evaluation challenge is that attitude changes such as a shift in behavior regarding racial inequality are not easily evaluated (Bennett & Bennett, 2004). It is my argument that an effective evaluation of diversity training and, in this study, diversity certificate training outcomes is required if the institution espouses its institutional core value of diversity and the program's goals are to train individuals to embrace the core value and foster a culturally responsive workplace.

Research Purpose and Questions

As demonstrated in the review in this chapter, some key large-scale training evaluation reports lack evidence of training transfer of skills, learning and behavior. The literature review in Chapter Two of this paper offers a discussion of existing frameworks for evaluating training, as well as insight on the nuances of evaluating diversity certificate training. However, there is scant literature on evaluating diversity certificate training with particular attention to training transfer, its unique barriers, and a potential framework to help practitioners and scholars identify the barriers.

The purpose of this study is to understand how managers evaluate the individual and institutional outcomes of their diversity programs. This study looked at diversity

certificate training specifically since the goals and nature of the programs were similar, therefore, the evaluation approaches may have been more similar than disparate types of diversity training. This study employed a qualitative grounded theory approach to develop a framework to understand current evaluation practices, perceived evaluation barriers, and evaluation opportunities that emerged. Such a model may increase an evaluation's formative feedback for program improvement as well as increase accountability for the training and its associated outcomes. Social theory reinforces the need to demonstrate the transfer of training learning and ultimately behavior change since institutions are comprised of people, and it is the actors within the systems who generate the cultural normative behavior. Without a thorough evaluation that measures both the transfer of learning and behavior as well as transfer climate, practitioners may have little confidence in their training's outcomes.

Research Questions

This study explored the following research questions:

1. What is diversity certificate training, and how does it differ from other types of diversity programming?
2. How do practitioners evaluate their diversity certificate training programs?
3. What barriers inhibit measuring diversity certificate training evaluation outcomes?
4. What do practitioners perceive as components in a thorough diversity training evaluation?

Glossary

The following terms are referenced in this paper. I offer the following definitions for a common understanding of the terms as they relate to the information presented in this paper.

Complex adaptive system: A cluster of individual parts that interact with each other, and over time system-wide patterns appear…the system is constantly shifting. Stable, permanent reality is impossible (Eoyang & Holladay, 2013, pp. 15-17).

Complex systems: Complex systems are comprised of a series of components, including agents, heterogeneity, self-organization, feedback, and an emergent behavior (Wolf-Branigin, 2013, p. 175).

Cultural discrimination: One group within a culture controls the power to define cultural behavior and the form that behavior should take. Power and perpetuation of oppression are fostered in the media, literature, art, music, language, morals, customs, and beliefs to define an agreed-on way of life (Whitley & Kite, 2010, p. 17).

Developmental evaluation: An evaluation approach cultivated by Michael Quinn Patton that recognizes the programs and the institutions in which the programs exist are dynamic, adapting to complex, changing environments (Patton, 2011).

Diversity: The opposite of racism; it is the work of dismantling racism, removing institutional barriers that serve as obstacles for full participation by people of color (Barndt, 2007, p. 223).

Discrimination or racism: Differential treatment of a person based primarily on the person's membership in a social group, such as based on racial or ethnic affiliation (Whitley & Kite, 2010, p. 12).

Dynamical change: Complex change that results from unknown forces acting unpredictably to bring about surprising outcomes (Eoyang & Holladay, 2013, p. 62).

Dynamical systems: Systems are dynamical when they change in response to nonlinear, high dimension, and/or discontinuous forces. The sciences of complexity and deterministic chaos are being developed to understand these emergent and unpredictable systems. The traditional evaluation method that is most able to capture the dynamical nature of social systems involves stories and storytelling. The emerging science of complexity may provide other tools to track and understand patterns of dynamical change. (Williams, 2004)

Evaluation: The systematic collection of information about the activities, characteristics, and outcomes of programs to make judgments about the program, improve program effectiveness and or inform decisions about future programming (Patton, 2002, p. 10).

Formative evaluation: Initially coined by evaluation scholar, Michael Scriven, the purpose of formative evaluation is program improvement versus decision making (as in summative evaluation) (Russ-Eft & Preskill, 2009, p. 18).

Impact evaluation: An evaluation that assesses a program's effects and the extent to which the program's goals were achieved (Mertens & Wilson, 2012, p. 559).

Institutional discrimination or racism: Discrimination that occurs when institutions sanction beliefs about group superiority. This form of racism is rooted in the norms, policies and practices within a social institution such as a family, religious institution, educational or criminal justice system (Whitley & Kite, 2010, p. 16).

Massively entangled: Boundaries in human systems are massively entangled. One individual participates in multiple natural systems: work group, family, faith community, and so on. Emergent patterns within each of those boundaries shapes the individual who, in turn, shapes patterns as they emerge in the other contexts (Eoyang, 2006, p. 128).

Multiculturalism: The notion that individuals are motivated to retain their cultural heritages and that intergroup relations are optimized when people retain their cultural heritages in a pluralistic society (Whitley & Kite, 2010, p. 577).

Prejudice: An attitude toward people who are members of a specific social group (Whitley & Kite, 2010, p. 11).

System dynamics: Having value at a specific point in time, systems consist of elements that change over time due to internal and external exacerbations (Williams & Hummelbrunner, 2011, p. 46).

Training Transfer: Applying the knowledge, skills, and attitudes acquired during training to the work setting (Russ-Eft & Preskill, 2009, p. 89).

Conclusion

In this chapter, I presented the social problem that institutional discrimination persists in education. An institutional response to address discrimination is diversity training. The reports reviewed in this chapter illustrated that few diversity training and evaluation practitioners use multiple methods to measure the outcomes of their training programs. They focused on the numbers of policies changed or quantities of participants. The reports did not convey evidence of individual or institutional outcomes from diversity training beyond policy enactment.

Chapter Two reviews the literature of scholarly approaches to evaluating training, including a comparison of key training evaluation frameworks from leading scholars in human resource development, education, and evaluation studies. In Chapter Three, I review the methodology employed for this study, including information on sampling and study limitations. Chapter Four offers an overview of this relatively new type of diversity training model, including program descriptions, demographic information, program evaluation practices, evaluation barriers and evaluation opportunities. Chapter Five reviews and discusses the findings, reveals the theoretical framework developed for this study, presents suggested evaluation methods for DCT evaluations, and concludes the paper.

Chapter Two: Literature Review

Chapter One presented an analysis of diversity training practitioner reports. The review illustrated that frequently these reports do not reflect training impact on individuals, nor do they address training transfer of skills, learning, and behavior back to

the workplace. The reports listed numbers of policies enacted or numbers of participants, but fell short of offering evaluative evidence of training impact. This is a common issue in training evaluation. Bersin (2008) stated that

> . . . while an estimated 74% of organizations believe it is important to measure the job impact of training programs, only 14% actually do. Seventy-two percent of learning executives believe that it is extremely important to measure the business impact of training—while only 10 percent have any such measurement program in place. (p. 2)

Furthermore, research indicates that only between four and five percent conduct a return on investment as part of their evaluation strategy (Bersin, 2008, p. 4). Nevertheless, Kulik and Roberson (2008) acknowledged that "diversity training is often seen as the cornerstone of diversity initiatives" (p. 277) with institutional goals that stress the importance and benefits of diversity in the organization as well as an opportunity to inspire employee behavior change.

The purpose of this literature review is to provide a critical analysis of existing training evaluation frameworks, to review scholarly sources that illustrate the motivations and common attributes of diversity training, and to offer contextual information to set the foundation for this dissertation. Over the course of the last three years, I searched scholarly sources from the field of program evaluation, including the *American Journal of Evaluation, Program Planning and Evaluation, New Directions in Evaluation,* and several books authored by scholars in the field. I also reviewed literature from the human resource development field, including training development, business management, and organizational development sources. The third primary source I tapped in this literature review was diversity management articles and books.

There are two components to this literature review. In the first component, I review literature to give the study a description and context of diversity training and its associated evaluation challenges. Searching specifically for literature on evaluating diversity training revealed a dearth of scholarly studies relating to practices or frameworks for diversity training evaluation. This gap in the literature underscores the need for additional research. Second, I review training evaluation approaches that are not diversity-specific since there are few published practices specifically relating to diversity program evaluation. I reviewed multiple studies conducted using different approaches as well as descriptions of the methods from the scholars who developed the various frameworks.

The Nuances of Diversity Training

Diversity training differs from other types of training because culture and worldview play a significant role in diversity training and participant outcomes (Fowler & Blohm, 2004; McDermott & Samson, 2005; Park & Denson, 2009; Salas, Salazar, & Gelfand, 2013). Guess (2006) and Deardorff (2008) asserted that the training participant's background, knowledge about race, racial identity and willingness to learn and expand skills and behavior toward inclusion are significant issues with diversity training and its outcomes. Training effectiveness is impacted by whether or not the participant chooses to participate or is mandated to do so (Kulik, Pepper, Roberson, & Parker, 2007). Those who choose to participate, according to the study authors, are less likely to need the diversity training and have a lower measurable training impact because they likely were more advanced at the beginning of the training.

Another way in which diversity training differs from other training are the trainer skills required for successful delivery. Clements and Jones (2008) emphasized that trainers need conflict management skills, knowledge of one's own prejudice, and an ability to ask tough questions. The authors deemed that successful diversity trainers were resilient, sensitive and motivated, and well trained in the subject. The authors also noted that training must be conducted in various ways to meet the needs of the learner. Gallos (1997) also recognized that diversity training requires a different approach than other types of training. He stated that "teaching workplace diversity demands a different pedagogical stance and personal connection to the teaching task" (p. 3).

Bell and Kravitz (2008) contended that diversity training can have a positive, neutral or negative impact on participants and organizations depending on many variables, such as an individual participant's background and the organizational climate. These are issues in every type of training, but since "[t]eaching diversity is substantially more complex than teaching other kinds of courses" (p. 302), the issues specifically related to evaluating diversity training are heightened. The authors concluded that "few studies have assessed changes in skills and many of those studies used self-assessments. Self-assessments indicate positive changes, but objective measures do not" (p. 303).

The Importance of Context in Diversity Training Organizations

Culture and climate within an organization conducting diversity training are critical concerns within that organization. Relating to the organization's culture are the motivations to conduct diversity training, which may affect training and results as well. Some organizations viewed a diverse cultural client base as a leading motivation to train

employees (Kulik & Roberson, 2008). Other scholars posited that organizations may be motivated by external pressures, internal advocacy, the need for increased diversity, and the corporate culture, all of which play a role in whether or not an organization implements diversity training (Dobbin, Kim, & Kalev, 2011).

Balancing diversity training with the needs of staff and the diversity identity within a corporation is challenging. Cole and Salimath (2013) illustrated that corporations must balance diversity training to have effective training that does not turn people off, yet enough to support minorities in the workplace. Another consideration for workplace diversity programming is that leadership support is crucial (McGuire & Bagher, 2010). McGuire and Bagher argued that organizations should approach diversity training with a desired outcome to improve performance rather than “simply a social ideal” (p. 499). They also acknowledged that diversity training cannot be solely the human resource department’s responsibility, but that everyone in the organization should share responsibility. Salas, Salazar, and Gelfand (2013) asserted a similar conclusion, stating that “the broader organizational context can shape the effect of diversity training on behavior” (p. 11).

The Unique Challenges of Evaluating Diversity Training

Diversity training evaluation has unique challenges, and many evaluation reports and, as noted earlier, scholarly studies routinely omit evidence of how training affects participants and their transfer of skills, knowledge, and behavior to the workplace. Little is published on why this issue persists, which is a gap this study pursued. Similar to the practitioner reports reviewed in Chapter One, some scholarly studies lacked empirical

evidence of outcomes or the theoretical approach of how they evaluated the programming or course (Blitz & Illidge, 2006; Greene & Siskind, 2006; Jeyasingham, 2012; Peacock & Daniels, 2006). If studies omitted evidence, it is unknown how the authors reached conclusions or the extent of the effect their diversity programming had on participants. Roberson, Kulik, and Tan (2014) acknowledged that a significant challenge in evaluating diversity training is measuring individual outcomes of diversity skills. They underscored the importance of measuring transfer of skills, learning, and behavior over time to fully understand the outcomes of the training versus other factors that may result in a shift in organizational climate. Furthermore, they posit that three learning outcomes are rarely assessed within one study, including cognitive, affective, and skill-based outcomes. They stated that "[p]ost-training changes in skill and the transfer of diversity skill to the job have been the most neglected" (p. 355) and that they are challenging to evaluate.

Kulik and Roberson (2008) acknowledged the dearth of published research assessing diversity training effectiveness. Furthermore, they asserted that published research is "not easily accessed by HR practitioners or diversity managers" (p. 267). This may account for the lack of connection between the practitioner reports reviewed in Chapter One and possible rigorous evaluation designs that could guide the evaluation process. The authors recognized that evaluations were conducted and that published evidence exists, but the evaluations tend to be limited to quantitative measures. This is consistent with my analysis in Chapter One of this paper. Evaluation evidence tends to be based on survey data or counts of policies enacted, employee diversity, or other measures. "These outcomes tell us little or nothing about the processes by which diversity interventions affect demographic composition" (Kulik & Roberson, 2008, p. 268).

Pendry, Driscoll, and Field (2007) concluded that little is known about what makes diversity initiatives effective due to the lack of published articles on the topic and the research approach that does not contribute to an understanding of how and why programming works or fails to work.

Shapiro (2006) defined two types of diversity training approaches. One approach was an effort within an institution to shift the informal policies and practices of public institutions toward diversity. This shift was achieved through engaging training programs that focus on institutional leaders who collaboratively review the structure of racial privilege and oppression within the institution. In contrast, a prejudice reduction approach sought to address internalized oppression and ignorance through deeply personal workshops and dialogues and, ultimately, to shift attitudes and behavior.

An ideology persists because it was initially built upon the group's interests (Shapiro, 2006). Stated in a different way, Ting-Toomey (1999) built on Berger and Luckmann's (1966) social theory suggesting that institutional discrimination is community-prescribed discriminatory practices that are not isolated incidences, but routine among individuals protected by the community. Changing the "laws" does not guarantee a shift in behavior. Ting-Toomey suggested that to reduce prejudice in an institutional setting requires a "mindset analysis" to confront biases and ethnocentric attitudes, question stereotypes, and learn to use mindful or neutral language (pp. 169-170). However, the lack of evidence in the practitioner reports of a shift in individuals' attitudes toward race or of individuals undergoing any sort of mindset analysis leaves the

reader wondering whether or not the diversity policies enacted within institutions will be enforced and, ultimately, result in eliminating racial bias within the institutions' culture.

Shapiro (2006), Delgado and Browne (1993), and Pendry and Driscoll (2011) articulated that a lack of understanding regarding the program's theory in diversity programming and evaluation may be a concern. It is important to recognize that the nature of diversity programming is different from other types of programming since it is a contested social issue (Barndt, 2007; Faegin, 2010); people enter into the training with different attitudes (Deardorff, 2008) and life experiences (Hurtado, Engberg, Ponjuan, & Landreman, 2002); and participants in training and diversity work may have different motivations (Edwards, 2006). Some training participants may be unaware of diversity issues as a social concern and, as a result, may be unwilling to accept the tenets of the training. Evaluation scholar Carol Weiss (2002) acknowledged the tenuous nature of such social programming:

> Social programs that evaluators usually deal with differ in many ways from other organizations. Their aim is not so much to satisfy participants' expressed wants as to change participants' wants—and their knowledge and behavior…whether the participants want that outcome or not. (Weiss, 2002, p. 211)

Scholars at Murdoch University (2003) in Australia conducted a meta-analysis of diversity program evaluation reports. The authors stated that evaluations focused on what worked versus what did not work and failed to look at differences in participants' backgrounds, beginning knowledge, and feelings about racism, delivering the same educational strategies to all participants. They also cited a lack of post-intervention contact to see if attitudes and behaviors had changed. This report advocated for longitudinal evaluations to see if participants' behavior and attitudes reverted back to their pre-training disposition. In their analysis, as with most other studies I reviewed, they

did not make the connection to program theory, but offered more of a critique of the evaluation process.

Shapiro (2006) asserted that practitioners are challenged to see larger social changes and that change efforts take place at individual, group, or systemic levels. She made the case that "changing individuals involves strategies that shift attitudes and perceptions, feelings, behaviors and motivations of participants in an intervention. Programmes [sic], like the prejudice reduction …invoke a wide range of psychological and therapeutic theories in facilitating change during small-group interventions" (p. 5).

Pendry and Driscoll (2011) proposed five guiding principles to improve diversity training assessment. First, they stated that it is important to know not just whether or not the training works, but how and why. This is the crux of the problem I examine in this study. How and why training works or does not work is integral to accountability, decision making, and program improvement. Yet few examples illustrated an evaluation of how and why a training program works. To assess this, Pendry and Driscoll suggested developing a hypothesis to guide the assessment. However, they acknowledged that this is not often used because a hypothesis is rare in diversity training work, and most training programs do not develop in a quasi-experimental way. They stated that “the norm is to provide the training, not to assess it” (p. 2.06) and that assessment measures do not relate to the goals of the training.

A second guiding principle the authors recommend was to collect baseline data to facilitate measuring whether or not change has occurred in relation to each participant’s starting point. Their third point was that the goals, the actual training, and the outcome

measures must be aligned. “Many organizations do not match training needs to the training undertaken, or either one of these to assessment measures” (p. 2.08), which then leads to measurement issues. Short- and long-term considerations regarding when to assess the training outcomes were the fourth principle. They suggested that, when possible, assessments should occur over time to “establish the exact nature of any changes” (p. 2.13) as long as such assessment is feasible. The fifth point related to external factors that may affect diversity training outcomes. They offered the example that if management did not support diversity training, then the training was conducted to fulfill legal requirements and may impair training outcomes.

The authors asserted that considering differing motivations of training participants was another factor that may affect outcomes. Collecting demographic data may help understand differences in motivations, but it may also inhibit survey responses. Another element of this fifth principle was the importance of considering outside training or experiences that affect a participant’s views on diversity.

Beaudry’s (1992) article in *New Directions for Program Evaluation* reviewed research on evaluating diversity programming. He found three constraints that inhibited teacher adoption of multicultural education practices: individual, demographic, and institutional factors. “Program designers have been able to disseminate knowledge but seem to have had less impact on teachers’ classroom behavior, a finding consistent for pre-service and in-service teachers” (p. 74). This meta-analysis reinforced the importance of evaluating the individual’s shift in behavior toward diversity since knowledge can be learned, but shifting an individual’s views resulting in behavior change is not as easily

achieved. Gurin (1999) posited that it was not enough for individuals to take a diversity course, but it was the nature of the interactions within the course that determine whether or not participants shift their behavior for desirable civic outcomes. However, neither article offered a paradigm or tools to conduct an evaluation on an individual's shift in behavior resulting from diversity programming. Conversely, the articles highlighted the need to develop tools or an evaluation paradigm.

Trenberry and Paradies (2012) used an organizational assessment approach to manage diversity interventions in the workplace. The authors reviewed eight tools developed in the United States and Canada. Seven of the tools focused on assessing cultural competency in health care and service organizations, and one tool was for diversity-based organizational assessment. The authors were unable to find any organizational assessment tools that had a diversity or discrimination focus. They concluded that "organizational assessment tools are notably lacking in the diversity field. Current approaches to organizational assessment within cultural competence literature and diversity management lack an explicitly focus on addressing systemic racism" (p. 21). They also concluded that "further research is required to ensure that domains used in current organizational assessment tools are consistent with theoretical and practice-based standards" (p. 22), thus again reinforcing the need for applying program theory when evaluating diversity certificate training programs.

Deardorff's (2008) model of intercultural competence included a base of requisite attitudes: respect or the ability to value cultural diversity; openness to learning about other cultures; and curiosity about other cultures with an acceptance of ambiguity and

uncertainty. Knowledge, comprehension, and skills were also integral to intercultural competence, including self-awareness about one's culture, understanding of culture, and that which affects one's worldview. The skills needed for intercultural competence included listening, observing, interpreting, analyzing, evaluating, and relating. Deardorff's model also highlighted both internal and external outcomes. Desired internal outcomes were adaptability, flexibility, and an ethno relative view. Empathy toward others was also a key internal outcome.

External outcomes included behaving and communicating using one's intercultural knowledge to achieve one's goals. The requisite attitudes may be a relevant factor to consider in the measuring the outcomes of diversity training. Another facet of programming to consider in the evaluation process was the participants' experience prior to engaging in the diversity initiative. Hurtado, Engberg, Ponjuan, and Landreman (2002) developed a conceptual framework to understand how precollege experiences predisposed students to three different outcomes. They concluded that a participant's background characteristics, including his/her pre-intervention environment, pre-intervention engagement in diversity activities, and pre-interaction with people of color, should be considered when assessing the effects of diversity programming on college students.

Carol Weiss also offered insight into the importance of focusing on important aspects of a program and its functioning. She suggested:

> A theory-based evaluation sets out which events and linkages should be studied. The evaluator does not try to look at everything that goes on but concentrates on the phases of the program that have been defined as critical…leading to detailed understanding of what is going on at each phase of the program. (2002, p. 205)

Bickman and Reimer (2011) propounded that in social psychology, when a shift in attitude or behavior is the program's goal, one way to identify potential places for the shift to occur is when participants experience dissonance. In their example, they explained that when "the performance outcome is inconsistent with the self-standard, it creates dissonance arousal and in turn motivation to reduce dissonance" (p. 121). The authors suggested looking for the points that create dissonance with participants to see how the participants deal with it. This is similar to Pendry, Driscoll, and Field's (2007) assertion that social psychological theory and methods may help guide diversity training evaluation design. Tools such as the identity awareness or cultural competence self-assessments, for example, may aid in evaluating pre- and post-training attitudes, knowledge and behavior.

Diversity training has unique issues such as the political climate that make evaluating its outcomes challenging. Bunch (2007) underscores the importance of assessing organizational culture and training transfer as key elements in understanding the extent to which training affected its participants and their work environments. According to the literature, measuring training transfer and measuring the training transfer climate are two significant elements that few diversity training programs practice. Likewise, studies that measure diversity training outcomes concluded that training transfer may not persevere due to political influences, media, leadership, and education (Hogan & Mallott, 2005; Rosen & Saloman, 2011).

In this section, I reviewed scholarly articles on evaluating diversity training, concluding that although some research on evaluating diversity training has been

published, there is still much unknown about how to design an effective evaluation appropriate for diversity training. In the next section of this chapter, I present four existing models training evaluation design models that are not diversity training specific.

Kirkpatrick's Four-Level Training Evaluation Model

In 1959 Donald Kirkpatrick first published his four-level model. Kirkpatrick's model, and it persists as the most frequently cited model. Bober, Teja, Foxon, and Koszalka (2008) conducted a review of 57 journal articles from the training literature that referenced a training model. Of the studies with training evaluation models reviewed, they found 77 percent included Kirkpatrick's model (p. 5). They also recognized, however, that typically organizations referenced his model, but did not incorporate all four levels into their evaluation.

His first level, *reaction*, measured the training participant's feelings about the training. Kirkpatrick likened this to measuring customer satisfaction when his role was the training consultant and he needed to assess his customers' feelings about the programming he delivered. He argued that the training reaction should be "favorable" (Kirkpatrick, 2006, p. 21) for the trainer and training to continue, explicating, "It is important not only to get a reaction but to get a positive reaction…the future of the program depends on positive reaction" (p. 22). The notion of training as a favorable experience is counter to some scholarly research on diversity training content and delivery.

Kirkpatrick's second level, *learning,* encompassed the participant's knowledge, skills and attitudes developed during the training. He offered guidelines (1994) for evaluating learning. He suggested using a control group if possible. This may be difficult since often trainees are required to attend or encouraged to attend training based on deficiency of skills. Second, he recommends using a paper-and-pencil baseline measure prior to training and then comparing with a post-training measure to understand knowledge and attitudes as well as skill development.

Behavior was Kirkpatrick's third level. He warned against the trainer's tendency to measure only behavioral changes based on the training experience. Kirkpatrick's concern was that if no change in behavior was measured, the organization and training sponsors may believe the training had no effect on participants. Conversely, he suggested that knowledge and skill learning does not matter unless it translates to behavioral changes in the workplace. He also acknowledged that it is "extremely difficult to create a compelling chain of evidence leading from training to results" (Kirkpatrick, 2006, p. 83) without measuring behavior change.

He presented four conditions needed for behavior change to occur. First, he stated that the person must want to change his or her behavior. Second, a person must know what to do with the newly-learned behavior. The work climate must also be conducive to change. The last condition was that the person must be rewarded for demonstrating the changed behavior. Kirkpatrick further expounds on conditions within the workplace climate needed to allow for the change in behavior. He described five different climates based on the training participant's boss: preventing, discouraging, being neutral (neither

supporting nor hindering training transfer of behavior), encouraging, and requiring. Kirkpatrick suggested that trainers be aware of the climate prior to conducting the training.

Kirkpatrick modified his fourth level, *results,* after his original four-level model was introduced. In his most recent (2006) book, he included productivity, work performance quality, reduced costs, reduction in accidents, reduced attrition, and increased profits as measures of results from training. He offered six guidelines for evaluating training results. The language he uses to describe his guidelines illustrated the malleability of his four-level model. Similar to his recommended approach for measuring learning, first, he suggested using a control group if practical. Again, this may be a challenge since random assignment is not likely practical when training related to specific workplace skills and behavior needed (Weiss, 2002), and the attributes of a control group would have to closely mirror the participating group. Second, Kirkpatrick noted that time must elapse for the training participant to practice what she or he learned. The third guideline was to measure pre- and post-training to understand behavior and job performance changes. He also recommended repeated measures at different time intervals to understand results, especially long-term. His fifth guideline was to measure cost versus benefits, probing on the financial outcomes relating to employee productivity and other measures, versus the cost of training delivery. His sixth guideline was to "[b]e satisfied with evidence if proof is not possible" (p. 65). He explained that if practitioners attempted to measure results, but were unable to find any effect, they should review the training outcomes for levels one through three and identify other ways in which evidence

can be gathered. He did not address the possibility that training had no effect on participants.

Kirkpatrick stressed the importance of conducting evaluations at all four levels "or as many as you can" (p. 123) and conducting them in a "linear fashion" resulting in a chain of evidence to support the learning and behavior change resulting from the training. By conducting a cost-benefit analysis in level four, Kirkpatrick underscored that the four-level process will illustrate the effect of learning on the organization's bottom line. He warned against conducting a return on investment evaluation without the supporting levels one through three data; as other scholars also highlight, cost effectiveness cannot be used on its own to illustrate causation (Buckley & Caple, 2004; Lewin & McEwan, 2001).

Kirkpatrick's four-level model has been widely heralded as the pre-eminent model in training evaluation (Brinkerhoff, 2006; Clements & Jones, 2008; Fitzpatrick, Sanders, & Worthen, 2004; Mertens & Wilson, 2012; Russ-Eft & Preskill, 2009), however, few organizations conduct evaluations using all four levels (Russ-Eft, 2010). In their meta-analysis of behavior modeling training, Taylor, Russ-Eft, and Chan (2005) reviewed 117 studies and concluded that none of the training evaluation studies reviewed measured all four of Kirkpatrick's levels, despite the model's prevalence in the literature. Training evaluations tend to address levels one and two, but not levels three and four (Russ-Eft, 2010). Kraiger, Ford, and Salas (1994) emphasized the need to use a theory-based approach to training evaluation, referring to models such as Kirkpatrick's as "unidimensional" and "simplistic" (p. 312). Brown and Sturdivant Reed (2002) asserted

that Kirkpatrick's model lacked the attention to external factors that could influence the transfer of skills, knowledge, and behavior from training to the job. Kaufman, Keller, and Watkins (1995) criticized Kirkpatrick's model, suggesting that it is not widely used due to perceived complexity of levels two through four. Thus, despite their broad appeal and frequent reference, the four levels are rarely used in practice. Parry (2000) echoed Kaufman et al., stating that conducting a return on investment evaluation was difficult and many organizations cited that the benefits of training were soft and difficult to quantify in financial terms.

Hamblin's Five Levels

Hamblin (1974) developed a training evaluation framework to supplement Kirkpatrick's model, stating that his model is "not intended to take over from others" (p. 13), but to add a layer onto the existing four-level model. His fifth level examined the economic outcomes of training, including "ultimate value variables" or the "human good." His model recognized three conditions required to illustrate whether or not learning objectives were met. First, he stressed that trainees must have the aptitude to do the required tasks, to learn the needed knowledge, skills, and attitudes. Second, he recognized the importance of the work climate and measuring whether or not it is conducive for the learner to practice the new skills, attitudes, and behaviors. Third, the author asserted that trainees must react favorably to their training experience. In contrast to Kirkpatrick, Hamblin acknowledged that this does not mean they must like the training. A trainee's reactions to the training should be aligned with the learning objectives.

Phillips' Five Levels

Phillips and Stone (2002) asserted that organizations fund training at the expense of other organizational needs, and therefore the business impact and return on investment must be included in the evaluation. The authors advocated that trainers should become business partners. Training should be integrated into the organization's strategic plan and include a comprehensive measurement process, and relationships must be established with key stakeholders within the organization. Their first four training evaluation levels mirrored Kirkpatrick's. Phillips and Stone recognized that their model expanded on Kirkpatrick's and was not intended to replace it. Kirkpatrick's fourth level was measuring "results," which was very broad. Phillips and Stone narrowed the focus for their level four to business outcomes and measuring specific organizational outcomes. Their fifth level was measuring the return on investment for the training. The authors of this five-level model integrated stakeholder engagement in their evaluation process. They cited the necessity of asking clients or customers of the organization as a means to see if there is an organizational change as a result of the training

Phillips and Stone acknowledged that the majority of training evaluations used levels one and two, reaction and learning, but did not delve deeper into levels three, four, and five. They stated that levels four and five consumed the most organizational resources and therefore were less likely to be conducted. They suggested that training programs meeting one or more of the following criteria should utilize all five levels in their evaluation process. First, if the life cycle of the program was twelve to eighteen months, all five levels should be considered when designing the evaluation. Second, if the

program implemented was intended to address key organizational strategies or goals, the evaluation should adhere to the five levels. Third, when the program's cost was 20% or more of the training budget, the evaluation should be more in-depth than a program that was less costly. Finally, the fourth criterion was that management has expressed interest in the program and it is a highly visible training program.

Guskey's Five Levels

Thomas Guskey (2014a) drew on Kirkpatrick's four-level training evaluation model to develop his five-level professional learning evaluation framework. His first two levels were similar to Kirkpatrick's first two levels. The first level, participant reactions, addressed whether or not participants liked the training, whether or not they felt their time was well spent, or whether or not the participant felt the room and refreshments met their expectations. Measuring the participants' learning, Guskey's second level, delved into what skills and knowledge the training participants felt they gained during the session.

Guskey's third level differed from Kirkpatrick's model and the other training evaluation frameworks in that it probed into how the organization supported the training and whether or not the organization was prepared for change that might result from the training. In his research, Guskey recognized that organizational support and readiness for change were critical, otherwise "valuable improvement efforts fail miserably because of the lack of…support from…leaders" (Guskey, 2014b, p. 12) or that lack of support can "sabotage any professional development effort" (Guskey, 2002, p. 47). He cited another example of institutions that failed to provide resources for implementing change, and thus

the initiative failed. When measuring organizational support, Guskey suggested questioning whether or not sufficient resources were made available, if and how problems were addressed, and measuring the extent to which implementation was supported. It is noteworthy that Guskey's work is in educational settings versus corporate training programs, thus the notion of organizational support relates to diversity programming in higher education.

Guskey's fourth level, use of new knowledge and skills, is similar to Kirkpatrick's third level. However, Guskey explicitly recognized that this information cannot be gathered immediately following a training session (Guskey, 2002) but that enough time must transpire for training participants to have the chance to practice their learning.

His fifth and final level, Guskey suggested measuring student outcomes. This related specifically to his work context of teacher training programs. The author asserted that by understanding the ultimate impact on students of the teacher's professional development, the training evaluation would demonstrate that students were more confident or attendance was improving. However, with diversity training, it would be challenging to isolate the effects of a diversity program on the large student body within a college or university. A comparable analogy might be changes to the campus climate survey responses over time. However, with many diversity efforts and many other variables in a complex, dynamic institution, it may be unrealistic to measure and isolate the effects of one training program through the perspectives of college students.

In a 2012 article, Thomas Guskey asserted that evaluating training should always begin with the outcomes of the program. By keeping the program goals in mind when

developing and implementing the evaluation, success measures should drive the data gathering and analysis. In the same article, he also conceded that understanding the outcomes based on evidence gathered can be “challenging and a complicated task” (p. 41).

Brinkerhoff’s Storytelling Methodology

Robert Brinkerhoff (2006) developed his methodology to give organizations a protocol for illustrating the participant outcomes through storytelling. He acknowledges that training is only one of many variables that affects participants in potentially achieving their goals, thus a challenge to isolating the training outcomes versus external factors. Rather than conducting a quasi-experimental or experimental design evaluation to control for external factors, Brinkerhoff’s methodology is to gather stories from participants who articulate the how they were or were not able to apply their learning.

Brinkerhoff criticized models such as Kirkpatrick’s that proposed measuring training success by relying on a participant’s perception of the specific training as a valuable experience, recognizing that whether or not a training participant enjoyed the training has little bearing on transfer of skills, learning, and behavior in the workplace. Instead, his methodology focuses on how training participants would respond to a few brief questions about the content of the training, how, if at all, the content was learned and transferred to the workplace, and then, ultimately, what potential outcomes occurred in the workplace as a result. Brinkerhoff affirmed that by gathering qualitative responses from at least one participant, the evaluator or training practitioner would have “a credible and defensible answer to the impact question” (p. 9). This process of inquiry can yield

information-rich perspectives to illustrate the impact of the training without the concern for additional factors that may affect outcomes. Evaluation participants have the opportunity to describe how their learning affected their work performance, from their perspectives.

One of the strengths of this model was its attention to other factors that affect the training outcomes. Brinkerhoff acknowledged that "training alone is never the sole factor in bringing about improved performance, and is often not even the major contributor" (p. 12). Thus, this methodology was not intended to illustrate the sole causation for changes in workplace performance, as was implicit in other models.

Another difference between Brinkerhoff's approach and others is that he explicitly addressed systemic factors that enable or impede training transfer of skills, knowledge, and behavior. He stated that 80% or more of long-term training impact was determined by system factors such as transfer climate. Just 20% of the success was determined by the training program (p. 38). Training impact cannot be sustained without support from managers and other key players. Brinkerhoff mentioned the importance of acknowledging the team support versus focusing solely on the training, asserting that these approaches are "divisive" and "exacerbate the political isolation of the training function" (p. 41).

Brinkerhoff attended to the use of evaluation feedback, contrasting his methodology with other frameworks. He illuminated the fact that models such as Kirkpatrick's focus trainers as the primary intended users of the training feedback versus senior leadership and supervisors who serve to coach their staff. This is a significant

concern, particularly for diversity training since training transfer is affected not merely by the training or learned skills and behavior, but by the climate in which the employee works. Brinkerhoff also viewed training evaluation as an opportunity to build organizational learning competence. A four-level model will help trainers understand skills learned and behaviors transferred to the workplace perhaps. However, it will not highlight barriers or opportunities for improvement in the workplace. Stories articulated by staff may illustrate their challenges, thus offering leadership the opportunity to address issues, to change the climate, and to increase the likelihood of employees demonstrating the desired behaviors. Without such feedback, it would be unlikely an organization would change.

Brinkerhoff's methodology relied on participants' willingness to talk about the impact of their training on work performance. It is unknown whether or not diversity training participants would be willing to participate in a qualitative inquiry, particularly when the responses may relate to their work climate. When I searched for diversity training evaluations that employed Brinkerhoff's storytelling approach, the search did not yield results. I incorporated a question about using narratives in people's evaluation designs into this study, and the findings on this topic can be found in Chapter Four.

Conclusion

In this chapter, I reviewed diversity program evaluations and training evaluation models. I contrasted four approaches to evaluating training programs. None of the models addressed issues of longitudinal assessment that would measure training transfer of learned skills into the workplace over time. The existing models probe into how the

training ultimately affects the organization, but Kirkpatrick's model focuses on return on investment and how the skills benefit the bottom line of the company. The models did not incorporate systems concepts such as the fact that diversity training participations within a complex system are massively entangled or that systemic change occurs slowly and over time. Furthermore, the impacts on the higher education universities relating to diversity issues are harder to measure because training programs seek to shift the personal values of their participants. The nature of this training differs from other types of human resource and staff development types of training. The lack of explicit direction of how and where to measure impacts at the individual and institutional levels are significant flaws in the existing model designs.

Little is known about barriers for diversity training professionals. Since evaluation reports exist and are made public, there is evidence that evaluations are in fact conducted. The diversity practitioner reports presented in Chapter One, however, did not adhere to any of these training evaluation design models that I presented in Chapter Two. The scholarly articles presented earlier in this chapter also lacked reference to specific training models. This illuminates a gap in the literature and begs the question: Do diversity training practitioners utilize a framework to guide their evaluation design? If not, why not?

Chapter Three: Methodology

In this chapter, I outline the methodology and the methods I used for my study. I discuss research approaches and my reasoning for choosing grounded theory to develop a diversity certificate training evaluation framework that reflects current practices. Having identified gaps in scholarly literature and concerns regarding practitioner diversity program evaluations, the purpose of this study was to examine how current DCT program managers and evaluators evaluate their programs.

Creswell (2003) conceptualized three central questions regarding research design: what knowledge claims the researcher makes, including a theoretical perspective; what inquiry informs the procedures; and what data collection methods and analysis will be used. He stated that the three elements of inquiry combine to form different approaches to research (p. 5). A postpositivist approach seeks to "make claims and then refine or abandon them for other claims" (p. 7). It does not offer multiple realities or a socially constructed reality. A constructivist approach assumes that individuals endeavor to understand the world in which they live, relying on study participants' perspectives about the object studied (p. 8). Taking a qualitative approach as a constructionist is important for this study because, as Merriam (2009) suggested, institutions and their approaches to programming such as diversity work are fluid, not a fixed phenomenon to be measured. "One of the assumptions underlying qualitative research is that reality is holistic, multidimensional, and ever-changing; it is not a single, fixed, objective phenomenon waiting to be discovered, observed, and measured as in quantitative research" (Merriam, 2009, p. 213).

In this study, I endeavored to understand how diversity certificate training managers evaluated the individual and institutional impact on participants. Although training evaluation frameworks exist as addressed in Chapter Two (Guskey, 2014a; Hamblin, 1974; Kirkpatrick, 2006; Phillips & Stone, 2002), few studies have applied such a framework to diversity training evaluation. Therefore, the gap in scholarly knowledge about this subject would make testing a hypothesis of why such a condition exists or to develop of theory of what the barriers are prior to gathering data a challenge. Therefore, this work must be rooted in participants' practices, experiences, and views.

Grounded Theory

For this study, I used a grounded theory approach to develop a diversity certificate training evaluation framework. According to Charmaz (2014), grounded theory encompasses a qualitative, systematic, iterative inquiry. From this multi-stage inquiry process, the researcher will develop a theory or framework based on participants' perspectives of the subject at hand.

A Brief History of Grounded Theory

In the 1960s, Barney Glaser and Anselm Strauss collaborated on research projects that concluded with co-authoring *Awareness of Dying* (1965) and *Discovery of Grounded Theory* (1967). From their work emerged a new methodology for approaching research and, through an inductive process, developing an emergent theory about a phenomenon or problem. Through his teaching qualitative data analysis and grounded theory over time, Strauss' approach diverged from the original grounded theory approach that he and

Glaser developed in the 1960s. In 1987, Strauss authored *Qualitative Analysis for Social Scientists,* a book based on two decades of continued research and teaching.

Strauss partnered with Juliet Corbin to publish books to help students and scholars conduct rigorous qualitative research including grounded theory studies. Their 1990 publication, *Basics of Qualitative Research: Techniques and Procedures for Developing Grounded Theory*, caused a ripple among some grounded theorists. Barney Glaser published a response to Strauss and Corbin's (1990) book, suggesting that their book was a significant departure from Glaser and Strauss's (1967) original definition and intention of grounded theory. Glaser felt that a theory emerging from a study must be grounded solely in the data gathered, never forced or informed in part by previous conceived notion or scholarship on the topic. Glaser perceived Strauss and Corbin's approach as "forced." Yet Strauss and Corbin's 1997 publication reinforced that the book was a guide, not a "recipe book," but a fluid process, a "useful set of tools for analyzing qualitative data" (p. xi). Strauss and Corbin authored their books based on their experiences teaching students to conduct not only qualitative studies, but studies using a grounded theory methodology.

Glaser (1992) defined grounded theory as a "systematically and inductively arrived at [theory] through covariant ongoing collection and analysis of data. It has a fresh start, open to the emergent. One does not begin with preconceived ideas or extant theory and then force them on data for the purpose of verifying them" (p. 15). In contrast, Corbin and Strauss (1998) defined grounded theory as

> [t]heory derived from data, systematically gathered and analyzed through the research process. In this method, data collection, analysis and eventual theory stand in close relationship to one another. A researcher does not begin a project with a preconceived theory in mind (unless his or her purpose is to elaborate and

> extend existing theory). Rather, the researcher begins with an area of study and allows the theory to emerge from the data. Theory derived from data is more likely to resemble the "reality" than is theory derived by putting together a series of concepts based on experience or solely through speculation (how one thinks things ought to work.) Grounded theories, because they are drawn from data, are likely to offer insight, enhance understanding, and provide a meaningful guide to action. (p. 15)

The two definitions offered here are similar. Both definitions recognize the importance of developing theory through data collection and analysis, informed primarily by the study participants and observations versus preconceived notions. However, Strauss and Corbin allow more flexibility in the process than Glaser.

Glaser stipulated that researchers do not conduct a literature review prior to the study since it would influence their view on the phenomena or problem. Strauss and Corbin (1998) in contrast posited that "literature can be used as an analytic tool," suggesting that the researcher think in theoretical terms. The authors stated that "the literature can provide a rich source of events to stimulate thinking about properties and for asking conceptual questions. It can furnish initial ideas to be used for theoretical sampling" (p. 47).

For this study, I had already conducted a literature review. She returned to the literature periodically during the data analysis phase of the study as well, once she had identified emergent themes and sought to review relevant literature. Thus, this study did not adhere to Glaser's prescribed process relating to the literature review. Other scholars, including Strauss and Corbin (1998) and Charmaz (2006), allow a preliminary literature review since generally their approaches to grounded theory are less prescriptive and more as guidance for the researcher. This is a strong contrast with Glaser's (2005) perspective

on grounded theory as a methodology. “There is a need not to review any of the literature in the substantive area under study.” Furthermore, he continued to underscore the importance of not turning to scholarly or other types of information that might “contaminate, be constrained by, inhibit, stifle or otherwise impede the researcher’s effort to generate categories, their properties, and theoretical codes from the data” (Glaser, 2005, p. 31).

Strauss and Corbin (1998) underscored the importance of a process that systematically interrelates relationships that emerge during the study to ultimately develop a framework that explains relevant social, psychological, educational, or other phenomena. To accomplish this, the authors advocated looking at theoretical comparisons. This, they suggested, must go beyond analyzing the symbolic meaning of language and object to understand the phenomena. They advocated interpreting through similes and the contextual meaning in which the simile is given. The analysis in grounded theory cannot rely solely on what is stated and its literal interpretation; it must delve deeper into what Strauss and Corbin label property and dimensional levels. Rather than naming and classifying themes, grounded theory requires looking at multiple possible interpretations of what and how statements are conveyed to determine the meaning.

The notion of looking symbolically as well as at multiple interpretations from the perspective of informants was critical in this study. Since diversity training takes place in a political climate this study drilled deeply into statements about barriers to conducting multi-level evaluations, thus adhering to Charmaz’ (2014) as well as Strauss and Corbin’s (1998) dimensional level of interpretation and analysis.

Grounded theory as a methodology received scholarly criticism. In an article authored by grounded theorist, Antony Bryant (2009), the author contrasted Glaser with Strauss and Corbin. He asserted that the original grounded theory method (GTM) was "a naively inductivist method," but that it was "Charmaz' work in particular that initiated the development of a more profoundly justified base for the method" (p. 6). For this study I chose Charmaz's process because it was a constructivist approach. I was intentional about following the process according to this scholar because I did not want to mix approaches, thus potentially muddling the data gathering, analysis, and interpretation process.

The criticality of choosing one methodological grounded theory approach is emphasized by grounded theory scholars to avoid "skips and dips" between different grounded theory approaches that may result in theory "slurring" (Gynnild, 2011; Martin, 2011). Fendt and Sachs (2008) advocated that researchers review all approaches, but choose the method that is most appropriate for the scholar, depending on the researcher's epistemological and ontological stance.

Grounded Theory can be a method of analysis as well as the result of analysis (Mertens, 2009). The initial publication offered a guide for systematic qualitative inquiry that could generate theory and challenge the dominant positivistic paradigm. According to Charmaz (2014), grounded theory gained increasing acceptance in the 1990s, even from quantitative researchers who use a mixed-methods approach. Charmaz underscored the construction of original data analysis versus a "formulaic prescription" (p. 3). Thus,

the methodology offers a set of guidelines and strategies, but how the study is conducted is determined by the researcher.

Charmaz (2014) distinguished grounded theory from other research strategies, stating that:

> Grounded theorists: 1) Conduct data collection and analysis simultaneously in an iterative process; 2) Analyze actions and processes rather than themes and structure; 3) Use comparative methods; 4) Draw on data (e.g. narratives and descriptions) in service of developing new conceptual categories; 5) Develop inductive abstract analytic categories through systematic data analysis; 6) Emphasize theory construction rather than description or application of current theories; 7) Engage in theoretical sampling; 8) Search for variation in the studied categories or process; 9) Pursue developing a category rather than covering a specific empirical topic (Charmaz, 2014, p. 15).

Charmaz (2014) acknowledged that researchers frequently engage in the first five actions, “but do not make the remaining actions evident” (p. 15). Furthermore, she stressed the importance of using inductive data to develop “analytic categories” that was different from coding and sorting as was the case with other approaches using qualitative methods. Thomas (2006) defined inductive data as a process where the researcher or evaluator a) condensed raw data into a brief summary; b) established a relationship between the evaluation or research questions and the summary of the findings; and c) developed a framework of underlying structures based on the study informants’ experiences.

Another element of grounded theory is its iterative process. During the interview phase of this study, I followed Charmaz’s (2014) grounded theory process, writing memos after each data-gathering encounter to capture additional thoughts and understandings along the way. In the iterative process, I compared memos, transcripts,

and secondary source information that is another key element, employing constant comparison in grounded theory. Although little is published specifically about diversity certificate training evaluation, literature on elements of training evaluation exists. Thus, I continued reviewing literature on evaluating diversity training throughout this study.

For this study, informants were asked about their diversity training evaluation practices. The responses were condensed into summaries in the form of memos. Through analysis across memos and data gathered, I was able to develop the theory behind current evaluation processes. Following Charmaz's (2014) inductive, iterative process I was also able to identify gaps between scholarly literature on training evaluation and the actual practice from the ten informants in this study.

Sensitizing Concepts

Charmaz noted that sensitizing concepts spark initial ideas or concepts to pursue and may be used as tools to further define their data. "Sensitizing concepts may guide but do not command inquiry much less commandeer it" (p. 30). She posited that these concepts help the researcher start the inquiry; they do not conclude it. The literature review, for example, should be limited to information needed to develop the study, yet not influence its direction or outcomes. Grounded theorists must be open to what they see and sense during the research process. Grounded theory is not a process of confirming existing theory, but developing new concepts, frameworks, and theories. The grounded theory process starts with sensitizing concepts as a springboard, but researchers must focus on the data and observations during the iterative process. Charmaz asserted that sometimes sensitizing concepts or initial research interests were not addressed in the

qualitative data gathered, and those initial concepts might be disregarded. Grounded theory should focus on the data gathered versus the researcher's theory or preconceptions of what the inquiry may yield.

Memo Writing

Memo writing is a key component of Charmaz's (2014) approach to grounded theory. The memo might contain what the researcher senses while gathering data, including observations beyond interview responses. The memo-writing process also gives the researcher space to compare ideas along the discovery path, versus waiting until data collection is complete. The researcher should also note gaps in understanding to guide the iterative inquiry process. "Memo writing encourages you to stop, focus, take your codes and data apart, compare them, and define links between them. Stop and catch meanings and actions" (p. 164). In addition to memo-writing, Charmaz recommended keeping a reflexive journal to focus on the data and process versus preconceived notions the researcher may have about potential conclusions or theories. She also encouraged incorporating notes about methodology in memos.

Charmaz offered memo-writing guidelines, probing what people are doing, saying, and connections or comparisons the researcher made during the data gathering. She posed provocative questions: "What are people saying or trying to say? What do they remain silent about? What accounts for their silence? When, if at all, do secrets shape what's happening?" (p. 169). The guide served as a reflection to motivate memo writing. The memo-writing process is a significant part of grounded theory because it stimulates the reflective nature of this research approach and inspires the researcher to pause data

gathering for the purpose of comparison and contemplation. During this study I included reflexive questions in my memo-writing process to keep my preconceived ideas about diversity certificate training evaluation in check during the data gathering and analysis phases of this study.

Coding

In grounded theory, coding defines what the data are about and links data collected with theory development. Charmaz (2014) describes two steps in grounded theory coding. First she states that the initial phase "involves naming each word, line, or segment of data" and then the second phase is "a focused, selective phase that uses the most significant or frequent initial codes to sort, synthesize, integrate, and organize large amounts of data" (p. 113). She underscores the importance of remaining open to new directions emerging from the data. Reviewing data periodically is important in this approach as well. The data are not coded one time and then deemed complete, but revisited to see if the researcher's reading of the data changed based on new data gathered. To maintain the study participants' perspectives, Charmaz suggests coding gerunds versus coding topics. She states that focusing on participants' actions will encourage the theory development as a key heuristic device. Starting at the data collection phase, constant comparative methods are another significant piece in grounded theory. Charmaz also suggests comparing each interview with other interview data and making sequential comparisons. Holding back assumptions is important and considering the participant's view of his or her situation before making a judgment about their attitudes or actions is also critical.

For this study, I followed Karen Charmaz's constructivist grounded theory approach. I developed a memo-writing template and recorded my thoughts and feelings after each data gathering session. I also incorporated Charmaz's points about keeping myself in check and coding and recoding based on the constant comparison and iterative process.

Sampling

To select participants for this study, I employed Patton's (2002) purposeful sampling method to ensure that information-rich data are collected. In my initial search of institutions, I identified programs with a certificate-type of diversity training model. After researching three institutions with diversity certificate training programs, I returned to the literature to determine what has been studied about this training model. I found little published scholarly work specifically examining diversity certificate training in higher education. Thus, I determined I would focus on the diversity certificate training (DCT) model.

Chapter Four offers an extensive description and definition of diversity certificate training. However, initially I identified key tenets of the training that would affect the evaluation design. First, diversity certificate training typically occurs over several weeks or months. It is not a one-time training program. Second, it is a newly emerging model with little published about its impact. Third, I determined that since the model concludes with a certificate, the programs were more likely to have participants demonstrate what they learn versus a one-time diversity training session.

After narrowing my search to this specific model, I conducted a web search of institutions in higher education that have a diversity certificate training program. I made multiple attempts and searched using different terms including “diversity certificate training,” “diversity training program,” “diversity training,” “diversity programming,” “diversity program,” “diversity workshop,” and “diversity certificate.” My results yielded a total of fifteen institutions. Upon closer examination, I determined that four institutions were not appropriate to include in this study. One program was developed and implemented by the institution, but was exclusively for for-profit corporations’ employees. This study focuses on training stakeholders who work at higher educational institutions, so that institution was eliminated. A second institution’s program was solely online. This model of online delivery is different than workshops conducted in person that feature discussion and interaction as key elements of the program. Two other programs were geared for students who receive academic credit. This is a different model, different motivation, and is intended to yield different outcomes. Thus, the four institutions were eliminated.

To begin my research, I reviewed each of the eleven websites to understand the program and glean any information about how they evaluate the diversity certificate training program. I also searched each institution’s website to understand how they define diversity, how the institution’s leadership articulates its commitment to diversity, and additional demographic information about the institution. From this exploration, I then honed in on the institution’s diversity staff. I identified one or more key people at each institution who were responsible for the diversity certificate training. Initially, I emailed each of the eleven institution contacts and invited them to participate. Then I contacted

each of them by phone. After multiple attempts over the course of two months, nine institutions responded to my invitation to participate in this study. Two institutions never responded, despite broadening my inquiry to other people, including the chief diversity officer at each institution. Once the commitment was solidified, I sent each contact a consent form and scheduled the interview.

Data Collection

Interviews were conducted between March and June 2015. The duration of each interview ranged from 53 minutes to one hour and 22 minutes. There was no compensation for participating in the study, but I did agree to share a copy of the paper with participants. Immediately following each interview, I constructed a memo that contained the key points of the interview and also documented the intangibles from the interview such as points that were emotional during the interview. I noted questions I had and points to follow up when reviewing secondary data from the institution. Part of the memo process was reflexive to understand my emotional interaction and reactions to the study participants and their work at hand. The purpose of this step was to keep my sense of self or "I" (Peshkin, 1988) in check.

Throughout the first round of interviews, I employed constant comparison of existing and newly gathered data and reviewed memos to understand how different organizations approach their diversity training evaluation. I attended specifically to differences in their program and evaluation design as well as commonality among multiple organizations.

Data Analysis

Interviews were conducted by phone, and each interview was audio recorded. Once the interviews were completed, I transcribed the audio recording. I used QDA Miner, a qualitative analysis software to organize, store, code, and analyze my data. Each document was catalogued to ensure a systematic process and management of the data. Scholars recognized the importance of a data management process (Creswell, 2003; Merriam, 2009; Patton, 2002) and the impact a data management plan might have on the analysis phase of the study. I worked through the iterative, grounded theory process to develop a diversity training evaluation framework of current practices based on the information gathered for this study from informants and secondary data review.

Study Limitations

Diversity certificate programs are relatively new, developed within the last ten years. None of the participants in this study had participated in any other study of this nature. A key limitation in this study is that it reflects the perspectives and experiences of the informants at the nine participating institutions at the time the study was conducted in the spring of 2015. The study focused on the commonalities and differences in how they developed, conducted, and evaluated their DCT program. As explained in detail in Chapter Four, the institutional focus on gathering formative feedback, especially since some programs are in their first or second year, the evaluations are still designed to primarily provide formative feedback for continuous program improvement. Thus, they are constantly changing both the program itself as well as their evaluation approach.

The purpose of this study was to understand how diversity certificate program managers evaluate their training. An outcome from this study is a theoretical evaluation framework, building on Kirkpatrick's 1959 framework. This study will not establish how the framework might be implemented or what impact the framework might have on an actual evaluation design for diversity certificate training in higher education. This study does not delve into what is or is not effective in diversity certificate training either. It is a study about how we evaluate DCT programs, not an evaluation of the programming itself.

Credibility is a consideration in qualitative research. According to Patton (2002), credibility in qualitative research requires four elements: rigorous methods, disclosure of researcher experience, training, and bias (p. 552), therefore, I disclose my researcher bias later in this section. The final element is the philosophical belief in the value of qualitative inquiry, including holistic thinking and a "fundamental appreciation of naturalistic inquiry" (p. 553). Furthermore, Patton suggested attending to integrity in the analysis process by generating and assessing rival conclusions, looking for data that support alternative explanations (p. 553). Patton also advocated making the researcher's individual bias explicit (p. 555), which I attended to in my reflexive process throughout this study and state later in this section.

Creswell (2003) offered strategies "to check the accuracy of findings" (p. 196), including triangulation of different data sources to build themes; member-checking to bring themes or data analysis back to study participants to solicit feedback; presenting "negative or discrepant information" that is counter to the identified themes, which I

addressed in the member checking, and review of secondary sources including evaluation reports, instruments, and institutional websites.

I followed a systematic data gathering process as described previously, including conducting interviews based on a pre-developed, semi-structured interview guide according to Kvale and Brinkmann's (2009) suggested format that will include interview questions and probes. They posited that conducting semi-structured interviews offered the flexibility for the researcher to use her judgment to direct additional probing questions. Additionally, I piloted each instrument and incorporated feedback prior to submitting the instruments to the University of Minnesota's Institutional Review Board.

In his article, "Interviewing the Investigator" Chenail (2011) advocated that his technique of interviewing himself as a researcher can be useful in creating instruments and assess researcher biases. He noted the role of piloting the study as a means to identify researcher bias prior to conducting interviews. He stressed that during the interview process as well as the whole research study process, the researcher should "[i]dentify personal feelings" (p. 260) relating to the interview subject matter, the process, or the relationship to study participants.

Researcher Positionality

Before, during, and after conducting this study, I recognized that I have a bias regarding diversity certificate training. Since I have personally participated in diversity training in the past and I have also conducted an extensive literature review to prepare for this study, my belief is that diversity certificate training like the programs that

participated in this study do in fact have an impact on some participants. After concluding the data gathering, I believe this model may be very promising as a change agent within higher educational institutions. This may, however, be an unfounded bias since there is little to support my assertion that the programs are promising based on the current evaluation practices.

In this chapter, I outlined the methodology and methods for my study. I established how I used a grounded theory approach for this study, resulting in a framework to guide diversity certificate training evaluation design (see Chapter Five). I wrote memos after each data gathering encounter as additional evidence relating to the framework development. Mid-way through data collection, I revisited scholarly literature to compare my findings and framework development with existing frameworks and theories that guide evaluation. Additionally, I practiced reflexivity to understand my motivations, emotions, and sense of self affects the study as I progressed through the process. In the next chapter, I present the findings from the nine interviews and secondary source review for this study.

Chapter Four: Study Findings

The purpose of this chapter is to present the study's findings. In the first section of the chapter, there is a description of participating institutions, including the documented institutional value of diversity and support for programming. The second section of the chapter describes the diversity certificate training (DCT) programs. Included in this section is a table comparing DCT programs with general diversity programming, a description of program evolution and history, and a description of DCT components.

Information about the program staffs' skills and rationale for adopting this model of training is also described. The third section in this chapter presents a review of current DCT evaluation practices, concluding with the barriers study participants articulated about evaluating their programming.

About the Participating Institutions

Eleven institutions with diversity certificate training (DCT) programs were identified as potential participants in this study. Nine of the eleven participated. As of May 2015, because the programming existed in a limited number of institutions in the United States, the information about the institutions is reported in a way that will reduce the likelihood of revealing their identities. The informant names used in this study are pseudonyms and were randomly assigned by the researcher. Every effort has been made to protect the identity of both the participating individuals in the study as well as the institutions they represent.

Institutions with DCT programs were located in all regions across the United States. Of the nine participating institutions, one was in the western United States; three were in the South; three were in the North; and two were in the East, indicating that diversity certificate training programs were not a regional phenomenon. Table 1, the Institution Demographics Table, compares institutions by student and staff size as well as Carnegie classification of institution size. Seven of the nine participating institutions were doctoral universities, of which eight were labeled highest research activity. Diversity certificate training has been implemented in a variety of institutions, public as well as

private, small and large, with student populations ranging from fewer than 10,000 to more than 50,000.

Table 1.

Institution Demographics

Study Code	Informant	Employees	Student Pop.	Type	Carnegie Basic Type
DCT101	Adam	5-15	26-50	Public	Doctoral Universities: Highest Research Activity
DCT102	Beth	15 <	50 <	Public	Doctoral Universities: Highest Research Activity
DCT103	Carrie	5-15	50 <	Public	Doctoral Universities: Highest Research Activity
DCT104	Deanna & Elena	> 5	> 10k	Private	Master's Colleges & Universities
DCT105	James	> 5	26-50	Public	Doctoral Universities: Highest Research Activity
DCT106	Gabriela	> 5	26-50	Private	Doctoral Universities: Limited Research Activity
DCT107	Peter	5-15	26-50	Public	Doctoral Universities: Highest Research Activity
DCT108	Isabelle	> 5	10-25	Public	Doctoral Universities: Highest Research Activity
DCT109	Tonya	5-15	50 <	Public	Doctoral Universities: Highest Research Activity

Student population and number of employees reported in thousands. Faculty, staff, and student workers are included in the employee totals.

The Carnegie Foundation for the Advancement of Teaching classifies higher education institutions (carnegieclassifications.iu.edu) for the purpose of comparison in research. In this study, the researcher queried the “basic” classification using the Carnegie Classification web-based search to serve as a comparison across participating institutions in this study. The search was conducted in March 2015. The researcher also referenced student body population from this listing as a standard comparison.
Informant pseudonyms used in this study were assigned by the researcher to protect the identity of the study participants.

Institutions Demonstrate Support for Diversity

Regardless of institution size or demographic location, one critical issue for institutions conducting DCT programming is the need for ongoing support from institutional administrators. Research suggests that institutional support for programming through financial, policy, and vocal endorsement from key institutional leaders increases the stability and the impact of the programming (Guskey, 2014b; Hogan & Mallott, 2005; McGuire & Bagher, 2010; Rosen & Saloman, 2011). Therefore, at the onset of the study, websites for each participating institution were reviewed relating to how the institution addressed diversity. Speeches such as annual campus addresses or convocations, as well as institutional reports such as strategic plans, were included in this review.

Each participating institution's website contained statements in support of diversity initiatives and creating a climate of inclusion. Institutions included diversity in their strategic plans and described diversity as a core value. Leaders at each institution also either demonstrated support by participating in the training itself, by attending the concluding ceremony, or by making statements about the importance of diversity and inclusion on their campus. Table 2 compares diversity statements and leadership support based on interviews and the researcher's review of secondary documentation. Also referenced in the table is financial support for the programing. Each institution received funding for staff salaries and programming through their departmental budget. Since the administration in the institutions funded the programming through their annual budget allotment, they financially supported the diversity certificate training programs.

Table 2.
Institutional Commitment to Diversity and Inclusion

Institution	Diversity Commitment	Leadership Support	Financial Support
DCT101	"Diversity is a top priority across campus." (DCT101 website)	Chancellor highlighted "goals for inclusion" in the institutional goals. (DCT101 Website)	Included dept. budget (DCT101 Interview)
DCT102	"Enhancing competency within our community. Prioritizing the use of resources both to raise awareness of how others may perceive the campus climate and to take steps to improve it." (DCT102 Campus Climate Report)	"We must intentionally reject complacency about diversity and campus climate…to become a more welcoming, respectful and diverse place to learn, teach and work. This is a key element of the strategic plan." (DC102 Administrator's Speech)	Included dept. budget (DCT102 Interview)
DCT103	Institution that strives to "advocate for and educate about the university's goal of becoming more diverse and inclusive." (DCT103 Website)	President's stated goals included "Become more inclusive and diverse" (DCT103 Website).	Included dept. budget (DCT103 Interview)
DCT104	Institutional values of diversity, "embracing differences and uniqueness through sincerity, awareness, inclusion and sensitivity." (DCT104 Website)	"Focused on growing the international student population as a means of increasing diversity on an otherwise predominantly Euro-centric population." (DCT104, President's Speech).	Funded by dept. budget and by regional coalition. (DCT104 Interview)
DCT105	Institution's mission is to "improve the human condition…and to serve a diverse [student body]." (DCT105 website)	The Chancellor and the President speak at the every DCT reception. (DCT105 Interview)	Included dept. budget (DCT105 Interview)
DCT106	Diversity as "core value" for the institution (DCT106 website).	University president signs every certificate; president attends the certificate ceremonies. (DCT106 Interview)	Initially received special grant; now in dept. budget. (DCT106 Interview)
DCT107	"Foster understanding of and respect for cultural differences." (DCT107 website)	Support for diversity programming to leverage institutional change is "backed by the highest office in the University." (DCT107 Interview)	Included in dept. budget (DCT107 Interview)
DCT108	"Uncommon commitment to diversity" that "pervades all that we do." (DCT108 website)	Vice chancellor and other administrators participated; one high-level administrator planned to "mandate [the training] for [her] direct reports." (DCT108 Interview)	Included in dept. budget (DCT108 Interview)
DCT109	"Create an environment that respects and nurtures all members of the student, faculty, and staff community…value a diverse community." (DCT109 website)	"As your President, I hope I am always open to that which challenges me and makes me uncomfortable…as I respond with the core value of respect in all that I do…[Together] we can eliminate fear and lead with hope." (DCT109 President's Speech)	Included in dept. budget (DCT109 Interview)

In conclusion, each institution highlighted a commitment to diversity on its website and in other ways. As addressed in the literature review in this study, institutional and leadership support is critical for the success of a program. In sum, if an institution's administration does not demonstrate support for the programming or the programming does not receive ample funding it is less likely to have an impact (Guskey, 2014b; Hogan & Mallott, 2005; McGuire & Bagher, 2010; Rosen & Saloman, 2011).

Diversity Certificate Training Program Description

DCT programming is one of many types of initiatives that operationalizes institutional diversity goals and/or strategic plans. In the previous section there was a review of statements from university presidents, chancellors, and institution websites on the importance of diversity work as well as the value of a diverse campus community. The purpose of this section is to describe DCT program history, program goals, and program staff. There is also a comparison of attributes across programs as well as a comparison of DCT and general types of diversity training.

The DCT programs that participated in this study were developed between 2004 and 2015. The most recent program (DCT104) launched one month prior to the interview for this study in 2015 (see Table 3 later in this chapter for the specific year for each institution). Unlike other initiatives or training programs, the DCT model is relatively new, with an average implementation date of 2011. Three of the nine programs started between 2004 and 2010 (DCT101, DCT103, & DCT108); the remaining six started between 2012 and 2015.

DCT Programming in Contrast to Other Diversity Programming. Table 3 generalizes information with a comparison between general and diversity certificate training programs. The information used for the general diversity programming in Table 3 was developed based on the literature review of scholarly and practitioner accounts of diversity training efforts. The DCT information was gathered from informant feedback for this study.

DCT programming differed from general diversity program in specific ways (see Table 3). First, DCT programs had a series of interactions with participants with the incentive of receiving a certificate upon completion of the requirements. Second, DCT programs were directed at administrators, faculty, and staff and generally not at students, whereas other types of diversity programming were open to everyone within a university community. Although the program goals were similar for all types of diversity initiatives, the rationale of choosing the DCT model for this study was based on the repeated interactions required to achieve the certification. Implicit in the model was the recurring interaction with trainers and coaching that occurred, building relationships between diversity staff and program participants who would not only learn the information, but also transfer their learning and new skills to the workplace. Another attribute of the DCT model versus others was that when building a relationship with participants over the course of the programming, there was a longitudinal opportunity to evaluate the impact of the training.

One of the drawbacks of DCT programming versus general programming was that the workshops were at set times and required employees to leave the workplace to

participate. Three institutions offered DCT program components online in addition to the in-person workshops, but each of these institutions still required some face to face interaction to fulfill the certification requirements.

Table 3.
Diversity Programming Comparison

	Diversity Certificate Training Programs	General Diversity Educational Programs
Program Design	DCT programs have a required number of workshops to achieve the certificate.	Other educational programs do not have the certificate incentive.
Voluntary	Training participation is voluntary.	Some diversity training is mandatory, such as new hire training and is intended to avoid a lawsuit.
Participants	Participation in DCT programs in higher education is open to faculty and staff. Some programs also invite students and alumni to participate.	General diversity training is usually open to all university stakeholders, including students.
Training Topics	Topics included in DCT programs focus on differences around racial, gender, sexual orientation, religious, age, veteran status, ability, and socioeconomic status. Some programs cover identity awareness in relation to the "otherness," power and privilege, or intercultural sensitivity.	General diversity programs include similar topics to the DCT programs. However, since general programs do not have any attendance requirement, participants who attend general training programs may not receive exposure to a breadth of topics, relying on their own interests to guide their participation.
Program Requirements	Active participation in a prescribed number of workshops is required at every institution. Some programs require a portfolio, journaling, community service, capstone project, or written statement about the participant's experience in the program.	With other diversity educational programming, there are generally no requirements unless it is mandatory training.
Program Components	DCT participants complete a series of workshops over the course of several months. Other common components of the DCT programming include self-reflection, problem-based learning, the opportunity to practice using new skills in a safe environment, and coaching.	The program components vary based on the event. They may or may not include opportunities for participants to interact in a meaningful way, for example.
Program Goals	DCT goals include increasing knowledge about diversity; building communication and conflict management skills; and ultimately, transforming the institution toward inclusion, tolerance, and respect of differences.	The goals of general diversity programming are to increase awareness about diversity. Individual programs may have additional goals.
Necessary Conditions	DCT programs require a specific amount of time away from the workplace, therefore institutional support as well as high-level leadership support is paramount for the development, implementation, and outcomes of the program. Participants must be self-motivated and open to change.	General diversity programs may not impact the daily work schedule as much as DCT programming because participants have more flexibility to attend as their schedule permits.

DCT Program Evolution and Goals

Program staff developed the certificate training programs as a way to delve deeper into diversity programming beyond mandatory training programs that are required by law. Each study participant articulated their program goals in similar ways. The main purpose of DCT programming was to have repeated workshops intended to educate participants about diversity, build skills needed for working in a diverse institution, and offer tools and opportunities to practice what participants learned during the workshops.

Study participants described in depth what their goals were for the training program specifically relating to institutional change. Not only did Beth (DCT102) seek to increase knowledge about equity and diversity issues among participants, but she also sought to increase a willingness for participants to speak up and take action to "effect change on equity and diversity issues [and to increase] recognition that equity and diversity is a core part of people's work at the University…for institutional change." Other study participants also referenced the goal of DCT participants to deepen their understanding about diversity so they will take action and become agents of change (DCT105, DCT104, DCT107) and work more effectively together within the institution (DCT108) while engaging in more effective communication and building conflict management skills (DCT109).

Taking into consideration the key goals for the programming, program managers described the connection between their institution's goals and the DCT program evolution. Deanna (DCT104) stated that her department was "the unit that the institution looks to bring its [diversity] goals to life." Similarly, Gabriela (DCT106) described her

institution's proactive approach to educating its faculty and staff about diversity. She felt that the mandatory training her department did was "a negative thing" so they shifted "the perspective [from] a compliance model to proactive…to build a more multicultural, inclusive, informed organization." Peter (DCT107) stated that his program was developed in response to "the administration's call to action for departments to take action, not just 'broad strokes,' but demonstrate that they develop programs that lead to change." Likewise, Adam (DCT101) conveyed that the training program at his institution "evolved based on a need to address institution-wide benchmarks regarding diversity." He developed this type of training to serve as a concrete way in which their department addressed diversity within their institution. Thus, these examples illustrate that the program staff sought to align the DCT program's goals and outcomes to the larger institutional commitment to diversity.

Theories Incorporated into DCT Design and Delivery

During each interview, participants were asked about any theory or framework they used to design their program. The majority of the institutions reported they did not use a framework or theory as they developed their DCT workshop sessions or overall program. Staff drew on their own knowledge and experiences in adult education, student affairs, diversity programming, or human resources/training to develop the workshops and certificate program format. The four models referenced explicitly during the interviews included the following: the ally model (Beth, DCT102; Adam, DCT101); Bloom's cognitive affective learning theory (Carrie, DCT103); Kolb's experiential learning model, as well as Storti's cross cultural theory (Tonya, DCT109).

Although four study participants referenced theories they incorporated into their programming, they also mentioned the lack of existing resources available to guide them. Beth (DCT102) mentioned that when she first started to develop the diversity certificate training program, she looked for models, but found few she felt would help guide her. Gabriela echoed this concern stating:

> Theories? That's interesting! [Laughter] I'd have to say no. The politically correct thing would be to say yes, I looked at this, I looked at that, and there were models of what other people were doing, but for me it was my experience, and I thought about what I would need to know and how did I start off in this work? (Gabriela, DCT106)

Gabriela's sentiment about the lack of existing models was also apparent during the scholarly literature review prior to and during this study. The dearth of literature on developing and implementing diversity certificate training in higher education presented DCT staff the challenge to come up with their own materials, drawing on their own skills and knowledge.

Diversity Certification Training Program Components

Table 4 depicts elements of each program, including program requirements, the year the program started, and participation counts for each institution. Although the program requirements varied by institution, each program required participation in a set number of workshops, corresponding with a specific amount of time allocated for the programming. The numbers of participants varied widely due in part due to differences in the size of the institutions. Adam's (DCT101) institution was the first program, implemented in 2004, and Elena and Deanna's (DCT104) was the most recent program that was implemented the month prior to the study interview in 2015.

Table 4.
Diversity Certificate Training Program Attributes

Interview Code	Program Components	Requirements	Time Limit	Year started	Est Annual Participants	Total Since Inception	Certificates Awarded
DCT101	Workshops in-person; portfolio, personal impact statement; community service	20 hours; 5 required courses; minimum 2 hours community service; portfolio submission	None	2004	150	1,800	112
DCT102	Workshops in-person; dialogue circles; coaching	Basic certificate is 18 hours; advanced is an additional 12 hours; third level is dialogue circles (54 hours total)	None	2012	275	1500	300
DCT 103	Four online modules or in-person workshops; additional film series and other events; coaching	8 hours; 4 sessions; multiple types of certificates available	None	2007	4,500	10,000	6,000
DCT104	Program is part of a regional coalition; variety of programming	Participants are selected based on their interest and commitment to diversity.	None	2015	18	18	18
DCT105	Workshops in-person; coaching; pre and post self-assessment	11 seminars total; Six required seminars and five additional electives	None	2012	50	150	50
DCT106	Workshops in-person; reflection paper, capstone project	Five required two elective courses, reflection paper. Advanced cert. has five req., four elective courses, and Capstone Project and paper.	Complete w/in 18 months	2014	40	120	40
DCT107	Workshops	Completion of one core course and five electives	None	2012	50	1100	316
DCT108	Workshops; online and in-person	Seven core courses and two electives; 15 - 21 hours	Complete w/in two years	2010	204	648	219
DCT109	Workshops; online and in-person; written learning statement; capstone	10 courses for a total of 30 hours; plus 4 learning opportunities; written statement about what they learned	None	2013	35	100	40; 4 completed capstones

Unique Program Practices

During the course of the interviews, study participants described their programming. Across the nine programs, some facets of the programs emerged as unique attributes that may warrant further investigation.

Table 5.
Comparison of Unique Program Attributes

Institution	Unique Program Attribute
DCT101	Integrated an introduction to the DCT program into new hire training to encourage participation.
DCT102	Multiple tiers of training (introductory; advanced) followed by dialogue circles. The final tier is a train-the-trainer model that was in development at the time of the interview.
DCT105	After program completion, participants were labeled "cultural ambassadors" and invited to participate in ongoing learning opportunities or serve as volunteers to support campus issues.
DCT106; DCT109	Capstone project that gave participants an opportunity to integrate knowledge and strategies into a practical, problem-solving project. DCT participants presented their projects at the end of the program. Part of the programming was also to document what participants learned and how they felt during the programming in a journal.
DCT101	Community service project and a personal impact statement about what participants learned were compiled into a portfolio.
DCT105	Coaching as a necessary element of the programming; staff serve as coaches when challenges arise for participants outside of the workshops.
DCT103	Online and in person workshops; offered multiple types of certificates with different topical foci.

DCT Participants

All programs invite faculty and staff to participate. Some programs extended the invitation to students (DCT102, DCT103), and one institution even reached out to alumni to participate (DCT102). The goal of the program was to give skills and tools to enable action toward inclusion, therefore, most programs focused on faculty and staff primarily

since students are transient and alumni may not be connected to the campus beyond any longer.

Training Participants Have Diverse Starting Points and Participation is Voluntary

One challenge study participants recognized in this programming was that participants have varying levels of knowledge, experience, and comfort regarding diversity and racial justice at the onset of participation. Meeting people where they are and developing training that engages people in different ways was a consideration when developing the programming (DCT104, DCT102, DCT107, DCT106). One study participant described:

> …[It's] all about meeting people where they are. If you're talking over people, they won't get it. Then when you get people who are comfortable with diversity and they connect with others, then they go back to their department and they're not with them, they can see that there's some place on campus where they can talk to like-minded people (Deanna, DCT104).

Described in a different way, Peter (DCT107) developed programming where "anybody could come and feel comfortable enough to engage in one or more of those classes." Since the programs were voluntary and participation was diverse, DCT staff planned for different levels of comfort and knowledge around diversity issues.

About Diversity Certificate Training Staff

During the interviews conducted for this study, participants described their background, experience, and roles in developing the diversity certificate training program and evaluation at their institutions. Staff tenure at the institutions in the DCT program role or another role ranged from four years to more than thirty years. Table 5 depicts each

informant's length of employment in the field or at his or her specific institution and job focus.

Table 6.
Diversity Certificate Training Staff Job Focus

Interview Code	Informant Pseudonym	Staff Tenure	Job Focus Current (past)
DCT101	Adam	4 years	Training & Staff Development
DCT102	Beth	4 years	Diversity (Student Affairs)
DCT103	Carrie	25 years	Diversity (Psychology)
DCT104	Deanna	30 years	Int'l Education & Diversity
DCT104	Elena	4 years	Diversity
DCT105	James	20 years	Diversity (Student Affairs)
DCT106	Gabriela	28 years	Diversity
DCT107	Peter	6 years	Diversity (Int'l Development)
DCT108	Isabelle	1 ½ years	Diversity (Lawyer)
DCT109	Tonya	4 years	Training and Development

Two of the programs (Tonya, DCT109; Adam, DCT101) were developed and administered in the human resources department. An additional program (James, DCT105) was described as a "partnership" between human resources and the office of diversity and equity. Six programs were developed by the diversity department staff. In summary, the ten participants at nine institution in this study represented diverse backgrounds and different journeys to their positions as diversity certificate training managers. They each developed their programming by incorporating their own skills and knowledge, which are presented in detail later in this chapter.

Diversity Certificate Training Evaluation Practices

The first part of this chapter was to develop the context for DCT programs by including statements of institutional support, describing DCT evolution and goals, and documenting similar and contrasting attributes of DCT programming. The purpose of this section is to illustrate evaluation practices based on study participant descriptions of how they designed, conducted, and used their evaluation results. Consistently across all interviews, study participant responses focused on formative evaluation or gathering immediate feedback following each workshop to improve the program.

The first part of the section describes their evaluation efforts, followed by a discussion of four emergent themes from the qualitative analysis of the interviews and secondary sources: evaluation practice had a formative focus for continuous process improvement; participants relied on informal evidence to know the programs were affecting participants; study participants recognized a lack of evaluation beyond the exit survey; and study participants perceived they had barriers to conducting evaluations beyond a workshop survey.

Comparison and Contrast of Current DCT Evaluation Approaches

At eight of the nine of the institutions, program staff administered an exit survey at the end of each workshop to gather feedback. Since DCT104 had just implemented their programming at the time of the interview, they had not begun gathering any evaluation data. Deanna and Elena were planning to design their evaluation during the summer of 2015. Table 6 lists evaluation methods beyond the end of workshop survey. A common theme across the additional evaluation measures was that some institutions

received information from DCT participants, but did not systematically analyze the information. Therefore they did not include it in their annual analysis of the program outcomes.

Table 7.
Evaluation Practices

Institution	Evaluation Practice
DCT102, DCT103, DCT105, DCT107	Practitioners conducted a meta-analysis across workshops for a holistic view of their programming. However, the data gathered had a formative focus versus focusing on long-term program outcomes.
DCT101	The program required a number of participation hours. Practitioners reviewed documentation of the hours completed to verify that participants met the requirement.
DCT101	The program required participants to demonstrate a commitment within their department to diversity in their personal impact statements.
DCT105	The practitioners assessed the outcomes not only across workshops, but also followed a cohort through the program.
DCT102	The program practitioner sent certificate program participants a survey after completion of each workshop. Completion of the survey via email was required to receive credit for attending the workshop. This institution had nearly a 100% response rate as a result.
DCT102, DCT108	University human resource web-based programs managed attendance and workshop surveys.
DCT109	Practitioner followed Kirkpatrick's training evaluation model, but implemented only levels one and two.
DCT109	The department hired a graduate student to design and evaluate the program, thus they utilized available resources within the university.
DCT101, DCT106, DCT109	The program required completion of a portfolio or journal. The practitioner conducted a review of the portfolios. However, the review did not follow a rubric and focused primarily on completing the portfolio versus measuring the content of the portfolio relating to program outcomes.
DCT103	Integrated into the end of each workshop was an oral "debrief." However, this debrief was not recorded, nor were substantial notes taken as a means to incorporate the discussion into the evaluation. The practitioner mentioned that she incorporated formative feedback from the session debrief into her program planning.
DCT105	Practitioner administered a survey at the onset of program participation and at the end of the program. The survey contained a list of personal reflection questions about the individual's comfort level relating to diversity and was compared pre- and post- to see if there were changes.
DCT103	The practitioner administered a one-time survey three months after participants completed the certificate. However, the response rate was very low so she did not continue this practice.
DCT107	Practitioner mapped participation by department and by job level to highlight which departments had or had not participated. They also identified within each department the different job levels of participants.
DCT107	The practitioner probed reasons that some prospective participants were unable to participate due to departmental or systemic barriers.
DCT106	Practitioner conducted an informal systems analysis of who does and does not participate across departments, job levels.
DCT104	Program was implemented just a month prior to the interview. Evaluation data had not yet been collected; the evaluation had not been designed at the time of the interview.

Thus, the evaluation practices varied beyond the workshop survey. However, as discussed in the following section, the evaluation practices at each institution had limitations. The four emergent themes are addressed in the following section.

Formative Focus for Continuous Program Improvement

With the exception of DCT104 that had not yet implemented an evaluation, all eight other DCT programs administered a survey at the end of each workshop. The survey was perceived as a primary source of feedback for the programming. Study participants described how they used the data gathered from the surveys to make program improvements.

Several of the informants highlighted changes they made to their programs as a result of feedback from participants. Adam (DCT101) commented that his focus was "a little bit on the content, but it's mostly about how we can improve the program." He further explained that they evaluated everything they do, and they "get feedback from every participant. That's helpful and what we get back from the trainer we can go back and look at…what we can do to improve it the next time." They collected feedback via survey after each workshop. Isabelle (DCT108) and Adam (DCT101) both referenced assessing trainers through their evaluation feedback. For example, Adam referenced making a change with an instructor based on course feedback and his own observations during training sessions, noting, "We're able to assess and know that this instructor wasn't helping us accomplish our goals." Likewise, James (DCT105) described how formative feedback aided in a change to both the programming as well as a trainer,

stating, “I had an economics professor for a socioeconomic session, but it was good, but it wasn't human enough.”

Based on formative feedback, Beth (DCT102), James (DCT105) and Peter (DCT107) changed course content. Peter described the content change to be more “thought provoking” after receiving feedback that the course was too much like “Diversity 101.”

Participants Rely on Informal Evidence to Evaluate Their Programs

Using their own intuition or informal, anecdotal evidence to know the DCT programming was affecting participants emerged as a theme across interviews. Study participants perceived that their programming was having an effect on participants and, ultimately, on their institutions. They articulated concerns about how to measure the outcomes and the barriers they faced through a systematic process. However, they also expressed that they rely on their intuition as a means to feel their training is making a difference. For example, Deanna (DCT104) mentioned that she was relying on her intuition to know programming makes a difference. “We haven't even thought about evaluating it. I know it has to be done, but so far I'm just going by my gut…I haven't really thought about evaluation (Deanna, DCT104).

Both Carrie and Beth were emotional when they shared their experiences interacting with program graduates. They felt their program affected participants when they had encounters with current and past participants who shared their experiences with them. Carrie (DCT103) explained:

> There is something else, and I don't know how you capture this. When I walk across campus, I may have ten people that come over and give me a hug, and I know that sounds silly, but that means you've emotionally touched another person, a human being, and this happens day in and day out. So whatever's happening in the classroom is affecting people so much at an emotional level. When you have a faculty member who's teaching a course, you probably don't go over and hug them, but that happens with this work, and I don't know how to define it or describe it or measure it.

Likewise, Beth (DCT102) talked about a "file of unsolicited comments" that program participants send to show her how the training affected them. She also told a story of running into two women at Starbuck's who told her they had been through the training and "loved it." She said this happened to her frequently around campus. Another example Beth (DCT102) gave related to how the certificate affected a past participant and her ability to get a new job. "I'm seeing tangible ways that this is benefiting people. It's become a kind of currency. I know people who've gone outside the U[niversity] and used it when they're on job searches."

James (DCT105) remarked that he had "a lot of informal conversations" about the program and its outcomes for participants. He valued building long-term relationships with participations and maintaining an open dialogue when they needed him. Through this he felt that he received substantive feedback about the program as well, but he did not consider this part of the evaluation since it was informal.

Program participation as informal evidence that the program affects participants. Another way in which study participants relied on informal evidence was that their programs were at attendance capacity consistently. The fact that departments continued to send their employees was an indicator to Adam (DCT101) that the program was valued and successful. Isabelle (DCT108) also referenced a waiting list as an

indicator of the program's worth and reputation on campus. She said that she intuitively knew the program was working because they have waiting list for their courses, but also from the feedback she received:

> There's about 20 to 30 people on the waitlist for every class. We had registration on September 8, every single classes offered, there are 23 classes that were offered, every single one was full to capacity in one day. There's obviously a need for it [even] after five years. (Isabelle, DCT108)

Study participants said they "felt" the positive outcomes for participants, even if they were not able to demonstrate it through evaluation measures. Peter (DCT107) did not feel the need to systematically reach out to participants because of ongoing interactions he had with certificate recipients:

> I'll be honest with you. I don't have to [go back to participants over time] because participants never stop coming, which is incredible. You have six required courses, and I have people who've taken 14 courses. They never stop coming. They never stop providing their feedback.

As Deanna put it, "I know there's a change. I know that something positive is happening" (Deanna, DCT104). "For me the answer is absolutely I can see that this is having an organizational impact, but now we just have to figure out a way to capture it and determine what's going on" (Gabriela, DCT106).

The role of the DCT ceremony as informal evidence of program outcomes. Another informal way in which study participants intuitively felt their program affected participants was through testimonials offered during the certificate awards ceremony at the end of the program. All nine of the programs had a ceremony to recognize the achievements of the DCT participants. James (DCT105) described how a graduate expressed her program experience:

> There was one woman in particular [who expressed what she learned from] this program [was] how to identify micro aggressions, and one of the aha moments she had is that she realized she's lived and worked in an environment full of micro aggressions, and she's just tolerated that the last few years, but now she's having to make choices as to which one she lets go, which one she will try to confront, and how she's going to confront it, and that's one of my goals is that we very much apply the concepts to our day-to-day work. That's our latest example, but that's exactly what happens all the way through.

Tonya (DCT109) also had DCT participants tell the "story of their journey" during every certificate ceremony, which she described as "very touching." Gabriela's (DCT106) institution had training participants put ribbons on a tree. The tree full of ribbons symbolized for them the transformation within their employees and their institution. She did not consider this exercise part of their evaluation, but found it nonetheless very meaningful.

Coaching as a means for informal feedback. Three study participants (DCT102, DCT103, DCT105) referenced how ongoing interactions with program participants gave them a sense of how the program affected participants. James (DCT105) articulated his ongoing role as a coach for current and former program participants. He mentioned that it was one way he felt he knew the program was making a difference for participants:

> Yeah, I do a lot of informal conversations. One of the great things about being the program administrator or manager and I try to get other experts to facilitate is that I can have a relationship with the participants…We can talk honestly about what's happening with the sessions because I'm not the facilitator so they can talk with me or about the experience that they're having…I go to their office and sit down with them and talk about their learning objectives. It's great 'cause it gives me a personal opportunity to make a connection with them.

James (DCT105) offered an example of coaching a specific employee in her time of need when the participant's son had just come out as gay to her. He felt this demonstrated not only his relationship with participants, but also what the woman had

learned during the training because she was critically thinking about the situation. James commented that the training was “very much work-focused, but there [were] times that it [became] part of people's personal life as well.” James gave the example to show why he felt the programming was affecting participants in multiple ways.

Other study participants also discussed how their relationships and interactions outside of the program helped them understand how their program affected participants. Carrie (DCT103) mentioned that she worked with participants on an ongoing basis, and, as a result, she instinctively felt the program had a powerful effect on participants. She said, “I don't know how to define it or describe it or measure it, but I know it happens, and people will call you and ask your advice on things so something changes.” In sum, coaching and ongoing interactions sometimes served as proxy for other evaluation measures as a means of demonstrating to the DCT practitioner that the program was affecting participants. Informants focused on the helpful interactions, productive coaching and examples that demonstrated the training or coaching affected participants in a transformative way. They did not share examples of negative outcomes or interaction with participants.

Study Participants Acknowledged their Lack of Evaluation beyond the Exit Survey

Participants in this study readily acknowledged their lack of evaluation for their DCT programs beyond a survey at the end of each workshop. Some laughed when they were asked about how they evaluated their programs (Carrie, DCT103; Deanna, DCT104; Isabelle, DCT108 & Tonya, DCT109), and others expressed frustration with their lack of ability to measure DCT program outcomes (Beth, DCT102; Carrie, DCT103;

James, DCT105). They recognized that relying on surveys at the end of the workshop limited the amount of information they received to understand whether or not there were any potential program outcomes for participants, whether positive or negative.

For example, James (DCT105) felt that the lack of evaluative evidence about whether or not the training made a difference could be frustrating, stating that “[a]fter many years of feeling like I was working really hard and many hours, I was making very little progress. Like three steps forward and five steps back. I was so frustrated. I need something to show that I’m making a difference.” Carrie (DCT103) wondered why no one at her institution was asking for more data to support the program outcomes.

Program managers received funding through their departmental budget for the DCT programs with the exception of DCT104 that also received funding through a regional collaboration that supported the programming. Since the funding was part of the larger budget, study participants described how they reported evaluation results to their administration with a focus on the number of participant and certificates awarded (DCT101, DCT102, DCT103, DCT105). When asked about how they reported program outcomes to their administration, James (DCT105) commented that “what's in [his] head [was] data and numbers take precedent.” Adam (DCT101) similarly stated that the administration at his institution was only “interested in how many people attend they've asked for attendance numbers, and we provide that every year.”

In conclusion, DCT program practitioners recognized their lack of evaluation methods and underscored their interest in doing more to evaluate their programming. However, they also recognized and readily contributed their perceived barriers to

designing and implementing more evaluative measures during the study interviews. The next section addresses their perceived barriers.

Perceived Barriers to Evaluation beyond Exit Surveys

When asked what potential barriers study participants perceived that inhibited their diversity certificate training program evaluation, the responses ranged from lack of time or staff to conduct the evaluation to the politics and fear around evaluating this particular type of programming.

Study participants readily referenced a lack of staff (Adam, DCT101; Beth, DCT102; Carrie, DCT103; and Tonya, DCT109), a lack of time (Adam DCT101, Beth DCT102, and Gabriela DCT106), which they felt also related to not having staff dedicated to conduct an evaluation of this programming, or lack of resources (Adam DCT101). Another example supported this point. "I'm basically the only trainer these days, so it's hard to gather that [evaluation] information. I've done 17 workshops for the last two weeks, and the workshops are two to three hours long" (Carrie DCT103). Beth (DCT102) also felt that diversity staff were "really committed" and tended to "bite off more than they could chew because the need is great and the demand is great" with regard to developing and conducting their training and evaluating it. Adam said, "It's kind of keeping the certificate program going versus expanding to do all the evaluation we want to do."

> If it's between doing more programming, more workshops, [or] evaluation…I want to get out there. I want to get the work done…but one of the barriers is that I think the majority of the people who do this work don't come to it from an evaluation lens (Beth, DCT102).

Like Beth (DCT102) and Carrie (DCT103), Gabriela (DCT106) commented that she had been too busy building the program to focus on evaluation. She was on leave from her university for six months so that compounded the issue. She mentioned, “We’re growing so quickly. It’s just organic, and we have to catch up with ourselves.”

Deanna (DCT104) mentioned that time within the scope of their programming was an issue. She felt they could not dedicate time for evaluation during the program itself for program participants to give feedback.

> I would say [evaluation] issues going forward, one commodity we have much less than money, we don't have is time and having people find time to learn and attend and meet to talk about what they've learned and to create the meetings or the spaces in their own units. That's a huge obstacle…Time is the commodity. (Deanna, DCT104)

Informants stated that a lack of tools (Adam, DCT101; Isabelle, DCT108) available was a barrier, and others perceived that the outcomes of this type of training were hard to measure (Carrie, DCT103; Gabriela, DCT106; Tonya, DCT109). Deanna and Isabelle felt a lack of skills to effectively evaluate DCT was a barrier, as Isabelle (DCT108) described:

> It's definitely a lack of resources. I know basic tools, like doing a pre-and post-evaluation and how to do an assessment like a metric. I would want to pull someone [who] has more of a data background than my background to get something that's like proven results, but the time to do that and the lack of resources of people who have their skill set is not currently feasible at our institution (Isabelle, DCT108)

One study participant conveyed that she was unsure how to evaluate the outcomes of her DCT program. “I have to tell you it's kind of tough to evaluate, at least it has been for us to really know how people are using this information” (Carrie, DCT103).

Tonya was also unsure how to measure the impact of the training, particularly the impact on the institution. She referenced using Kirkpatrick's four-level evaluation approach when designing their evaluation, but was unsure how to incorporate his fourth level (results) and particularly whether or not training affected the institutional culture, policies, or procedures:

> Well, it's like any training, [participants] learned this, but looking at that, at the organizational impact, sometimes it can be very difficult to measure. You can say that people say they've changed, but actually putting a number to it can be a lot more difficult (Tonya, DCT109).

Fear that the evaluation would not show an impact and the possibility of the program funding being cut were concerns that both Beth (DCT102) and Gabriela (DCT106) referenced. The fact that diversity issues and shifting the culture within the institution were highly emotional and political was an issue as well. Gabriela (DCT106) articulated her concern:

> I think that's an important piece of diversity [that] people grapple with. People are reluctant to define measures of success, and I don't know if it's a fear of losing money, ROI, how to show a return on investment. I think just the terms and diversity work seems like, how do you put a price tag on social justice, equity, moral issues of representation? For me that is the reluctance for diversity practitioners that say we need to measure success. What's the consequence if we can't measure success? Does the program go away? (Gabriela, DCT106)

Gabriela continued the conversation, stating that she was concerned about how to productively measure diversity programming since it is so complex:

> This is a big challenge in our work. For me that's the danger. Once you find a finding, once you see it, see a problem or challenge, then you're bound to address it. I am constantly saying, "How do we measure?"" How do we know?" . . . What are measures of success? (Gabriela, DCT106).

Carrie (DCT103) was concerned that no one in her institution was asking for more accountability in her program outcomes. She intuitively knew the program made a difference, but also commented:

> If you were to say to me, “Where are we at today eight years after I've been here? Where [are] we versus where we were before?” I'd say we've really upped the game in terms of awareness, and awareness is about the lowest level. Awareness and appreciation. I would see if there have been some behavioral changes, too. There maybe have been [some changes], but I have no data that supports that, and that's really sad. It's amazing that nobody says to me, “To continue to do your work, we need to see your outcomes, something to support your outcomes.” (Carrie, DCT103)

James (DCT105) expressed a different fear about incorporating results from their training program into their department’s annual report. Since his institution conducted a pre- and a post-self-assessment for each participant and the participants received a numeric rating, he feared that sharing results, particularly when the cohort was small, might breach confidentiality.

> . . . [W]e do an annual report for the division, and we’re looking for the appropriate way to incorporate this into the annual report. We want to do it in a way that focuses on the positive as opposed to zero out the any particular cohort. The cohorts are pretty small so it tends to get rid of the anonymity of it. We don't want anyone to feel like we’re calling you out…like saying you didn't know much about diversity when you started out. We could say that to the cohort and have a conversation with the cohort, but I'd hesitate to have that in a public domain document because then there [are] going to be people reading it saying, “One of my coworkers went to that cohort, what areas of diversity did she not understand?” I try to respect people's individuality and confidentiality within this. (James, DCT105)

Training Transfer: Measuring Behavior Change over Time

Chapter Two addressed training transfer theory. This is a critical element in evaluating training outcomes because it is the process of transferring learning and skills

into behavior applied in the workplace. If a program is to have meaningful outcomes, this element of the training impact must be measured over time, allowing training participants to practice and apply what they learned (Bunch, 2007; Hogan & Mallott, 2005; Rosen & Saloman, 2011). Measuring training transfer cannot be satisfied with an exit-type survey immediately following a training workshop because participants have not had the opportunity to apply their learning.

There was one study participant who recognized the importance of measuring training transfer. When asked if they go back to training participants over time to see if or how they are able to use their newly developed skills, none of the nine institutions' representatives said they were able to measure long-term impact effectively. Tonya referenced information collected that helped them understand how participants have incorporated the training into their work, but did not offer details.

Carrie worked with her institution's assessment department and developed a survey. She felt that the survey "would have been highly successful if we could just get [former DCT participants] to understand the importance of that. Several times we tried [asking] 'please take five minutes to help us out.'" They only tried to conduct a survey to all former participants once and had a low response rate.

Lack of Skills

Beth was heartfelt in her response when asked about evaluating their training program. She explained, "Let me be totally honest. That's never been my strongest suit. My strongest suit is educating, designing the programming, and I'm pushing myself to be

more attentive to evaluation." Gabriela (DCT106) was animated in her expression of concern about measuring "deep-seated" values:

> What I struggled with…if [training participants] come in voluntarily, they choose to be part of it, and it's from that positive frame that they're more open to learning. So then it's switching the paradigm--"this will benefit me, this will benefit my organization, beyond my work it will benefit my community, my church, my kids." We all need to be more diversity competent and understand our own assumptions and how we navigate the world. And this gets to again how we measure it, and people are really reluctant to measure those deep-seated values (Beth, DCT102).

Similarly, Carrie (DCT103) stated:

> I think there's a belief that what you do with diversity and inclusion kinds of education is deep down in the individual, and it's deeply personal, and things happen, and it can be tough to pull that out. People change, but they don't know how to share that with you. These are the kind of things you get when you talk about evaluating this. If we say learning occurs, that's permanent change in behavior to me, but I don't really know that, as honest as I can be, they give me some indication. I get somebody who sends me an email, and they tell me about what they learned in class and a story about what happened and an application of what they learned in your class, and I say those are hallelujahs for me, but I can't really say word systematic about documenting that.

Peter (DCT107) shared that he felt the lack of tools was a challenge for evaluating their programming. He explained, "I honestly think that one of our weaker areas…we're challenged by that we can certainly develop more, is more tools to get feedback outside of the anecdotal stuff." Peter (DCT107) described a course-by-course review, with a formative focus. He felt that they needed to conduct another full program evaluation versus their primary focus that had been evaluating each session.

> Once, kind of in-house about a year and a half ago we asked everybody about the entire program. We got some really good feedback, and some of the new courses we offer, some of the new thematic areas came out of that. But we will be looking at soon is seeing if we can go back to another one of those overall general program evaluations. One of our shortfalls now is that we evaluate on a course by

> course basis, and in a little over a year and a half we haven't done a full program evaluation. (Peter, DCT107)

Deanna (DCT104) stated they “didn’t have an evaluation of the program; it’s really just in the beginning [of the program]. I don’t have an evaluation.” But she and her colleague, Elena, felt that they would evaluate the program “[m]aybe in eight months. I look at the people who were inducted and ask them, ‘Do you feel like you’re getting enough? What else can we give you?’” (Deanna, DCT104). Isabelle (DCT108) expressed regret that they currently did not “go back and check to see whether or not they’re able to use their skills in the workplace.”

Study participants expressed interest in doing more to identify program outcomes. Across all nine interviews, informants commented that they were interested in doing more to evaluate their training programs. Gabriela (DCT106) commented that it was important to delve deeper into the outcomes of their DCT to “demonstrate that progress and added value to the organization, to be able to say ‘this is the impact’ would be a tremendous thing.”

Peter (DCT107) admitted that they needed more tools, but he was very interested to learn how to do more as well, although his comment focused on program improvement:

> I spoke to my chief diversity officer. We’re really excited about learning what else is out there and how we can learn and how we can grow and how we can modify…We are at a place of learning that we don't believe [we’re] perfect. We believe there [are] things that were missing so please include us in whatever communication we can get out of the process.

As referenced in the section on informal evidence, study participants intuitively felt something was changing. Carrie (DCT103) expressed interest in increasing her institution's ability to measure the change:

> Some people walk out the door and I think there's some kind of thought process. What I tell people, my job is to have an emotional shift. It's about a sharing of perspectives. I wish I knew when people walked out the door what was the perspective that challenged you, but it's hard to do on the spot because people are processing what's going on. I don't know if there's a way to do a follow-up, but just what was the point of view that triggered, made you think if that happens, that's learning to me… It would be great to say, 'Hey, we can evaluate this. Here's a way to do this.' It's unbelievable that none of us know how. We just keep doing it. (Carrie, DCT103)

James (DCT105) commented that he was "frustrated" and that he desired some evidence that would show the program was making an impact on participants. He expounded on his interest and evaluation needs versus reporting the numbers that the administration required:

> You know that something I haven't included in the assessment piece just because when those reports filter upward from the chief diversity officer and upper administration, they tend to put a lot of weight on the percentages and the numbers in the raw data, and they're not so keen on 'tell me a personal story of how this impacted a person's life.' It's like, I want to know what people are learning and what this translates for students and staff and retaining talent. I'd like to find a way to incorporate that. . . That's really where the core of my heart is, but what's in my head is data, and numbers take precedent. (James, DCT105)

Lightheartedly, Isabelle (DCT108) mentioned during her interview, "I want to say, we currently don't do [multi-level evaluation]. I think we should. I felt something in your questions, if you have any suggestions [laughing], any recommendations?" During the interview she commented that she took notes based on the questions asked. Gabriela also felt the interview for this study spurred her into thinking more about how they might evaluate her program.

> You have me thinking about what I need to do to move forward. I took a few notes on some of the comments [about] the logic model. That's given me a few ideas to move forward to addressing it this summer. Then again, how do we show our value? I'm not afraid of the conversation, and if it means that this program is eliminated, that means there are resources I could put into something else. I think evaluation is very needed. Many of the people came out of this work organically by doing it, by living it, by experiencing it, and so because it's experiential that we do need the tools…I think people are uncomfortable with those hard measures still because diversity has moved beyond, beyond to inclusion. We're struggling with how to measure it. (Gabriela, DCT106)

Isabelle (DCT108) described her interest in doing more, especially relating to the training transfer from the workshop to "actualizing" learning into the workplace:

> Off the top of my head I would want to know how it is actualizing into their daily working environments . . . Let's say they complete the program, and I would want another assessment done six months or a year later. How have these concepts actually been integrated into daily working? How [do] you interact? How do you include in work with students based on these concepts that you're learning? (Isabelle, DCT108).

Chapter Four of this study has presented the study findings based on interviews conducted with DCT staff at nine institutions around the United States. The chapter included descriptions of the intuitional and leadership support for diversity based on secondary source review, descriptions of the DCT program evolution, goals and staff who developed it, and the evaluation practices articulated by the study participants. Within the evaluation practices, four emergent themes were addressed, including a focus on formative feedback, reliance on informal or intuition to know the program is affecting participants, lack of evaluative measures to support program outcomes, and perceived barriers to conducting a more in-depth evaluation. Despite barriers and concerns, study participants expressed interest in learning how to do more to evaluate their programming. Two of the participants stated they took notes during the interview for this study based on the questions asked.

Chapter 5: Discussion

In this concluding chapter of the study, a brief summary and a discussion of the findings are presented. An emergent framework was developed based on the information gathered for this study. Suggested evaluation design and approaches are included in this chapter as well.

Table 8.
Study Findings

Question	Findings
Research Question 1: What is diversity certificate training (DCT)? How is this model different from other training?	Diversity certificate programs in higher education have multiple workshops on topics related to diversity. The programs are intended to give participants knowledge, skills and opportunities to practice newly learned skills. Participants attend a specific number of workshops to fulfill the program requirements. DCT programs are voluntary.
Research Question 2: How do practitioners evaluate their DCT programs?	The primary evaluation focus was formative for program improvement. The program staff administered a survey at the end of each workshop. Study informants described using their intuition or informal feedback as ways of knowing their program was making a difference.
Research Question 3: What barriers inhibit multi-level DCT evaluation design?	Study participants articulated a lack of staff, skills, or available tools as barriers. The political nature of DCT programming was also a concern regarding evaluating and sharing the results. Study participants felt that the outcomes of this type of programming were hard to measure.
Research Question 4: What do practitioners perceive as components in a thorough DCT evaluation?	DCT program practitioners underscored a deep interest in doing more to evaluate their programs. They were unsure or unable to address how they would design and conduct a different evaluation. Some referenced methods such as conducting focus groups or incorporating a pre- and post-assessment.

Discussion of the Findings

Institutions emphasized their valuing of diversity and their deep need to address diversity issues on their campuses. This crystalized during the secondary source review of websites and speeches at each of the nine universities that participated in this study.

The fact that institutions and their executive administration supported diversity as a core value was important for the diversity program development and ongoing financial support. It was also important for participants to buy into the program and feel it was worth their time.

However, the main reason this theme was highlighted throughout the study was that the administration's support for diversity within the institution directly connected to the DCT program's purpose. Diversity staff developed DCT training to teach faculty and staff the skills and knowledge to help them transform personally, but also to ultimately transform their institution's culture. If the intention of the program was to have an institutional impact, then the resulting evaluation design must explicitly recognize that the training was not about individual outcomes, but also about how individuals and groups of DCT participants effected change within their institutions. Moreover, it necessitates that the evaluation design framework recognize that the training involves groups of participants from across the university. As participants moved through the workshops, the programs' intentions were that DCT participants would practice their new skills and transfer their learning into the workplace. If the commitment to institutional reform is sincere, this must be effectively measured and measured over time.

DCT programs have the unique attribute that the series of workshops take place over time and conclude with a certificate. General types of diversity programming do not necessarily have these two components. Longitudinal analysis could be built into the evaluation if program managers began looking at the individual participants versus the aggregated results from one workshop as the unit of analysis. They could track each DCT

participant through the program to measure any change, positive or negative. Only one informant in this study described administering a reflective pre- and post-self-assessment.

Some informants (DCT101, DCT106, DCT109) described a requirement for DCT participants to keep a journal or portfolio or to submit an impact statement to fulfill the requirements for the program. The informants stated that they conducted cursory reviews to check if the assignment was completed, but did not incorporate content review into their evaluation process. To consider the journals part of the evaluation, they would need to have a systematic review of the content since "[p]rogram evaluation is the systematic collection of information" (Patton, 2002, p. 10) versus their practices of a cursory review to determine whether or not there was enough substance in the journal to fulfill DCT requirements.

Study participants articulated and reinforced that their evaluation focus was primarily formative versus summative or relating to accountability. This was not surprising since some programs were implemented within just months before the interview for this study. Program staff were also focused on developing and conducting the training versus evaluating the training's outcomes, therefore reinforcing the practical need for formative evaluation versus attempting other types of evaluation.

One striking aspect in the findings of this study was the issue of accountability and the program's ability to demonstrate that it was meeting its goals of individual and systemic change. Study informants felt the program was part of their institution's goals and funded by their institutions. However, there was a gap between evaluating how, if at all, the institution was affected by individuals or groups of DCT participants. The focus

of accountability was on reporting the participant numbers. Therefore, according to the informants, as long as they were able to report the numbers of participants and immediate learning outcomes, those two measures were sufficient for their institutional accountability.

Informants offered stories and descriptions about how they rely on informal evidence to support their feeling that the program makes a difference. Informants drew on their own intuition to know their program was having the desired outcome. Their comments illustrated a concern about accountability. However, as described in Chapter Four, key barriers to conducting evaluations were the lack of evaluation skills and knowledge as well as a lack of available tools.

Building on the individual outcomes and evaluating training transfer, an evaluation framework might recognize that DCT participants across the institution work either together, perhaps as a cohort, or separately, but simultaneously to create slow, incremental change across the institution. This significant factor of individuals trained in groups and then returned to their individual workplaces was not reflected in any of the training evaluation models I reviewed in Chapter Two. By incorporating tenants of systems theory, including the concept of staff within complex institutions as massively entangled (Eoyang, 2006, p. 128) and the notion that change resulting from DCT programming may be very slow and incremental, the resulting evaluation design should recognize the dynamic complexity of universities and potential spillover effects beyond the institution.

Of the ten informants in this study, only one had developed a systematic review of who did and did not participate in the DCT programming. This might be another missed opportunity since institutions clearly stated their commitment to diversity. Programs were intended to effect change in the institution's culture of inclusion and equity. Based on the one practitioner example discussed in chapter four, the systematic review of participants was designed to help them understand who needed to be encouraged to attend the training versus explicitly determining who was participating and what effect if any, participation had across the institution. This example was the only evidence of an effort to look at the system across all informants in this study. This process of mapping participants could be expanded to include evaluative measures by mapping outcomes of aggregated individuals or teams of DCT participants concentrated in one area within the institution to measure what, if any, outcomes resulted from the training.

Study participants indicated barriers inhibiting them from evaluating their DCT programs. The list reflected issues such as lack of time or a lack of dedicated staff. Other barriers among study participants were the lack of available tools to help design and guide a DCT evaluation as well as the lack of staff skills and knowledge.

DCT Evaluation Practices Framework

The framework that emerged from the information gathered and analyzed for this study (Figure 1) explains the practices around evaluating diversity certificate training programs in higher education. The first circle highlights that other diversity initiatives exist within the higher education context, which is a consideration when the evaluation attempts to show causation. Current political and social events as well as other diversity

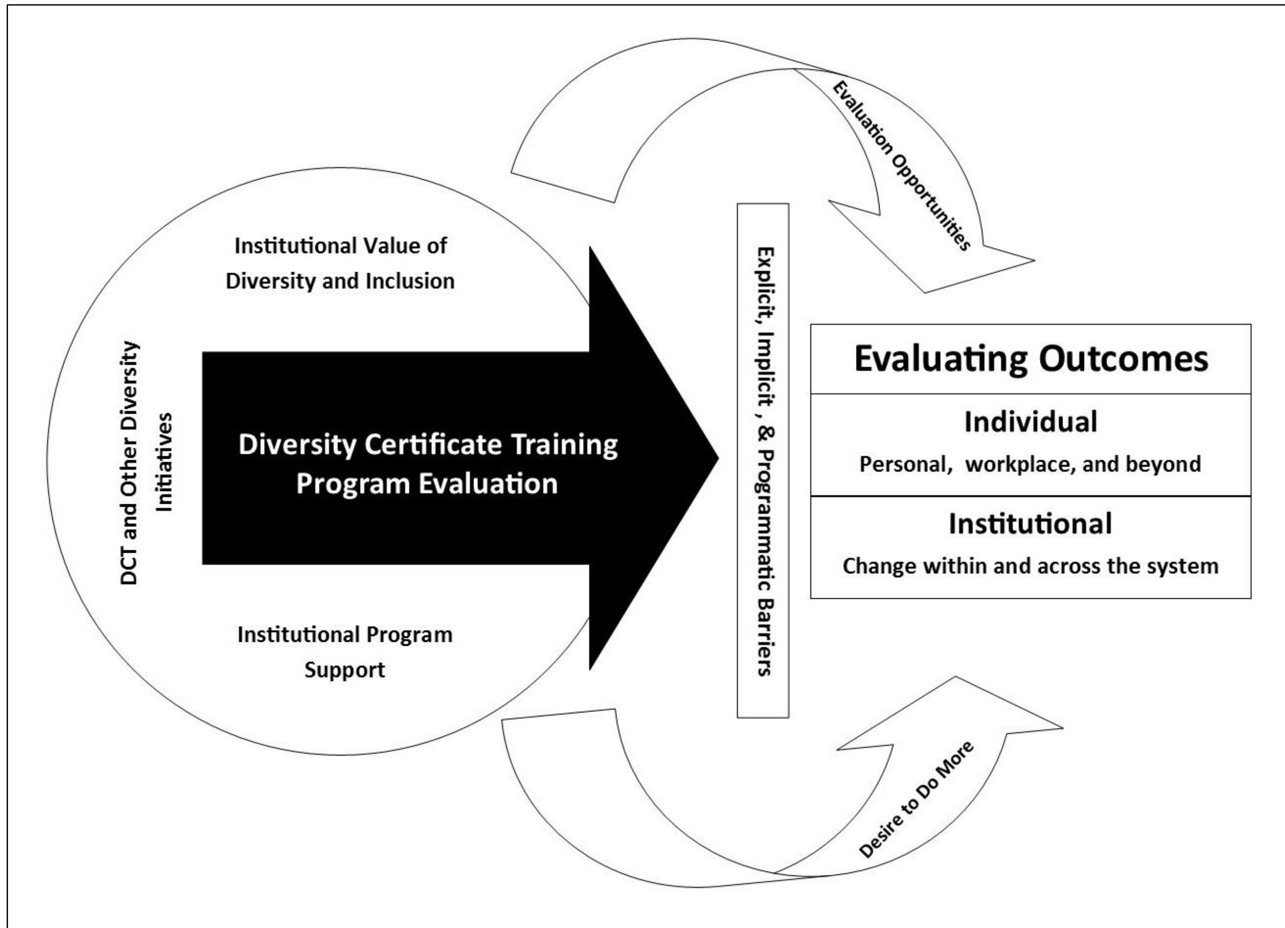
training, mandates, and initiatives within an institution may affect the experience as well as outcomes from DCT programming.

Another point highlighted in the first circle is institutional commitment to diversity. A consideration may be that as administrations and trends change, an institution's commitment to diversity may shift. Shifting priorities may affect DCT program funding. The shift may also require a different evaluation approach to fulfill the accountability needs of the administration.

The arrow pointing to the right represents the program itself. Moving toward the tip of the arrow, a narrow rectangular box depicts the barriers to DCT evaluation. Implicit, explicit, and programmatic barriers inhibit a robust evaluation design. However, the top and bottom arrows depicting evaluation opportunities and a desire to do more to evaluate DCT programming depict two ways in which DCT program managers and evaluators may overcome barriers and conduct the evaluation that yields ample feedback and information to fulfill their institution's accountability needs, offer formative feedback for program improvement, but also begin to demonstrate whether or not their program is having the desired short- and long-term outcomes.

Figure 1.

Diversity Certificate Training Evaluation Practices Theoretical Framework



Diversity Certificate Training Evaluation Design Opportunities

Study informants were unsure how they might enhance their evaluation efforts. This section offers suggestions based on informants' feedback as well as scholarly literature relating to evaluation design and methods. Table 8 lists different options of how to evaluate DCT programs. The evaluation options would yield different types of information that would shed insight into short, medium and long-term DCT outcomes.

The evaluation design table expresses when and how the evaluation would be and the unit of analysis. There are also three types of evaluations reflected in Table 8: formative, for program improvement; applied learning or training transfer for an outcome evaluation; or developmental evaluation, an approach that recognizes programs and the institutions in which the programs exist are dynamic, adapting to complex, changing systems (Patton, 2011).

The methods suggested to design the different types of evaluations include exit surveys, participant narratives, embedded assessments, tracking participant coaching, participant mapping, feedback loops (for change or no change in participants' behaviors), and meta-analysis across evaluation data. The unit of analysis may be an individual participant who is tracked during the program and for a period following completion. Alternatively, the unit of analysis might be across participants or over years. The timing of when to evaluate the program experience, outcomes and long-term effects would be during program participation, such as an exit survey or embedded assessment. A time series evaluation, taking measures over an extended period of time allows a longitudinal understanding of the program effects.

Table 9.
DCT Evaluation Design

What	Formative Evaluation	Applied Learning (Training Transfer)	Developmental Evaluation
How	• Workshop exit surveys • Individual participant narratives • Embedded assessments: (Journals, capstone projects, coaching)	• Individual participant narratives • Critical systems heuristic (advanced participants) • Embedded assessments: (Journals, capstone projects, coaching) • Participant coaching • Ripple effect mapping	• Individual participant narratives (identifying spillover effects) • Embedded assessments Participant mapping by department • Feedback loops • Meta-analysis across surveys, narratives, embedded assess-ments that illustrate examples of change across the institution over time
Unit of Analysis	Individual participants	Individual participants/ aggregated	Meta-analysis across participants
When	During program participation	During and after program completion at specified intervals	During and after program completion at specified intervals

Embedded assessments

Study participants described elements of their programming that could be used in their evaluations, but were not currently using them. This includes participant journals, personal statements and capstone projects. Patton (2002) defined program evaluation as the systematic collection of information about a program, therefore, using program components in the evaluation necessitates a systematic review. Development of a rubric with expected outcomes (e.g. how the participant describes diversity issues at work) would help guide a systematic review of content.

Coaching

Study informants described getting informal feedback when coaching DCT participants. This type of feedback was not included in their evaluation and was not collected systematically. By incorporating a feedback loop process (like a web-based form where the current or past participant would describe the situation or challenge and request help), DCT participants and graduates could receive the help they need, but the program would add this embedded feedback component that would allow the evaluation staff to review and track the feedback as part of the evaluation.

Time Series Evaluation

Tracking participants during and after the program for a set period of time as in a time series evaluation (Russ-Eft & Preskill, 2009) allows the evaluation to measure long-term outcomes. The downside of this method is cost and that it may affect the employee's productivity since there may be repeated interruptions over the course of the set time

period. However, by measuring the outcomes over time, using a time series evaluation approach would address training transfer and understanding what DCT learning and skills, if any, participants were able to apply after the training. Measuring diversity training transfer over time and evaluating how participants apply their learning is vital (Bunch, 2007).

Ripple Effect Mapping (REM)

REM is a participatory evaluation approach that utilizes mind mapping to understand program outcomes (Emery & Flora, 2006; Emery, Higgins, Chazdon, & Hansen, 2014). Using a reflective process, the facilitator guides participants through an appreciative inquiry exercise using mind mapping, a process to mentally envision changes and them articulate them to others. Participants are paired up and interview each other about their experience and how the program may have affected them in different ways. Then the group as a whole reviews the experiences together while a facilitator maps the connections among program experiences and effects, showing the "ripples" or causal pathways (Kollock, Flage, Chazdon, Paine, & Higgins, 2012). The process yields a map of the stories described by participants and illustrates the direct impact as well as spillover and unintended consequences of the program. The process is cost-effective since it takes approximately two hours to achieve a map of the program. A benefit of the method is that the facilitator is able to capture "impacts of complex work," making it a useful tool for diversity programming. The process can also spur additional conversation about the program, inspiring further action (Kollock, Flage, Chazdon, Paine, & Higgins, 2012).

In their research of the effects of a community engagement program, Welborn, Downey, Dyk, Monroe, Tyler-Mackey, and Worthy (2016) found ripple effect mapping as an evaluation tool to be "engaging and well suited to the very heart of civic engagement, the valuing of local knowledge and perspectives as designers of their own communities" (p. 394). Thus, the process yielded not only a causal map of the program's effects, but also engaged the community in an effective way.

The limitations of REM are potential bias based on participant selection. The REM participants may not have adequate information about the program. Another limitation is that the process may not offer critical feedback or negative outcomes from the program (Kollock, Flage, Chazdon, Paine, & Higgins, 2012).

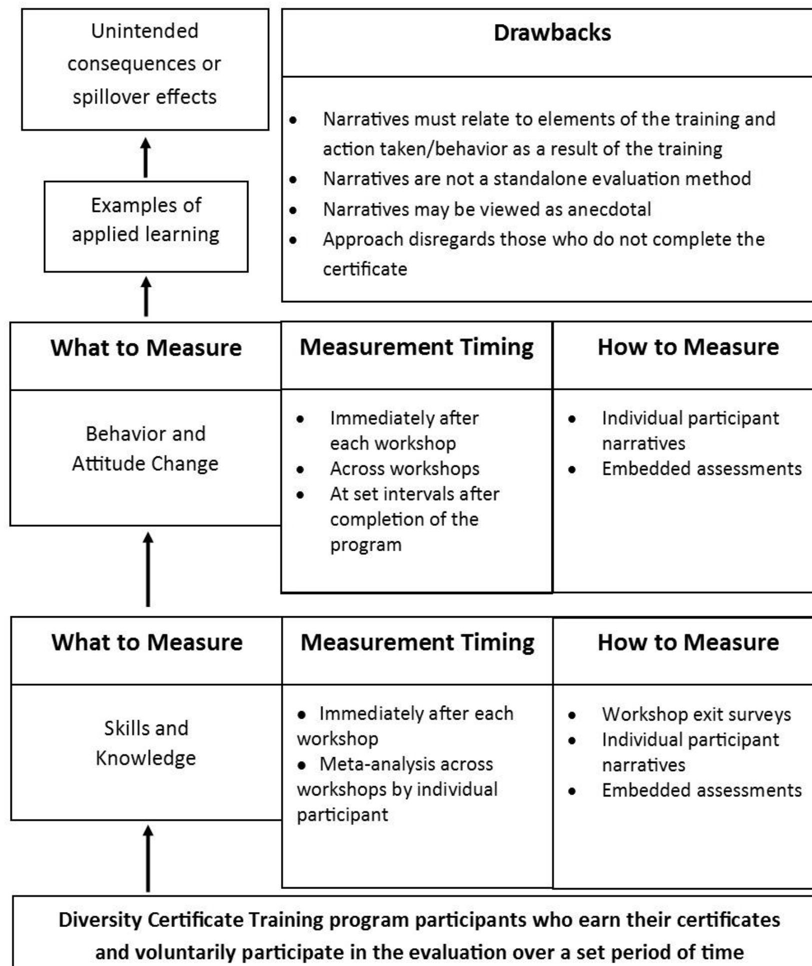
Participant Narratives

Robert Brinkerhoff's (2006) success case story method gives training participants an avenue to articulate their own experience during and after the training, from their perspective. The story telling process is a form of appreciative inquiry and should not be used as a stand-alone evaluative measure. However, rich stories may illuminate the many diverse ways in which not only individual employees, but also their work environments change as a result of DCT participation. Showing the value of the program through narrative and sharing it to the broader community may also help illustrate the program's public value (Franz, 2011).

Figure 2 depicts how and when to gather narratives about a DCT program. The figure also lists drawbacks for this method. An element of the narrative approach might

be to return to the same participants over time to learn how, if at all, their stories have changed.

Figure 2.
DCT Program Evaluation: Individual Participant's Narrative Approach



As with any evaluation approach, there are drawbacks to collecting narratives to tell the impact stories about a program. According to Brinkerhoff (2006), the narrative approach is not recommended as a singular method for reporting on a program's outcomes. Stakeholders and the broader public may consider it anecdotal, therefore triangulating the stories with other evidence gathered from participants is important. The

narratives are also a form of appreciative inquiry, and therefore, it is less likely that critical feedback or negative outcomes will emerge in a story regarding the program.

Developmental Evaluation and Systems Thinking

Some evaluation practices rooted in systems theory may be helpful if applied to DCT evaluation. Educational institutions such as schools and universities are complex systems (Boden, 2011; Henderson, Beach, & Finkelstein, 2011; Lohmann, 2006). Complicated systems are those that have many parts, but the relationship between and among the parts can be clearly defined. Conversely, complex systems weave together parts of whole, but the parts are massively entangled, emergent with patterns that cannot be distinguished from their components. In other words, systems that can be understood by review of their parts are complicated. If the whole system cannot be understood by the sum of its parts, it is complex. (Williams & Hummelbrunner, 2011).

In a higher educational setting, institutions have many complex parts that are constantly shifting. The University cannot be understood by the sum of its parts because they are dynamic, interrelated (massively entangled), and patterns change and emerge in different ways across the institution. Furthermore, institutions in higher education are dynamical systems that change and adapt in nonlinear ways. Relating specifically to diversity training and its potential impact on the institution, the program can demonstrate impact in a predictable way when studied. However, if a racially-motivated incident occurs such as the shooting of an African-American youth followed by a public outcry, this may affect DCT participants' experience, thoughts, and feeling about their own racial identity as well as the training program itself. Carrie (DCT103) mentioned incorporating

current events into her workshops. She specifically referenced discussing the 2015 arson incidents at Black churches in South Carolina. Types of incidences such as this are known in systems theory as "external exacerbations" that may impact the system as well as the program within the system. In this example, not only did the training itself adapt to a current condition, but the experience may have affected training participants and their ability to process the South Carolina incidences in that specific snapshot of time.

Higher education institutions are complex adaptive systems (CAS). In systems theory CASs are defined as a "cluster of individual parts that interact with each other, and over time system-wide patterns appear…the system is constantly shifting. Stable, permanent reality is impossible" (Eoyang & Holladay, 2013, pp. 15, 17). Another scholar suggested that programs and institutions are dynamic (Patton, 2011). Since universities are constantly shifting and DCT programs change to meet the dynamic needs based on current events, funding, and other forces, the evaluation approach should recognize this dynamic.

Employees are human and humans are massively entangled, meaning they do not work and live in isolation (Eoyang, 2006). When employees participate in DCT training, they may apply their learning in multiple aspects of their lives, including at home with their families, faith communities, educational institutions, and other areas of their lives. This may be measured through ripple effect mapping or narrative inquiry, but may be a consideration for the evaluation design since unless solicited from DCT participants, program managers and evaluators may otherwise never know the extent to which participants apply their learning beyond its intended outcomes.

Measuring systems' resilience through positive and negative feedback loops (change or no change) and attending to interaction and interconnectedness among staff (Patton, 2011) is another concept relevant to DCT evaluation. The main current practice common among DCT practitioners was an exit survey at the end of a workshop. The information gathered through that process may yield insight into short-term perceived learning outcomes. However, it would not capture long-term outcomes, especially to distinguish whether or not participants were able to apply what they learned or not. If they were unable to apply their learning in the workplace, probing about systems resilience or the barriers to transferring the skills and knowledge learned may help the institutional leadership understand what barriers within their system inhibit a cultural shift toward inclusion and acceptance.

Another systems exercise that applies to DCT program evaluation is participation mapping. This process recognizes individuals and departments that send or fail to send staff to the DCT program. However, attending to who does and does not complete the training is also critical. Probing participants who attend one or two sessions but do not follow through the end of the program may yield formative feedback on why the program did not work for some participants. Likewise, an inquiry geared toward participants who drop out of the program may help program managers understand what, if any, systems barriers persist, preventing participation in the program and ultimately, preventing change from occurring in a department.

A critical systems heuristic (CSH) (Ulrich, 1996; Ulrich, 2000; Ulrich & Reynolds, 2010) developed for DCT participants would facilitate a critical reflection on

the participant's job function and workplace. A CSH is a tool that provides a framework for reflective questioning regarding a program or institution. The process considers "boundary judgements" (Ulrich, 2000) that facilitate reflection about the underlying assumptions to understand the purpose of the system, whom the system serves, and how decisions are made (Ulrich, 1996; Ulrich, 2000). The heuristic could be developed by the DCT program trainer or by an advanced training participant as a means to critically analyze the current functions within the system or portion of the institution related to the participant's work function. The heuristic would be revisited over time to see what aspirations were elucidated previously using the heuristic and how, if at all, the operation changed to better meet the diversity and equity aspirations to ultimately more effectively serve diverse student and stakeholder populations within the higher education institution.

Ulrich's critically heuristic boundary questions (2000) cover four areas: sources of motivation, sources of power, sources of knowledge, and sources of legitimization. Within each area, questions are asked in what "is" and what "ought" to be modes, looking at the current perception about the phenomena and the potential of what the program could be doing differently. For example, a question under sources of motivation is "Who is (ought to be) the beneficiary of DCT programming? Whose interests are (or should be) served?" A DCT participant may consider his boss or supervisor the beneficiary of the program since the employee will learn new skills that may affect his job. However, if the participant considers who should be the primary beneficiary, she might realize that, ultimately, diverse students and staff should be the beneficiaries of her actions that result from the learning in the program.

The drawback to using a critical systems heuristic as an evaluation tool is that developing appropriate questions may be challenging for someone who is not well versed in this particular form of reflective questioning. Although Ulrich (2000) offered a template, the tool would need to be tailored to DCT programming or the employee's departmental functions. Furthermore, when implementing such a tool in a sensitive context such as discussions around diversity, equity, and oppression, participants must have ground rules and fully understand the tool's purpose and how to understand and use the responses to further their understanding of diversity and inclusion in their workplace.

In sum, there are opportunities for designing and conducting a mixed method evaluation that would yield both short- and long-term program outcomes. Using narratives or ripple effect mapping may yield information about direct and spillover effects that would otherwise not be known. Taking a systems approach, recognizing that universities are complex adaptive systems that are dynamic allows for system mapping or the development of a heuristic to look at the program and its outcomes in different ways to understand whether or not participants effect change within their workplace. Evaluating beyond the exit survey will inform program managers, university administrators, and other stakeholders about outcomes that, according to the current practices documented in this study, are yet unknown.

Suggestions for Future Research

There are many themes budding out of this study that warrant additional research. After extensive searching through scholarly literature, I found very little published about diversity certificate training programs specifically. There is information about diversity

programming in general, but not about how the DCT model is different and may have different outcomes. This is not a one-time workshop. DCT programming is typically not required, nor is its primary goal lawsuit avoidance. What emerged from the data in this study is that DCT workshops are repeated over time, allowing for time to practice new skills and ongoing coaching. One of the most important elements of the program is that it culminates in acknowledging that participants fulfill the requirements for the program by receiving a certificate at the end. Additional research is required for scholars to understand what, if any, difference in impact this model has on participants over other types of diversity programming models. Since my exhaustive initial search only yielded eleven institutions in the United States with this DCT model program, perhaps additional research would help other institutions adopt the model as well as learn from existing programs in a way that is not currently possible due to the lack of published information.

A future research study might contrast existing programs, delving into distinct elements of the programming such as requiring a capstone project or written journal. Cross institution evaluations of the programming would elaborate on how different program attributes affect program outcomes for individuals and their institutions. This would help researchers and program managers understand what elements of the DCT programs appear transformative and to what degree the outcomes are achieved.

My main recommendation for future research is to test the evaluation options to understand what works in a practical application. Without pushing institutions to adopt more measures and frame their evaluations to look at the various levels of impact, we

cannot be confident about the impact of the diversity certificate training model on individuals, their institutions, and potentially the community beyond the institution.

Conclusion

In this study, I highlighted the problem facing diversity training staff and evaluators. Current practices based on practitioner reports indicated that diversity programs were not measuring the outcomes of their program on individuals as a result of training. They reported changes such as enactment of new policies or the numbers of participants in the program, implicitly suggesting that these measures translated to changes in the organization's culture, without offering any evidence to support such a claim. Scholarly human resource literature included training evaluation models, such as Kirkpatrick's (2006) four-level model. After analyzing and contrasting additional training evaluation models (Hamblin, 1974; Phillips, 2002), I concluded that little scholarly research exists addressing how, if at all, these models may fit to guide multi-level impact evaluation considering the nuances of diversity certificate training programs in higher education.

The purpose of this study was to understand how training managers in higher education evaluate the impact of their diversity certificate training programs. It is important to note that the purpose of the study was not to assess the outcomes of the programs themselves nor to delve into which practices in the various programs may or may not be promising. The study employed a grounded theory approach to develop an evaluation framework tailored to diversity certificate training programs at nine higher educational institutions in the United States.

A secondary review of the programs' and institutions' websites illustrated that the universities valued diversity and sought to shift their institutional culture toward inclusion, tolerance, and equity. The staff I interviewed for this study reinforced that the purpose of diversity certificate training programs was aligned with their institution's strategic goals around diversity and inclusion. They viewed the program as an avenue for achieving transformation.

However, when informants described their evaluation practices, they focused primarily on gathering formative feedback for program improvement. When probed about evaluating impact on individual participants, some study participants described exercises that could be embedded assessments, such as reflective journals that might help illuminate potential shifts in attitude or behavior over time. However, when asked about how they incorporated these elements into their evaluations, the response was that they used them for formative feedback or they did not incorporate them beyond checking to ensure program participants completed their required tasks.

Study participants conducted formative evaluations and relied on informal ways of knowing their programs' outcomes. They described real-world barriers of "biting off more than they can chew" as a limitation of incorporating more into their evaluation process beyond an exit survey with a formative focus. Other barriers included lack of knowledge or evaluation expertise to implement a multi-level evaluation. Ultimately, this study revealed that current practices evaluating DCT programs were falling short of gathering in depth, longitudinal data to fully understand short- and long-term individual, departmental and institutional outcomes.

Epilogue

In an era where the deaths of African American youth at the hands of the police are a frequent occurrence as in the cases of Michael Brown in Ferguson, Missouri in 2014 (Healy, 2014); Freddie Gray in Baltimore, Maryland (Bever & Ohlheiser, 2015); Jamar Clark in Minneapolis, Minnesota (Graham, 2015); Sandra Bland in Wallen County, Texas (Montgomery, 2016); Tamir Rice in Cleveland, Ohio (Williams & Smith, 2015); Melissa Ventura in Yuma, Arizona (Downs, 2016); Alton Sterling in Baton Rouge Louisiana (Fausset, Perez-Pena, & Robinson, 2016); Pedro Villanueva in Los Angeles, California (Downs, 2016); Anthony Nunez in San Jose, California (Downs, 2016) or Philando Castile in Falcon Heights, Minnesota (Yuen & Feshir, 2016) among others. Institutional bias also persists as a significant barrier for full participation of those who identify as non-dominant in our society. Programs that enable people to understand their own racial identity awareness and work toward social justice to shift systemic bias are in higher demand than ever. As study participants described, one way they intuitively feel their program is having an impact is that their sessions are not only full, but have waiting lists. Do the programs stimulate change within the participants? Do participants transfer what they learned from the DCT workshops and practice their new skills in conflict management, cultural responsiveness, and cross cultural communication in the workplace? At this time little is known about such effects due to the lack of scholarly research and published evaluation for this program model.

The DCT program designers, managers, and trainers are to be commended for their efforts to transform their institutions toward tolerance, inclusion, and respect for all

students, faculty, and staff. Their commitment to equity and social justice within the institutions was palpable during the conversations I had with them. I am truly inspired by the practitioners who developed DCT programming and who commit their lives to equity and change.

Systemic change is slow and incremental. Social theory suggests that when we reach a normative tipping point, the culture within an institution will shift toward inclusion and push those unwilling to change to the margins of the organization. If we are to celebrate true, deep transformation, albeit slow and incremental, we must require accountability of diversity certificate programs. We must be willing to recognize the slow pace that the transformation requires since changing deep-seated values is an incredibly complex endeavor. We must be willing to challenge each other to commit to diversity as a personal as well as workplace shared value.

Most importantly, we must continue to develop a sense of our own identity awareness, especially committing to an authentic look at our own power and privilege. At some point in the very distant future, humans may learn to respect the inherent worth and dignity of each person they meet. They may be able to act with compassion, equity, and justice for all. It will be at that point when diversity staff know their work is done. Today, however, our society desperately needs diversity training models that have transformative outcomes for participants and serve as a catalyst for institutional change. Moreover, such programs need evaluations to demonstrate whether or not change is occurring at the individual and institutional levels.

Bibliography

Alexander, M. (2012). *The new Jim Crow: Mass incarceration in the age of colorblindness.* New York, NY: The New Press.

Banks, J. A. (2006). *Race, culture, and education: The selected works of James A. Banks.* New York: Routledge.

Barndt, J. (2007). *Dismantling racism: The twenty first century challenge to White America.* Minneapolis, MN: Augsburg Fortress.

Beaudry, J. (1992). Synthesizing research in multicultural teacher education: Findings and issues in evaluation of cultural diversity. *New Directions for Program Evaluation, 53,* 69 – 86.

Bennett, J. & Bennett, M. (2004). Developing intercultural sensitivity: An integrative approach to global and domestic diversity. In D. Landis, J. Bennett & M. Bennett (Eds.), *Handbook of intercultural training* (3rd Ed), (pp. 147 – 165). Thousand Oaks, CA: Sage.

Berger, P. & Luckmann, T. (1966). *The social construction of reality: A treatise in the sociology of knowledge.* Garden City: NY, Doubleday & Company.

Bersin, J. (2008). *The training measurement book: Best practices, proven methodologies, and practical approaches*, (1st ed.). San Francisco: Wiley & Sons.

Bever, L., & A. Ohlheiser. (April 20, 2015). Baltimore Police: Freddie Gray died from a 'tragic injury to his spinal cord'. *Washington Post* retrieved online from https://www.washingtonpost.com/news/morning-mix/wp/2015/04/20/baltimore-police-freddie-gray-arrested-without-force-or-incident-before-fatal-injury/?utm_term=.9b55a8c2029a.

Bicchieri, C. (2006). *The grammar of society: The nature and dynamics of social norms*. New York, NY: Cambridge Press.

Bickman, L., & Riemer, M. (2011). Using program theory to link social psychology and program evaluation, in M. Melvin, S. Donaldson, & B. Campbell (Eds.) *Social psychology and evaluation,* pp. 104-140.

Blitz, L., & Illidge, L. (2006). Not so black and white. *Journal of Emotional Abuse, 6*, 113-134.

Boden, G. T. (2011). Retention and graduation rates: Insights from an extended longitudinal view. *Journal of College Student Retention: Research, Theory & Practice, 13*(2), pp. 179-203.

Bonilla-Silva, E. (2010). *Racism without racists*. Lanham, MD: Rowman & Littlefield.

Brinkerhoff, R. O. (2006). *Telling training's story*. San Francisco: Berrett-Koehler.

Brinkerhoff, R. O. (2005) The success case method: A strategic evaluation approach to increasing value and effect of training. *Advances in Developing Human Resources, 7*, 86-101.

Brinkerhoff, R. O. (2002). *The success case method.* San Francisco: Berrett-Koehler.

Bryant, A. (2009). Grounded theory and pragmatism: The curious case of Anselm Strauss. *Qualitative Social Research, 10*(3). Retrieved online from http://www.qualitative-research.net/index.php/fqs/article/view/1358/2850#g12

Bunch, K. (2007). Training failure as a consequence of organizational culture. *Human Resource Development Review, 6*(2), 142-163.

Buckley, R. & Caple, J. (2004). *The Theory & Practice of Training, 5th Ed.* London, England: Kogan-Page.

Cabrera, N. (2014). Exposing whiteness in higher education: white male college students minimizing racism, claiming victimization, and recreating white supremacy. *Race Ethnicity and Education*, *17*(1), pp. 30-55.

The Carnegie Foundation for the Advancement of Teaching classifies higher education institutions. Retrieved online January 6, 2016 from: carnegieclassifications.iu.edu/.

Castro Valley Unified School District. (2008). *Equity Action Plan*, rev. 2008.

Charles Stuart Mott Foundation (2007). Crossroads Anti-racism organizing & training; teaching and training methodology; documentation and evaluation report. Retrieved online from http://crossroadsdiversity.org/wp-content/themes/crossroads/PDFs/Crossroads_Anti-Racism_Documentation_w-_Appendices.pdf.

Charmaz, K. (2014). *Constructing grounded theory.* (2nd ed.). Thousand Oaks, CA: Sage.

Chen, H. T., Donaldson, S. I., & Mark, M. M. (2011). Validity frameworks for outcome evaluation. *New Directions for Evaluation, 130*, 5-16.

Chenail, R. (2011). Interviewing the investigator: Strategies for addressing instrumentation and researcher bias concerns in qualitative research. *The Qualitative Report, 16,* 255-262.

Clements, PE. & Jones, J. (2008). *The diversity training handbook: A practical guide to understanding and changing attitudes, 3rd Ed*. London: Kogan Page Limited.

Cole, B. M., & Salimath, M. S. (2013). Diversity identity management: An organizational perspective. *Journal of Business Ethics, 116,* 151-161.

Cole, P., & Secret, M. (2012). Factors associated with work–family conflict stress among African American women. *Social Work in Public Health, 27,* 307-329.

Condron, D., Tope, D., Steidl, C., & Freeman, K. (2013). Racial segregation and the black/white achievement gap 1992-2009. *The Sociological Quarterly, 54,* 130-157.

Cooper, M. (December 12, 2012). Census officials, citing increasing diversity, say U.S. will be a 'plurality nation', *New York Times,* retrieved from http://www.nytimes.com/2012/12/13/us/us-will-have-no-ethnic-majority-census-finds.html?_r=0.

Cordes, A. (2009). A window and a mirror: Looking out, looking in. A case study of the Saint Paul Foundation: Documenting a comprehensive, institution-wide commitment to anti-racism. Saint Paul, MN: The Saint Paul Foundation. Retrieved from: https://www.facingrace.org/_asset/hf8y3s/facingrace_2009_Windows-Mirror_full.pdf

Creswell, J. (2003). *Research design: Qualitative, quantitative, and mixed-methods approaches,* (2nd ed.). Thousand Oaks, CA: Sage Publications.

Deardorff, D. (2008). Intercultural competence: A definition, model and implications for education abroad. In V. Savicki (Ed.) *Developing Intercultural Competence and Transformation,* (pp. 32-52). Sterling, VA: Stylus.

Delgado, G., & Browne, B. (1992). Anti-racist work: An examination and assessment of organizational activity. Oakland, CA: Applied Research Center.

Delgado, R., & Stefancic, J. (2001). *Critical race theory: An introduction.* New York, NY: New York University Press.

Douglas, D., & Halas, J. (2011). The wages of whiteness: Confronting the nature of ivory tower racism and the implications for physical education. *Sport, Education and Society, 18*(4), 453-474.

Downs, K. (July 14, 2016). Why aren't more people talking about Latinos killed by police? *PBS Newshour*, retrieved online from http://www.pbs.org/newshour/rundown/black-men-werent-unarmed-people-killed-police-last-week/.

Edwards, K. (2006). Aspiring social justice ally identity development: A conceptual model. *NASPA Journal, 43,* 29-60.

Emery, M. & Flora C. (2006). Spiraling-up: mapping community transformation with community capitals framework. *Journal of Community Dev Society. 237*(1) pp. 19–35.

Emery, M., Higgins, L., Chazdon, S., & Hansen, D. (2014). Using ripple effect mapping to evaluate program impact: choosing or combining the methods that work best for you. *Journal of Extension, 53*(2), retrieved online from http://www.joe.org/joe/2015april/tt1.php.

Eoyang, G. (2006). Human systems dynamics: Complexity-based approach to a complex evaluation, in B. Williams & I. Imam (Eds.) *Systems Concepts in Evaluation: An Expert Anthology.* Point Reyes, CA: American Evaluation Association.

Eoyang, G., & Holladay, R. (2013). *Adaptive Action: Leveraging uncertainty in your organization.* Stanford: Stanford University Press.

Equal Opportunity Employment Commission data report Retrieved June 3, 2014 from http://www1.eeoc.gov/eeoc/statistics/enforcement/state_13.cfm

Faegin, J. (2010). *The white racial frame.* New York, NY: Routledge.

Fausset, R., Perez-Pena, R., & C. Robinson. (July 6, 2016). Alton Sterling shooting in Baton Rouge prompts Justice Department investigation. *New York Times,* retrieved from https://www.nytimes.com/2016/07/06/us/alton-sterling-baton-rouge-shooting.html.

Fendt, J. & Sachs, W. (2008). Grounded Theory Method in Management Research: Users' Perspectives. *Organizational Research Methods,11*(3), 430–455.

Fernandez, R. & Fernandez-Mateo, I. (2006). Networks, race, and hiring. *American Sociological Review, 71,* 42-71.

Fitzpatrick, J. L., Sanders, J. R., & Worthen, B. R. (2004). *Program Evaluation: Alternative Approaches and Practical Guidelines, 3rd Ed.* Boston: Pearson Education.

Fowler, S., & Blohm, J. (2004). An analysis of methods for intercultural training. In D. Landis, J. Bennett & M. Bennett (Eds.), *Handbook of intercultural training.* (3rd ed.), (pp. 37 – 84). Thousand Oaks, CA: Sage.

Franz, N. (2011). Advancing the public value movement: Sustaining Extension during tough times. *Extension and Outreach Research and Scholarship*. Paper 7. Retrieved online from http://lib.dr.iastate.edu/extension_research/7.

Gallagher, C. A. (2006). The challenge to public sociology: Neoliberalism’s illusion of inclusion. In *Public sociologies reader*, (pp. 293-304). J. Blau and K. Smith (Eds.) Lanham, MD: Rowman and Littlefield, Publishers.

Gallos, J. V. (1997). *Teaching diversity: Listening to the soul, speaking from the heart*. San Francisco: Jossey-Bass.

Garces, L. & Cogburn, C. (2015). Beyond declines in student body diversity: How campus-level administrators understand a prohibition on race-conscious postsecondary admissions policies. *American Journal of Educational Research, 52*(5), 828-860.

Glaser, B. G., & Strauss, A. L. (1967). *The discovery of grounded theory: Strategies for qualitative research*. Chicago: Aldine Pub. Co.

Glaser, Barney G. (1992). *Basics of grounded theory analysis.* Mill Valley, CA: The Sociology Press.

Glaser, B. G. (2005). *The grounded theory perspective III: Theoretical coding*. Mill Valley, CA: Sociology Press.

Graham, D. (November 18, 2015). How did Jamar Clark die? *The Atlantic* retrieved from https://www.theatlantic.com/politics/archive/2015/11/jamar-clark-is-the-next-great-police-brutality-controversy/416418/.

Graham, R. (November 15, 2016). By normalizing hatred, we've already normalized Trump. *Boston Globe,* retrieved from: https://www.washingtonpost.com/opinions/global-opinions/donald-trump-is-normalizing-bigotry/2016/09/30/9d754b2e-8731-11e6-a3ef-f35afb41797f_story.html?utm_term=.4a2860cd5d5b

Greene, M. & Siskind, A. (2006). Forward: from a multicultural institution to an antiracist institution: A traditional Jewish organization meets the challenge. *Journal of Emotional Abuse, 6*, xxvii-xxxiv.

Guess, S. (2006). The social construction of whiteness: Racism by intent, racism by consequences. *Critical Sociology, 32*, 649 – 673.

Gurin, P. (1999). Expert report of Patricia Gurin. *Equity & Excellence in Education, 32,* 32-62.

Guskey, T. (March, 2002). Does it make a difference? Evaluating professional development. *Educational Leadership,* 45-51.

Guskey, T. (2012). Focus on key points to develop the best strategy to evaluate professional learning. *JSD, The Learning Forward Journal, 33*(4), 40-43.

Guskey, T. (2014a). Evaluating professional learning, in *International Handbook of research in professional and practice-based learning, Springer international handbooks of education.* (pp. 1215-1235). S. Billett, C. Harteis, H. Gruber (Eds.) Netherlands: Springer.

Guskey, T. (2014b). Planning professional learning. *Professional Learning: Reimagined, 71*(8), 10-16.

Gynnild, A. (2011). Book Review: Grounded Theory: A practical guide (Birks & Mills, 2011). *The Grounded Theory Review, 10*(3), 63–66.

Haider, A., Sexton, J., Sriram, N., Cooper, L., Efron, D., Swoboda, S., Villegas, C., Haut, E., Bonds, M., Pronovost, P., Lipsett, P., Freischlag, J., & Cornwell, E. (2011). Association of unconscious race and social class bias with vignette-based clinical assessments by medical students. *Journal of the American Medical Association, 306*(9), 942-951.

Hall, N. (2013). *Hate crime, 3rd Ed*. New York: Routledge.

Hamblin, A. C. (1974). *Evaluation and control of training.* Maidenhead, Berkshire, England: McGraw-Hill.

Hardie, J. & Tyson, K. (2013). Other people's racism: Race, rednecks, and riots in a southern high school. *Sociology of Education, 86,* 83-102.

Harper, S. (2012). *Black Male Student Success in Higher Education: A Report from the National Black Male College Achievement Study*. Retrieved online from http://works.bepress.com/cgi/viewcontent.cgi?article=1049&context=sharper.

Healy, J. (November 26, 2014). Ferguson, still tense, grows calmer. *New York Times,* retrieved from https://www.nytimes.com/2014/11/27/us/michael-brown-darren-wilson-ferguson-protests.html.

Henderson, C., & Beach, A., & Finkelstein, N. (2011). Facilitating change in undergraduate STEM instructional practices: An analytic review of the literature. *Journal of Research in Science Teaching, 48*(8), 952-984.

Henry, G. T. (2000). Why not use? In V.J. Caracelli & H. Preskill (Eds.), The expanding scope of evaluation use. *New Directions for Evaluation, 88,* 85-98. San Francisco: Jossey-Bass.

Hikido, A. & Murray, S. B. (2015). Whitened rainbows: How white college students protect whiteness through diversity discourses. *Race and Ethnicity in Education, March*, 1-23.

Hogan, D., & Mallott, M. (2005). Changing racial prejudice through diversity education. *Journal of College Student Development, 46,* 115-125.

Hurtado, S., Engberg, M., Ponjuan, L., & Landreman, L. (2002). Students' precollege preparation for participation in a diverse democracy. *Research in Higher Education, 43,* 163-186.

Ivancevich, J., & Gilbert, J. (2000). Diversity management: Time for a new approach. *Public Personnel Management, 29*(1), pp. 75-92.

Jeyasingham, D. (2012). White noise: A critical evaluation of social work education's engagement with whiteness studies. *British Journal of Social Work, 42*, 669-686.

Kaufman, R., Keller, J., & Watkins, R. (1995). What works and what doesn't: Evaluation beyond Kirkpatrick. *P&I, November-December*, 8 - 12.

Kirkpatrick, D. L. (2006). Seven keys to unlock the four levels of evaluation. *Performance Improvement, 45*(7), 5-8.

Kollock, D. H., Flage, L., Chazdon, S., Paine, N., & Higgins, L. (2012). Ripple effect mapping: a "radiant" way to capture program impacts. *Journal of Extension, 50*(5) article number 5TOT6.

Koumpilova, M. (July 17, 2013). St. Paul school board OKs groundbreaking racial equity policy. *Pioneer Press*, retrieved from http://www.twincities.com/ci_23674468/st-paul-school-board-oks-groundbreaking-racial-equity.

Kraiger, K., Ford, J. K., & Salas, E. (1994). Application of cognitive, skill-based, and affective theories of learning outcomes to new methods of training evaluation. *Journal of Applied Psychology, 78*(2), 311-328.

Kulik, C. T., Pepper, M. B., Roberson, L., & Parker, S. (2007). The rich get richer: predicting participation in voluntary diversity training. *Journal of Organizational Behavior, 28*(6), 753-769.

Kulik, C. T. & Roberson, L. (2008). Diversity initiative effectiveness: What organizations can (and cannot) expect from diversity recruitment, diversity training, and formal mentoring programs. *Diversity at Work.* Brief, A. P. (Ed.). Pp. 265-305. Cambridge, UK: Cambridge University Press.

Kvale, S., & Brinkmann, S. (2009). *Interviews: Learning the craft of qualitative research interviewing*, 2nd ed. Thousand Oaks, CA: Sage.

Leiderman, S., & George, M. (2005). Looking back: Project Change from 1991-2005. Center for Assessment and Policy Development. Retrieved online from: http://www.capd.org/pubfiles/pub-2005-04-01.pdf.

Lewin, H. M. & McEwan, P. J. (2001). *Cost-effectiveness analysis: Methods and applications, 2nd Ed*. Thousand Oaks: Sage.

Logan, E. (2011). *At this defining moment: Barak Obama's presidential candidacy and the new politics of race*. New York, NY: New York University Press.

Lohmann, S. (2006). The public research university as a complex adaptive system, In: *Proceedings of the European Conference on Complex Systems*, University of Oxford, September 25-29, 2006.

Martin, D. (2009). Researching race in mathematics education. *Teachers College Record, 111*(2), 295-338.

Martin, M.B. (2011). A Commentary on Ekins (2011).*The Grounded Theory Review, 10*(3), 57–59.

McDermott, M. & Samson, F. (2005). White racial and ethnic identity in the United States. *Annual Review of Sociology, 31,* 245-261.

McGuire, D., & Bagher, M. (2013). Diversity training in organsations: an introduction. *Journal of European Industrial Training, 34*(6), 493-505.

Merriam, S. (2009). *Qualitative research: A guide to design and implementation*. San Francisco, CA: Jossey-Bass.

Mertens, D. M. (2009). *Transformative research and evaluation.* New York, NY: Guilford Press.

Mertens, D. M. & Wilson, A. T. (2012). *Program evaluation theory and practice: A comprehensive guide.* New York, NY: Guilford Press.

Miller, L. (2014). ASTD's 2013 State of the Industry Report: Workplace learning. *T&D, November,* 41-45.

Milner, R. (2013). Why are students of color still punished more severely and frequently than white students? *Urban Education, 48,* 483-489.

Moniz, J. & Spickard, P. (2006). Carving out a middle ground. In Brunsma, D. (Ed.), *Mixed messages* (pp. 63-81). Boulder, CO: Rienner.

Montgomery County Public Schools. (2016). Our Call to Action: Pursuit of Excellence 2011-2016, retrieved online from http://www.montgomeryschoolsmd.org/uploadedFiles/about/strategicplan/OurCallToAction.pdf

Montgomery, D. (January 6, 2016). Texas trooper who arrested Sandra Bland is charged with perjury. *New York Times.* Retrieved from https://www.nytimes.com/2016/01/07/us/texas-grand-jury-sandra-bland.html

Morgan, P., Farkas, G., Hillemeier, M., & Maczuga, S. (2016). Science achievement gaps begin very early, persist, and are largely explained by modifiable factors. *Educational Researcher, 45*(1), 18-35.

Murdoch University, Centre for Social Change & Social Equity. (2003). Anti-racism: what works? An evaluation of the effectiveness of anti-racism strategies. Perth, Australia: Murdoch University. Retrieved online at: http://www.omi.wa.gov.au/resources/clearinghouse/diversity_what_works.pdf

Murray, L. (2004). The San Jose Unified School District Story: Implementing a college-ready curriculum for all. *Profiles in Leadership, 4th Ed*. pp. 87-94.

Muskal, M. (December 12, 2012). Whites to lose majority status by 2043, the census projects. *Los Angeles Times*, retrieved from http://articles.latimes.com/2012/dec/12/nation/la-na-nn-whites-to-lose-majority-status-by-2043-the-census-projects-20121212.

Myers, S. L. (2011). The Economics of diversity: The efficiency versus equity trade off in diversity management. In S. Chen (Ed.), *Theoretical perspectives and practical approaches,* New York: Nova Science Publishers, 55-70.

Ngo, B., Myers, S., Chang, Y., Orville, R., & Illes, M. (2012) The Higher Education Anti-Racism (H.E.A.R.T.) Team collaboration report.

Park, J. J., & Denson, N. (2009). Attitudes and Advocacy: Understanding faculty views on racial/ethnic diversity. *The Journal of Higher Education, 80*(4), 415-438.

Parry, S. (2000). *The ASTD Handbook of Training Design and Delivery*, G. M. Piskurish, P. Beckschi, & B. Hall, Eds. Pp. 147-157. New York: McGraw-Hill.

Patton, M. Q. (2002). *Qualitative research and evaluation methods* (3rd ed.). Thousand Oaks, CA: Sage.

Patton, M. Q. (2008). *Utilization-focused evaluation*, (4th ed). Thousand Oaks, CA: Sage.

Patton, M. Q. (2011). *Developmental evaluation: Applying complexity concepts to enhance innovation and use.* New York: Guilford Press.

Peacock, C. & Daniels, G. (2006). Applying an antiracist framework to a residential treatment center. *Journal of Emotional Abuse, 6,* 135-154.

Pendry, L., & Driscoll, D. (2011). Five guiding principles to help improve diversity-training assessment. *Training and Management Methods, 25,* 2.01-2.19.

Pendry, L., Driscoll, D., & Field, S. C. T. (2007). Diversity training: Putting theory into practice. *Journal of Occupational and Organizational Psychology, 80*(1), 27-50.

Peshkin, A. (1988). *In search of subjectivity. One's own. Educational Researcher, 17*, 17-21.

Phillips, J. & Stone, R. (2002). *How to measure training results*. New York: McGraw-Hill.

Portland Public Schools. (2011). Portland Public Schools Racial Educational Equity Policy 2.10.010-P adopted June 13, 2011.

Posselt, J., Jaquette, O., Bielby, R., & Bastedo, M. (2012). Access without equity: Longitudinal analyses of institutional stratification by race and ethnicity, 1972–2004. *American Educational Research Journal, 49*(6), 1074-1111.

Preskill, H. & Torres, R. T. (1999). *Evaluative inquiry for learning organizations*. Thousand Oaks, CA: Sage Publications.

Ramsey County Research and Evaluation Department. (2010). Bringing about organizational change: Highlights of anti-racism efforts by Ramsey County Community & Human Services: Accomplishments and lessons learned 2005 – 2010. Retrieved from http://www.co.ramsey.mn.us/NR/rdonlyres/8A612DA6-3D77-499B-B902-827887DA878D/29030/ARLTLessonsLearned911.pdf.

Roberson, L., Kulik, C. T., & Pepper, M. B. (2009). Individual and environmental factors influencing the use of transfer strategies after diversity training. *Group & Organizational Management*, *34*, 67-89.

Roberson, L., Kulik, C. T., & Tan, R. Y. (2014). Effective diversity training, pp. 341 - 365. *Oxford Handbook of Diversity and Work,* Q. Roberson (ed.). Oxford, England: Oxford University Press.

Ropers-Huilman, R., Winters, K., & Enke, K. (2013). Discourses of whiteness: White students at Catholic women's colleges (dis)engaging race. *The Journal of Higher Education, 84*(1), 28-55.

Rosen, Y., & Salomon, G. (2011). Durability of peace education effects in the shadow of conflict. *Social Psychology Education, 14*, 135-147.

Rothstein, R. (2013). Why our schools are segregated. *Educational Leadership, 70,* 50-55.

Rubineau, B., & Kang, Y. (2012). Bias in White: A Longitudinal Natural Experiment Measuring Changes in Discrimination. *Management Science 58*(4), pp. 660-677.

Russ-Eft, D. (2010). Expanding scope of evaluation in today's organizations. In J. L. Moseley & J. C. Dessinger, (Eds.), *Handbook of Improving Performance in the Workplace, 3,* 339 - 353. New York: Pfeiffer.

Russ-Eft, D., Bober, M. J., de la Teja, I., Foxon, M. J., & Koszalka, T. A. (2008). *Evaluator Competencies: standards for the practice of evaluation in organizations.* San Francisco: Jossey-Bass.

Russ-Eft, D. & Preskill, H. (2009). *Evaluations in Organizations: A Systematic Approach to Enhancing Learning, Performance and Change. 2nd Ed*. New York: Basic Books.

Saint Paul Public Schools (July 16, 2013). Racial equity policy. Retrieved from http://boe.spps.org/uploads/racialequitypolicy_spps_final_071613.pdf.

Salas, E., Salazar, M. R., & Gelfand, M. J. (2013). Understanding diversity as culture, in *The Oxford Handbook of Diversity and Work*, Q. Roberson (Ed.). Oxford, England: Oxford University Press.

Scriven, M. (1967). The methodology of evaluation. In R. E. Stake (Ed.), *Curriculum evaluation*. In *American Educational Research Association monograph series on evaluation, No. 1*, 39-83. Chicago: Rand McNally.

Shapiro, I. (2006). *Extending the Framework of Inquiry: Theories of change in conflict interventions.* Berghof Research Center for Constructive Conflict Management.

Snyder, T., de Brey, C., & Dillow, S. (2016) *Digest of educational statistics 2014, 50th edition*. NCES 2016-006. National Center for Education Statistics. Retrieved online from http://nces.ed.gov/pubs2016/2016006.pdf.

Stainbeck, K., & Irvin, M. (2012). Workplace racial composition, perceived discrimination, and organizational attachment. *Social Science Research, 41,* 657-670.

Stake, R. (2006). *Multiple case study analysis.* New York, NY: Guilford Press.

Strauss, A. L. (2010). *Qualitative analysis for social scientists*. Cambridge: Cambridge University Press.

Strauss, A. L., & Corbin, J. M. (1990). *Basics of qualitative research: Grounded theory procedures and techniques.* Newbury Park, Calif: Sage Publications.

Strauss, A., & Corbin, J. (1998). *Basics of Qualitative Research: Techniques and procedures for developing grounded theory*, 2nd ed. Thousand Oaks, CA: Sage.

Taylor, P. J., Russ-Eft, D. F. & Chan, D. W. L. (2005). A meta-analytic review of behavior modeling training. *Journal of Applied Psychology, 90*(4), 692-709.

Thomas, D. R. (2006). A General Inductive Approach for Analyzing Qualitative Evaluation Data. *American Journal of Evaluation, 27*(2), pp. 237-246.

Ting-Toomey, S. (1999). *Communicating across cultures.* New York, NY: The Guilford Press.

Trenberry, B., & Paradies, Y. (2012). Organizational assessment: an overlooked approach to managing diversity and addressing racism in the workplace. *Journal of Diversity Management, 7,* 11-26.

Trochim, W. K. (1998). An evaluation of Michael Scriven's "minimalist theory: The least theory that practice requires." *American Journal of Evaluation, 19*(2), 243-249.

Ulrich, W. (1996). *A primer to critical systems heuristics for action researchers.* United Kingdom: Centre for Systems Studies.

Ulrich, W. (2000). Reflective practice in the civil society: The contribution of critically systemic thinking. *Reflective Practice 1*(2), pp. 247-268.

Ulrich, W. & Reynolds, M. (2010). Critical systems heuristics. *In Systems approaches to managing change: A practical guide,* M. Reynolds & S. Holwell, Eds., 243-292. London: Springer.

University of Minnesota. (February 9, 2017). Last anti-semitic incident part of a national trend. Retrieved from: https://campus-climate.umn.edu/content/latest-anti-semitic-incident-part-national-trend.

Washington Post Editorial Board. (September 30, 2016). Donald Trump is normalizing bigotry. *Washington Post,* retrieved from: https://www.washingtonpost.com/opinions/global-opinions/donald-trump-is-normalizing-bigotry/2016/09/30/9d754b2e-8731-11e6-a3ef-f35afb41797f_story.html?utm_term=.4a2860cd5d5b

Washtenaw Intermediate School District, Ann Arbor, Michigan. (2013). *Creating culturally proficient communities*: *The racial and economic justice project.*

Weiss, C. (1998). Have we learned anything new about the use of evaluation? *American Journal of Evaluation*, 19, 21-33.

Weiss, C. (2002). What to do until the random assigner comes. In F. Mosteller & R. Boruch (Eds.), *Evidence matters: Randomized trials in educational research.* Washington, DC: Brookings Institution.

Welborn, R., Downey, L., Hyjer Dyk, P., Monroe, P., Tyler-Mackey, C., & Worthy, S. L. (2016). Turning the tide on poverty: Documenting impacts through ripple effect mapping. *Community Development, 47*, 385-402.

Whitley, B. & Kite, M. (2010). *The psychology of prejudice and discrimination*, (2nd ed.). Belmont, CA: Wadsworth.

Wijeyesinghe, C., & Jackson, B. (2012). Introduction in C. Wijeyesinghe & B. Jackson, (Eds.), *New perspectives on racial identity development* (2nd ed.), (pp. 1-10). New York, NY: New York University Press.

Wilder Research (2011). *One Minneapolis community indicators report*. Retrieved online from https://www.wilder.org.

Wildhagen, T. (2012). How teachers and schools contribute to racial differences in the realization of academic potential. *Teachers College Record, 114,* 1-27.

Williams, B. (2004). *Complex adaptive systems.* Retrieved online from http://www.bobwilliams.co.nz/Systems_Resources_files/CASmaterial.pdf.

Williams, B. & Hummelbrunner, R. (2011). *Systems concepts in action: A practitioner's toolkit.* Stanford: Stanford University Press.

Williams, T., & Smith, M. (December 28, 2015). Cleveland officer will not face charges in Tamir Rice shooting death. *New York Times*, Retrieved online https://www.nytimes.com/2015/12/29/us/tamir-rice-police-shootiing-cleveland.html?_r=0.

Wilson, W. J. (1978). *The declining significance of race.* Chicago, IL: University of Chicago Press.

Wilson, W. J. (2012). *The truly disadvantaged: The inner city, the underclass, and public policy* (2nd Ed). Chicago, IL: University of Chicago Press.

Wolf-Branigin, M. (2013). *Using Complexity Theory for Research and Program Evaluation.* New York: Oxford University Press.

Yen, H. (December 13, 2012). White population to lose majority status in 2043. *Boston Globe* retrieved from http://www.bostonglobe.com/news/nation/2012/12/13/census-whites-longer-majority/cldoCAQfjIWT34hnhmXFAM/story.html.

Yuen, L. & Feshir, R. (September 26, 2016). Philando Castile police shooting probe complete, sent to prosecutors. *Minnesota Public Radio* retrieved from http://www.mprnews.org/story/2016/09/28/philando-castile-investigation-done-prosecutor-review.

Appendices

Study Participation Consent Form

Evaluating Diversity Training: A Grounded Theory Approach
to Understanding Challenges and Opportunities

You are invited to participate in a research study that will help us understand the challenges and opportunities associated with diversity training evaluation. You were selected as a possible participant because your institution conducts diversity training and you were identified as a staff person responsible for the training and/or training evaluation. We ask that you read this form and ask any questions you may have before agreeing to be in the study.

This study is being conducted by: Molly Illes, Organizational Leadership, Policy and Development department at the University of Minnesota.

Phone: 612-227-8864 Email: cern0001@umn.edu.

Advisor: Jean King, OLPD, University of Minnesota

Phone: 612-626-1614
Email: kingx004@umn.edu.

Background Information

The purpose of this study is advance existing literature in two key ways. First, it is focused on diversity training in higher educational settings exclusively versus general diversity training. Second, although there is a wealth of knowledge relating to developing and conducting training, literature focused specifically on the challenges and opportunities to effectively evaluating diversity training is limited

Procedures

This study will advance existing literature in two key ways. First, it is focused on diversity training in higher educational settings exclusively versus general diversity training. Second, although there is a wealth of knowledge relating to developing and conducting training, literature focused specifically on the challenges and opportunities to effectively evaluating diversity training is limited.

If you agree to be in this study, we would ask you to do the following things:

An interview conducted by phone with an anticipated duration of up to sixty (60) minutes. For the purpose of accuracy in conveying your information in this study, the researcher would like to audio record the interview. No one other than the P.I. will hear

or have access to the recording. Audio recorded files will be deleted within three months of completing the transcript.

In addition, the P.I. would like to review documents such as training materials, evaluation instruments and evaluation reports. The purpose of reviewing the documents is not to assess the impact of your programming, but to analyze how you evaluate your training program. No one other than the P.I. will have access to or review the documents you submit.

Example Interview Questions

- About when did your institution start doing diversity training?
- Describe a typical training session.
- How did you design your diversity training evaluation?

Risks and Benefits of being in the Study

The study has risks: First, diversity training is diversity training in higher education is highly political in nature and if you feel stating something may put you at risk, we can skip the question or stop the interview. All information will be kept confidential. Neither your institution nor you will be identified in the study.

There is no direct benefit to you for participating in this study. If you choose, you can receive a summary of the study. Your participation may help advance our knowledge of opportunities and challenges associated with diversity training evaluation.

Compensation:

You will not receive compensation for study participation.

Confidentiality:

The records of this study will be kept private. In any sort of report the P.I. might publish, she will not include any information that will make it possible to identify a subject or the associated institution. Research records will be stored securely and only the P.I. will have access to the records. Study data will be encrypted according to current University policy for protection of confidentiality. Only the P.I. will have access to the interview audio recording and any training documents you submit for review. All electronic documents will be stored on a secure server and password protected. Any paper copies of documents, transcripts or correspondence will be stored in a secure, locked location. Once the study is completed, all electronic files will be deleted and paper documents will be shredded.

Voluntary Nature of the Study:

Participation in this study is voluntary. Your decision whether or not to participate will not affect your current or future relations with the University of Minnesota. If you decide to participate, you are free to not answer any question or withdraw at any time without affecting those relationships.

Contacts and Questions:

The researcher conducting this study is: Molly Illes. You may ask any questions you have now. If you have questions later, **you are encouraged** to contact the P.I. at the University of Minnesota, 612-227-8864 or cern0001@umn.edu or the study advisor, Jean King, at (612) 626-1614 or by email at kingx004@umn.edu.

If you have any questions or concerns regarding this study and would like to talk to someone other than the researcher(s), **you are encouraged** to contact the Research Subjects' Advocate Line, D528 Mayo, 420 Delaware St. Southeast, Minneapolis, Minnesota 55455; (612) 625-1650.

You will be given a copy of this information to keep for your records.

Statement of Consent:

I have read the above information. I have asked questions and have received answers. I consent to participate in the study.

Signature:____________________________ Date: __________________
(please type your name, date and return to Molly Illes via email)

Signature of Investigator:________________ Date: __________________

Thank you for participating in this study.

Diversity Certificate Training (DCT) Practitioner Interview Instrument

Thank you for your willingness to meet with me today. The purpose of this interview is to gather a description of your current diversity training program and to understand how you evaluate your training. I am interviewing several other diversity training practitioners. The information you provide will advance our understanding of challenges and opportunities relating to diversity training evaluation.

The interview will last approximately 60 minutes. There is no compensation for participating in this study. Although there is no direct risk in participating, some of the questions may be challenging to answer because of the political nature of diversity training. If any question makes you uncomfortable or you are unable to answer it, we can skip it. All of the information you share with me will remain confidential. I will not include information that would reveal you or your organization.

[Review consent form together.] For the purpose of transcription, I would like to record today's conversation. The audio recording will be stored in a password-protected location and only accessed by me. Once the study is complete, I will delete the recording. I will not transcribe any names or identifying information you may share during your interview. Do I have your permission to record our conversation today? Thank you.

Let's get started. First, I'd like to start by asking a few questions about your background.

1. About how long have you worked at [organization name]?
    a. Have you always worked in this capacity within this organization?
    b. [If less than five years] What did you do prior to coming to this organization?

Now I'd like to ask you some general questions about your diversity training.

2. If you were to describe your training to someone who's never attended, what would you say?
    a. When did you (or your organization) first develop this training?
    b. What do you believe are the key motivations behind offering this training?
    c. About when did your institution start doing diversity training?
    d. What have been some of the topics your training addresses?
    e. Describe a typical training session.
        i. (Duration, participant demographics, frequency of offering, multi-session versus one, time of day offered, mandatory versus voluntary).
    f. How is your training funded? Who funds it?

3. What are the primary goal(s) for this training program?
    a. What are common outcomes you've identified from your training?
    b. In what way, if any, did you develop your training goals in relation to organizational goals?

Now I'd like you to walk me through the process you use to evaluate your diversity training program.

How did you design your evaluation?

c. Who designed the evaluation?
d. Did you use any type of model to guide your evaluation? If so, describe the model.
e. Who typically conducts the evaluation? Analyzes the data? Writes the report?
f. [If evaluator is different from the person interviewed] In what ways, if any, do/did you work with [person named in a] to develop the evaluation?
    i. Do you have a logic model/theory of change for your training? If so, do you review logic model when creating the evaluation?

4. Do you go back to training participants over time to see how they are doing with the skills, learning, and behavior from your training? If so, how do you do this?

5. What do you think works well with your evaluation process?

6. What, if any, general training evaluation challenges do you have at [organization]?

7. Let's consider for a minute that the challenges you mentioned no longer exist. If you could design and conduct any type of evaluation of your diversity training, what would you do?

8. How do you know that your training is effective?

9. You mentioned a few minutes ago that this training started in [reference timing from 2a]. Has the way you deliver the training changed since you first started doing it?
    a. If yes, how did you decide to make the changes?

10. What do you do with the evaluation results?
    a. Within your organization, whom do you share your results with?
    b. Outside of your organization, whom do you share your results with?
    c. Can you describe any ways in which the results from your training are used within your department or institution? External to your institution?

11. In question two, you mentioned that [funder] funds your programming. Do they have reporting guidelines you need to adhere to? If so, what are they?

12. I have asked a lot of questions. However, are there any things you haven't shared today that you'd like to mention?
    a. Do you have any questions for me?

Thank you again for taking the time to speak with me today. I appreciate your dedication to diversity training and your willingness to participate in this study.