

Meeting Learning Objectives in a Multicultural Leadership Course: Using Assessment to Inform Pedagogical Practice

Leigh E. Fine

Assistant Professor
Kansas State University

Chance Lee

Instructor
Kansas State University

Abstract

It is becoming increasingly important for leaders to recognize and develop the skills needed to interact with diverse others (Karim, 2003). To this end, several leadership programs in American colleges and universities offer courses that explore the practice of multicultural leadership (Brungardt, Greenleaf, Brungardt, & Arensdorf, 2006); our institution's leadership program is no different in its goals of preparing culturally competent leaders. The multicultural leadership course in the Staley School of Leadership Studies at Kansas State University has two learning outcomes: "Understand the impact of cultural identity (Note that this manuscript uses the terms *social identity* and *cultural identity* interchangeably, which is common practice (Wren, 2002), life experiences, and world views on leadership relationships as it relates to privilege and inclusion," and, "Practice inclusive leadership through advocacy for social change." Such goals are desirable, but also lofty – how can we know if our course has had any effect on student learning toward these objectives?

Theory-informed assessment of the effects of multicultural leadership courses can meet many pedagogical needs while answering the demands of multiple stakeholders (Fink, 2006; Wholey, Hatry, & Newcomer, 2004). Knowing what changes occur in students' attitudes or behaviors during a multicultural leadership course would permit faculty, staff, and other practitioners to evaluate whether the experience may be able to meet desired learning goals. Future courses could be recalibrated to address deficiencies or strengthen successes. Assessment data also facilitates the sharing of the course's effects to administrators, prospective students, parents, or donors in efforts to communicate the value of intentional, curricular student leadership development.

Beyond these larger-scale needs for assessment, we have an interest in knowing to what degree the course contributes to our students' engagement and leadership development. Assessment can help us to identify what students are learning as a result of the course. Further, we hope that sharing our assessment structure, analysis, and results can help other leadership educators strengthen a culture of assessment in their own programs, particularly concerning multicultural leadership development.

To determine if we are making progress in meeting our learning outcomes, we developed an assessment tool to determine students' change over the course of the term across two theoretical constructs: diversity attitudes, which encapsulates the first learning goal of recognizing linkages between worldview and leadership (Bennett, 2004; Karim, 2003; Komives, Longersbeam, Owen, Mainella, & Osteen, 2006); and changes in leadership behaviors and attitudes, which speaks to the second learning goal of catalyzing social change-based leadership (Komives, Wagner, & Associates, 2009).

We use a pre-test / post-test design to compare students' reported values at the beginning of the semester to their values upon completion of the class (Fink, 2006). By comparing pre-test and post-test responses to questions from the assessment instrument, we demonstrate in what ways our course might influence student leadership development and multicultural competence, as well as areas we might wish to consider fortifying in future iterations of the course to better meet learning objectives. We then reflect on the effect the results will have on learning outcomes, curricular development, and the instrument itself, illustrating the cyclical relationship between theory, practice, and assessment in multicultural leadership education.

We have two aims here. Our first is to articulate what gains may be made by students who engage in our course to determine what exactly is gained by students who participate in such experiences. We believe such courses have educational value, but needed to find a way to articulate our course's success in helping students meet our stated learning objectives. Our second aim is to illustrate both the necessity, rewards, and limitations of engaging in critical engagement with the assessment-practice cycle (Maki, 2002). Here, we reflect on what changes we or others could make to our course to better achieve our desired goals, as well as how we might continue to strengthen our assessment efforts. We hope that our analysis here provides some starting points to assess multicultural leadership programs' efficacy in meeting student learning outcomes.

Literature Review

Multicultural, cross-cultural, and inclusive leadership. Those who study and practice leadership recognize the importance of cultural competence (Karim, 2003). Seemiller's (2013) leadership competencies identify constructs such as empathy, social justice, and inclusion as key, encouraging those who wish to learn more about leadership and develop their leadership skills to be able to do so within varying cultural and interpersonal contexts. Likewise, Komives, Wagner and Associates (2009) recognize that collaborating with others "within diverse groups requires each of us to enhance our ability to work effectively with, listen to, and develop trust with people who are different than us" (p. 213).

Global learning is also of interest to leadership scholarship, as it is generally considered to encapsulate intercultural, civic, and critical thinking competencies (Hartman, Lough, Toms, & Reynolds, 2015). Increasingly, scholars are advancing that global learning can be inclusive of domestic and international, as well as immersive and classroom-based, learning (Sobania & Braskamp, 2009). Global learning may serve to address the civic mission of institutions of higher education (Gerstl-Pepin & Aiken, 2012; Hovland, 2014) in a rapidly flattening global society (Friedman, 2007; Pope, Mueller, & Reynolds, 2009).

Although cross-cultural leadership and cross-national leadership could be conflated, other dimensions of cultural difference are gaining attention as meaningful determinants of power and privilege within leadership relationships (Pope, Mueller, & Reynolds, 2009). For instance, Hartmann (2015) argues that Americans generally value diversity and are energized by the potential cross-racial relationships have to transform notions of equality, citizenship, and belonging, all of which are related to the practice of effective leadership (Chrislip & O'Malley, 2013). Sexual identity, too, is another social identity that scholars increasingly see as a meaningful dimension of difference that affects leadership relationships (Lugg & Tooms, 2010).

The ability to assess student attitudes toward a variety of groups across multiple dimensions of cultural difference could provide a more comprehensive understanding of whether we are meeting our two learning objectives. First, analysis may indicate where students are growing in their ability to engage with others, which speaks to our learning outcome of understanding social identities, privilege, and inclusion, particularly across what dimensions of difference students perceive gaps (Komives et al., 2006). Second, because the skills to create change with a diverse constituency is a leadership imperative – as our second learning outcome states – it is all the more integral to be able to determine what experiences contribute to this learning outcome for those who endeavor to develop these skills with their students.

Assessment: The theory and practice cycle. Assessing student learning allows instructors to determine if a course is accomplishing desired learning outcomes; to measure the level at which students are performing; and to answer the data demands of various stakeholders such as donors, administrators, parents, and students (Kuh & Ewell, 2010; Wholey, Hatry, & Newcomer, 2004). Not only should courses include assessment as a means of addressing these needs, but it should also use the findings from assessment to constantly improve the course's ability to meet outcomes (Maki, 2002).

Measuring intercultural competence: Social distance theory. Our first learning outcome for the course holds that students should be able to grow in their understanding of the forces of identity, inclusion, and privilege. As the world continues to become metaphorically smaller with the advent of technologies that permit increased intercultural contact (Friedman, 2007), providing emerging leaders with the skills and capabilities necessary to create strong relationships with diverse people is a necessity for leadership educators.

Although there are several means of evaluating individuals' relative comfort or antipathy toward various social groups (Correll et al., 2010; IDI, 2015), existing instruments tend to provide a generalized picture of respondents' openness toward diversity. Because of this, it can be more difficult to pinpoint which student attitudes are shifting – toward which groups or on which topics. For instance, it is possible that a student may be open to religious differences, but still harbor racist attitudes – a nuance such instruments may not be able to capture. Therefore, we created our own instrument, using two bodies of theory to develop its items: social distance theory and leadership development theories.

Bogardus (1947) developed a means of determining survey respondents' relative comfort with various racial groups, which serves as the basis for many of our instrument's items related to attitudes toward people of different social identities. Bogardus's respondents rank their

relative comfort with a variety of racial and ethnic groups on a Likert (1932) scale. Social distance continues to be a generally-accepted method of analyzing comfort toward various social groups, not just as a function of race (Correll et al., 2010; Parrillo & Donoghue, 2005; Wark & Gaillher, 2007). If one of our learning objectives is to encourage the examination of behaviors, attitudes, and beliefs in such a way that permits students to engage meaningfully with diverse groups of people (Karim, 2003), social distance measures are one means of determining if any shifts have occurred for students along specific dimensions of difference through the course. Although a strict reading of our first learning objective could argue that understanding difference is not necessarily concomitant with more positive attitudes, understanding and empathy are intimately linked in the practice of leadership (Komives, Lucas, & McMahon, 2013; Komives, Wagner, & Associates, 2009). Therefore, we feel justified in using social distance as a measure of students' ability to understand the impact of worldview, identity, privilege, and inclusion, as it is an indicator of the ability to recognize others' realities as valid.

Measuring multicultural leadership skills: Social change. In addition to encouraging students' intercultural competence, our cross-cultural leadership course focuses on leadership development and social change in its second learning outcome. Particularly for programs that focus on leadership as relationships (Rost, 1991), thinking about how to interact with others to foster meaningful change is imperative. Both the course and the assessment instrument use the Social Change Model of Leadership (Komives, Wagner, & Associates, 2009) to articulate the value of including diverse others in the practice of leadership. This model examines leadership as a phenomenon that occurs at three levels: individual, group, and society. The Model encourages leaders to think about how to engage in practices that advocate for a common good within each level and in concert with the other levels. Given the course's particular focus on advocacy based on issues of inclusion and culture, several items we developed for the assessment instrument examine students' willingness to engage in social change behaviors to make the world more inclusive, such as building cross-cultural coalitions or participating in democratic processes.

Method

Sample: The multicultural leadership course. We explore how our cross-cultural leadership course meets learning outcomes related both to multicultural competence and social change leadership through the development of a survey instrument. Data come from 500 students enrolled in a multicultural leadership course in the Staley School of Leadership Studies at Kansas State University; the course is nested in an interdisciplinary, non-selective undergraduate minor program. This sophomore-level course's explicit focus is on developing students' comfort with various social identity groups while developing the leadership skills necessary to engage with a diverse population. Through intense dialogue and learning communities, which are high-impact learning practices (Kuh, 2008), we hope that students will grow in their appreciation for others' worldviews while expressing a desire for a more inclusive social world. The course relies on a modest level of theoretical content. Bennett's (2004) model of ethnocentrism, the Social Change Model of Leadership (Komives, Wagner, & Associates, 2009), the social construction of culture and identity (Berger & Luckmann, 1966; Fine, 2015), and Cullen's (2004) Ten Core Concepts are generally covered within the first few weeks of the course. Most of this content is covered very early on in the semester and then reinforced through

continually-utilized language as students engage with contentious issues and practical application throughout the rest of the course.

The bulk of the work during class time provides students with scenarios, conversations, experiences, or current examples that allow them to apply these concepts to their practice of leadership. One long-standing component of the course involves sharing class time with international students on campus and panels of people from diverse identity groups (Fine, 2015). Students also engage in discussion about contemporary events related to leadership and multiculturalism; these conversations encourage students to consider worldviews other than their own, to challenge one another's interpretations, to evaluate the role privilege plays in constituting and reifying power structures, and to consider how their actions contribute toward larger social patterns.

Assignments for the course encourage students to explore both cultural lenses and how to advocate for inclusive social change. Historically, the two pivotal assignments are a cultural immersion assignment, called a Cultural Plunge (Nieto, 2006), and a summative research project that examines how policy affects civic engagement or generalized well-being based on dimensions of cultural difference and individual experience (Dallimore, Rochefort, & Simonelli, 2010). The summative research project tasks students with developing a social change strategy to implement based on the data they collected through research and meaningful engagement with diverse others. Through these papers and additional reflective journal entries, we endeavor to challenge students to consider linkages between identity, the sociocultural, and the practice of leadership (Ford, Harding, & Learmonth, 2008; Karim, 2003).

The data here represent six terms' worth of responses from over two and a half years: from spring 2014 to summer 2016. Students were asked to complete the pre-test assessment instrument within the first two weeks of the course and the post-test within the last two weeks of the course. Data were gathered with pencil-and-paper until the spring 2015 term, at which time the survey was ported over to Qualtrics in an electronic format. Our assessment tool collects both quantitative and discursive data: the pre-test and post-test asked for respondents' demographic information, responses to open-ended questions about various multicultural leadership concepts, and responses to Likert scale questions about attitudes toward questions related to cross-cultural leadership practice. Here, we present the quantitative data.

Not all cases are used in each analysis. The primary limitation of these data is the lack of matched pre-test / post-test data for all students across all items of interest. This is due to students declining to answer a particular item, students forgetting to turn in assessment data for either the pre-test or post-test, or instructors' varied methods in gathering data using the survey instrument (return rates were improved for the instructors who set aside class time to gather the data). Additionally, some questions were added between waves of data collection as the survey instrument was refined, meaning that students who took the instrument earlier on in its development were not asked all the questions that students taking the most recent iterations were. The item with the greatest number of paired responses is (N = 500), whereas the lowest is (N = 177). Reliability and composite scale analyses are calculated using only those surveys that were complete for all items analyzed here (N = 159), while individual item analyses use all available

paired cases to maximize statistical power. Sample sizes for each analysis is noted on the results table.

Assessment Development, Data, and Analysis

Developing the survey instrument. Data for analysis come from questions we developed that are intended to measure social distance and social change, the two theoretical constructs of interest given our learning objectives. Table 1 lists the items we developed regarding three constructs of interest: leadership behaviors and social change, attitudes toward LGBT people, and attitudes regarding racial diversity, immigration, and nationalism. The latter two constructs represent the course's first outcome of promoting students' understanding of inclusion, diversity, and worldview; the first construct speaks to the course's second outcome of encouraging students to employ social change leadership and develop its associated skills.

The items employ a Likert scale (Likert, 1932) to gather information on the dependent variables. This practice is common in research that examines attitudes (Jamieson, 2004) and is in keeping with Bogardus's social distance measures. Students note to what degree they agreed with the statements posed on a scale from one to five, with one being "strongly disagree," three being "neither agree nor disagree," and five being "strongly agree." We then calculate mean values for each question by averaging the responses for all paired pre-test and post-test cases. Although it may be considered problematic to treat Likert scale responses as ratio variables when they are in fact ordinal (Jamieson, 2004), we believe that the analysis here still presents meaningful results that are interpretable as changes in students' reported values that is in keeping with the use of Likert scales in statistical analyses (Knapp, 1990).

We developed our own survey instrument because we felt that no existing multicultural inventory would best reflect our course's efficacy in meeting learning objectives. Because we are using our own survey, it is integral to determine if it yields reliable results. We address reliability by grouping together variables that are measuring the same rough construct. This grouping also addresses the fact that can be problematic to measure Likert scale items in isolation; creating composite variables that average the results across many related Likert scale items is a sounder analytical strategy (Gleim & Gleim, 2003).

Table 1. Survey Items Included in Each Construct

Construct	Item Prompts	Prompt Reverse Coded	Number of Items	Bonferroni Threshold			
1: Leadership Behaviors and Social Change	I am capable of changing the world.		12	p < 0.004			
	Including all types of people is important to me.						
	I would protest if I thought a law were unjust.						
	I want to make the world more inclusive.						
	It is hard for me to talk with people who disagree with me.	Yes					
	I like talking with others who are different from me.						
	A good leader rarely needs advice.	Yes					
	I am a leader.						
	When I argue with someone, my answer is usually the right answer.	Yes					
	A good leader is someone who listens.						
2: Attitudes Toward LGBT People	Being gay is morally wrong.	Yes	8	p < 0.006			
	Gay people make me uncomfortable.	Yes					
	Sometimes people need to explore what gender they are.						
	Gay people flaunt their sexuality.	Yes					
	I wish I knew more gay, lesbian, bisexual, or transgender people.						
	Same-sex marriage should be legal.						
	If people want to change their sex, medical insurance should cover the costs incurred.						
	I would be uncomfortable if someone of the same sex told me they found me attractive.	Yes					
	3: Attitudes Regarding Racial Diversity, Immigration, and Nationalism	We focus too much on race in American society.			Yes	9	p < 0.006
		Affirmative Action is unfair.			Yes		
Groups like the Black Student Union create more racial divisions.		Yes					
Racial tensions can best be handled by minimizing racial differences.		Yes					
We have overcome racism in the 21st century.		Yes					
Immigrants make this country stronger.							
Every country wishes they were the United States.		Yes					
People of color are treated just as well as (if not better than) white people in the United States.		Yes					
English should be the national language of the United States.		Yes					

Using social distance theory and the Social Change Model as starting points, we developed a battery of questions to measure students' attitudes across these constructs. To determine our survey's reliability, we calculate Cronbach's alpha values for the groupings of items we developed using data from the pre-test and post-test instruments. Results in table 2

show that alpha values for all three constructed variables are $\alpha > 0.7$, the generally-accepted minimum threshold for reliability (Gleim & Gleim, 2003). Our survey instrument also collects data on items that we hope can be used to construct scales related to religious tolerance, empathy, and civic engagement, which are other constructs of interest given our learning objectives. However, sample sizes are yet too small or items too few to yield alpha values of $\alpha > 0.7$; therefore, these results are not presented here.

Table 2. Cronbach's Alpha Values for Survey Item Constructs

Construct	α
1: Leadership Behaviors and Social Change	0.725
2: Attitudes Toward LGBT People	0.856
3: Attitudes Regarding Racial Diversity, Immigration, and Nationalism	0.705
N	159

Analytical strategy. We measure change in students' responses between the pre-test and post-test instruments across individual items and the composite scales. Because we are comparing the population mean at two different time periods, we use t-tests to determine if the differences between the mean values for each item and each composite scale are statistically significant (Welch, 1947). Even if we have a hypothesis as to which direction the change in students' responses will take, we choose to use more conservative two-tailed tests for significance in testing that the mean difference on all items is not equal to zero.

As we are comparing t-tests across individual items that measure the same construct, familywise error needs to be taken into account for these t-tests. The most common means of addressing familywise error is employing the more conservative Bonferroni correction when determining t-tests' significance for individual items (Bender & Lange, 2001). The Bonferroni correction sets the threshold for significance at the generally-accepted threshold divided by the number of items in the scale: in other words, ($p < [0.05 / i]$), where i represents the number of items grouped to form each scale. Table 3 lists the Bonferroni thresholds for each scale and its individual items, based on the number of items that comprise the scale.

Although the survey instrument did also collect background demographic data, such as gender, race, age, and major, we do not use analyses that might demonstrate inter-group differences (e.g., ANOVA, ANCOVA) to present a general overview of to what extent the course's learning objectives are being met. Further, as our work is descriptive and not predictive, we choose t-tests instead of regression techniques (e.g., OLS). Therefore, we believe t-tests to be sufficient for the research question here. Future work could examine how background characteristics might affect changes in students' responses, how changes across items predict future leadership behaviors.

Results

Individual item t-tests. Table 3 shows the results of t-test comparisons between pre-test and post-test scores across a battery of questions from the assessment instrument, grouped by composite scale.

Table 3. T-test Results Comparing Items Across Pre-Test / Post-Test Assessment Instruments

Item	N	\bar{x}_{pre}	\bar{x}_{post}	$\Delta\bar{x}$		s.e.
<i>1: Leadership Behaviors and Social Change, threshold: $p < 0.004$</i>						
I am capable of changing the world.	347	4.061	4.135	0.075		0.045
Including all types of people is important to me.	498	4.349	4.402	0.052		0.034
I would protest if I thought a law were unjust.	346	3.266	3.324	0.058		0.053
I want to make the world more inclusive.	343	4.061	4.207	0.146	*	0.049
It is hard for me to talk with people who disagree with me.	342	2.535	2.520	-0.015		0.060
I like talking with others who are different from me.	347	4.110	4.207	0.098		0.035
A good leader rarely needs advice.	180	1.439	1.322	-0.117		0.059
I am a leader.	178	4.157	4.230	0.073		0.050
When I argue with someone, my answer is usually the right answer.	179	2.922	2.810	-0.112		0.057
A good leader is someone who listens.	178	4.640	4.680	0.039		0.046
Good leadership means taking charge.	178	3.500	3.281	-0.219	*	0.074
If I'm right, there's little point in listening to others' arguments.	178	1.893	1.742	-0.152		0.058
<i>2: Attitudes toward LGBT People, threshold: $p < 0.006$</i>						
Being gay is morally wrong.	498	1.978	1.860	-0.117	*	0.039
Gay people make me uncomfortable.	500	1.902	1.838	-0.064		0.039
Sometimes people need to explore what gender they are.	497	2.928	3.155	0.227	**	0.047
Gay people flaunt their sexuality.	492	2.528	2.402	-0.126	*	0.042
I wish I knew more gay, lesbian, bisexual, or transgender people.	344	2.964	3.221	0.257	**	0.050
Same-sex marriage should be legal.	179	3.704	3.911	0.207	**	0.061
If people want to change their sex, medical insurance should cover the costs incurred.	177	2.373	2.684	0.311	**	0.068
I would be uncomfortable if someone of the same sex told me they found me attractive.	178	2.944	2.736	-0.208		0.085
<i>3: Attitudes Regarding Racial Diversity, Immigration, and Nationalism, threshold: $p < 0.006$</i>						
We focus too much on race in American society.	342	3.515	3.281	-0.234	**	0.062
Affirmative Action is unfair.	336	2.869	2.792	-0.077		0.053
Groups like the Black Student Union create more racial divisions.	497	2.911	2.853	-0.058		0.049
Racial tensions can best be handled by minimizing racial differences.	492	3.122	2.852	-0.270	**	0.055
We have overcome racism in the 21st century.	346	2.039	1.783	-0.256	**	0.056
Immigrants make this country stronger.	344	3.407	3.573	0.048	**	0.048
Every country wishes they were the United States.	177	2.062	1.859	-0.203	*	0.069
People of color are treated just as well as (if not better than) white people in the United States.	177	2.469	2.028	-0.441	**	0.062
English should be the national language of the United States.	177	3.746	3.537	-0.209	*	0.062

* = $p < \text{Bonferroni threshold for construct group}$, ** = $p < 0.001$, all two-tailed.

Regarding leadership behaviors, items in this scale that were significant using a Bonferroni correction are, “I want to make the world more inclusive” ($\Delta\bar{x} = 0.146, p < 0.004$) and, “Good leadership means taking charge” ($\Delta\bar{x} = -0.219, p < 0.004$). However, items in this scale related to more direct forms of social action did not change at a statistically significant level between the pre-test and post-test, such as, “I would protest if I found a law unjust” ($\Delta\bar{x} = 0.058, n.s.$). Further, some items did not register a statistically-significant change because students entered and left the course believing they had the capacity and desire to practice inclusive leadership. For instance, the item, “Including all types of people is important to me,” demonstrates high agreement from students during both the pre-test ($\bar{x}_{pre} = 4.349$) and post-test ($\bar{x}_{post} = 4.402$).

Students’ attitudes toward sexual and gender identity differences changed between the pre-test and post-test. Six of the eight individual items that comprise the scale were found to be significant in a direction that indicates more positive attitudes toward sexual minorities. Only two items did not demonstrate a level of change at a statistically significant level between the pre-test and post-test. “Gay people make me feel uncomfortable,” was not found to be significant as students report relatively high levels of disagreement with the item both during the pre-test ($\bar{x}_{pre} = 1.902$) and post-test ($\bar{x}_{post} = 1.838$); “I would be uncomfortable if someone of the same sex found me attractive” exhibits a change between iterations of the instrument, but not at a level that is statistically significant with the Bonferroni correction employed ($\Delta\bar{x} = -0.208, n.s.$).

Seven of nine items related to attitudes regarding race, immigration, and nationalism were found to be statistically significant, even after employing a Bonferroni correction. All changes are in a direction indicating more openness toward engaging with differences across racial dimensions: “We focus too much on race in American society” ($\Delta\bar{x} = -0.234, p < 0.001$), “Racial tensions can best be handled by minimizing racial differences” ($\Delta\bar{x} = -0.270, p < 0.001$), “We have overcome racism in the 21st century” ($\Delta\bar{x} = -0.256, p < 0.001$), “Immigrants make this country stronger” ($\Delta\bar{x} = 0.048, p < 0.001$), “Every country wishes they were the United States” ($\Delta\bar{x} = -0.203, p < 0.006$), “People of color are treated just as well as (if not better than) white people in the United States” ($\Delta\bar{x} = -0.441, p < 0.001$), and, “English should be the national language of the United States” ($\Delta\bar{x} = -0.209, p < 0.006$). The items, “Affirmative Action is unfair,” and, “Groups like the Black Student Union create more racial divisions,” did not exhibit change that is statistically significant at the Bonferroni threshold.

Composite scales. To avoid overstating any change in student attitudes by using single Likert scale questions, we sum the differences in students’ responses across all individual items in the scale, then use a t-test to determine if this sum is not equal to zero at a statistically significant level (Gleim & Gleim, 2003).

Table 4. T-test Results Comparing Constructed Scales Across Pre-Test / Post-Test Assessment Instruments

Construct	$\Delta\bar{x}$	t-test:	
		$\Delta\bar{x} \neq 0$	s.e.
1: Leadership Behaviors and Social Change	1.151	***	0.304
2: Attitudes Toward LGBT People	1.497	***	0.259
3: Attitudes Regarding Racial Diversity, Immigration, and Nationalism	2.302	***	0.299
N	159		

* = $p < 0.05$, ** = $p < 0.01$, *** = $p < 0.001$, all two-tailed.

Table 4 provides the results of these t-tests. Data show that the scales related to leadership and social change ($\Delta\bar{x} = 1.151$, $p < 0.001$); attitudes toward LGBT people ($\Delta\bar{x} = 1.497$, $p < 0.001$); and attitudes toward racial differences, immigration, and nationalism ($\Delta\bar{x} = 2.302$, $p < 0.001$) all change between pre-test and post-test at statistically-significant levels and in a positive direction.

Discussion & Conclusion

The first goal of this work is to determine if the course meets its stated learning objectives, based on findings from our assessment tool. Results give voice to the patterns we see in our students' attitudes and learning throughout the course while permitting us to explore how assessment of such can inform pedagogical practice. Our analyses show our students do indeed exhibit a good deal of growth over the semester while in the course in terms of understanding identity and advocating for social change, indicating the course may provide a net benefit. With regard to social distance from various social groups, which is how we operationalized the class's first learning objective, student attitudes toward various social identity groups do change between the pre-test and post-test. This may be due in large part to how the course curriculum approaches conversations around diverse social identities through the mechanisms of authentic storytelling (Fine, 2015), making these concepts more concrete and bringing them into the classroom. If the goal of the class is to lessen the social distance students express toward various social groups, finding means for people to share their stories may be the most effective means of accomplishing this goal (Wheatley, 2001).

At the same time, the assessment tool also identified gaps in student learning. For instance, although the composite variable related to change in students' self-reported leadership and social change behaviors was statistically significant and positive, many of the individual items shift very little largely because students already think of themselves as capable of exercising these skills. Despite such concerns, it appears the experience over the course of the semester leads to positive gains in students' ability to transcend (some) cultural borders, further supporting the value of multicultural educational experience in developing socially-responsible future leaders (Hartman et al., 2015; Komives et al., 2006).

Directions for curriculum revision. Our second goal here is to demonstrate how meaningful assessment of programs naturally leads to reflection on what is learned to promote curricular change that better meets learning outcomes. Appropriate application of the assessment cycle (Maki, 2002) necessitates that we examine how we will adapt our course after analyzing

the data. It is our hope that other institutions are likewise inspired to critically analyze multicultural leadership curricula, to broaden our understanding of what makes students multicultural leaders, and to provide new ways of assessing leadership students' learning. Given our findings, we plan on making the following adjustments to the curriculum of the multicultural leadership class and to future iterations of the assessment tool.

Pinpoint how the course influences student beliefs and attitudes to bolster students' leadership ability and lessen social distance. Findings here show that students are generally more aware of problems surrounding cultural divisions in leadership contexts. However, because these are summative, indirect assessment data, we cannot speak confidently to what aspect(s) of the course are exerting the greatest influence on student learning: collaborative work, hearing others' stories, engaging in reflective writing, or advocating for social change through the culminating research project. Conducting deeper formative, not just summative, assessment of our course may provide a means of identifying specific interventions that are particularly successful at shifting student attitudes and providing deep learning related to social change and advocacy (William & Black, 1996).

Critically examine the role of both leadership development and multicultural competence in the course. It can be a difficult pedagogical balancing act to create a curriculum that fosters both students' leadership development and multicultural competence. At times, we feel as though meeting both demands is difficult, given time constraints, the varied levels of pre-existing multicultural competence with which students may enter the class, and the expectations colleagues who teach other leadership courses have regarding our course's outcomes. If students report that they are more aware of their own ethnocentrism and privilege as a result of the course, a next logical question is how we as instructors might create an environment where students can directly leverage this knowledge to apply it to their beliefs about and practice of leadership, particularly as our students already report high efficacy in their abilities to exercise social change. The data here represent the genesis of ongoing conversations about what the student learning outcomes of the course are and if they are meeting students' needs as they enter a diverse and changing world – which is what good assessment should do (Oakleaf, 2009).

Continue to strengthen this assessment tool while promoting longitudinal assessment across the program. Employing measures of leadership that already exist and have been verified in future iterations of the survey can strengthen our tool's validity and give us access to new vocabularies to articulate how students are affected by the course. Measures of global learning such as Braskamp, Braskamp, and Engberg's (2014) Global Perspective Inventory or the AAC&U (2013) Global Learning rubric can provide valuable insight for future development of our assessment. We plan to include Kane et al.'s (2002) leadership efficacy measures and adaptations of IUPUI's Center for Service and Learning's (2015) Civic Minded Graduate measures of civic engagement, which may highlight ways in which our course does or does not affect other dimensions of leadership competence. Measures of social distance can also be supported by determining if students' generalized empathy changed during the course, as empathy is often cited as an imperative skill for practicing cross-cultural leadership that speaks to our first learning outcome (Komives, Lucas, & McMahan, 2013); to that end, we hope to include Davis's (1983) Interpersonal Reactivity Index measures as another means of determining how the course may facilitate students' development of cross-cultural leadership skills. These

items may provide new means of evaluating changes in student learning and leadership behavior that are used more broadly by other scholar-practitioners.

Additionally, we see room for revision across the constructs we already measure here. Before conducting these analyses, we had hoped we would be able to provide a comprehensive evaluation of our students' social distance toward myriad social identities. Our results here show that our survey instrument could use further revision if we wish to make such claims. Although we are able to tease out change in student attitudes as they relate to racial and LGBT issues, low reliability coefficients prevent us from making similar claims about students' comfort toward those of other social identities. Including more items in future iterations of the survey related to students' comfort with these groups could permit us to conduct similar analyses across these dimensions of difference.

Continue data collection, analysis, and dispersal. Gathering assessment data from our students gives us a new set of tools with which to further the field's collective goals of developing multicultural-competent leaders. Fortunately, these data provide future research opportunities. As mentioned above, the demographic data we collect alongside these scales could be used to determine whether particular subgroups of students show more change across various scale items than others. Because we hope to collect data on future civic engagement endeavors and empathy, it is possible that predictive analyses of future data could yield findings of interest as well.

Share our story. Ideally, assessment provides educators a way to share information with outside invested parties about what change occurs as a result of various educational interventions (Upcraft & Schuh, 1996; Wholey et al., 2004). Thanks to these data, we can now say that our students are more likely to recognize the importance of inclusion and express higher confidence in their social change leadership capabilities at the end of the course than they were at its start. This is a powerful story that has the potential to attract more students, to inspire donors to support our work, and to communicate to the greater educational community the need for developing multicultural competence in young leaders.

Limitations and future refinements of the assessment tool. The pre-test / post-test data collection design has limitations (Rosch & Schwartz, 2009). Students may be able to decipher what response the instructors hope will be selected, which can influence their responses. Also, a summative assessment design makes it difficult to pinpoint precisely what components of the course affected the students' change in responses – or if the course was responsible for the changes at all (William & Black, 1996). Future data collection could use formative assessment that examines the power of a particular intervention at key points throughout the course could complement the pre-test / post-test design and mitigate these concerns.

Survey response bias could emerge from what Ross and Schwartz (2009) call the horizon effect. A common issue in analyzing college students' leadership development as they take leadership courses is that their frame of reference shifts as their skills, abilities, and self-concept also change (Taylor, Russ-Eft, & Taylor, 2009). For instance, although the effect is not statistically significant, it might be distressing to us that the course appears to lower students'

desire to protest against unjust laws. However, if the horizon effect is present, this may not be because the course changed students' attitudes regarding protest as a valid form of civic engagement and social change action; rather, it may be because our course has caused students to be more honest with themselves (and us) about their willingness to protest. Triangulation or revision of items may be ways to adjust the survey instrument to lessen horizon effect biases in future iterations.

Perhaps our largest assumption is that our course is the primary contributor toward changing students' attitudes toward various cultural groups, which is a hefty assumption indeed. Throughout this work, we are careful to mention that students' attitudes change without attributing causality (solely to) our class. There are many college classes or co-curricular activities that can expand students' ability to engage with diverse others (Komives et al., 2006). Further, although our minor is interdisciplinary, there may be selection bias in terms of what student elects to take a course on cross-cultural leadership: a student who may be more open to developing her or his skills in this arena in the first place and, therefore, may show disproportionate growth across the dimensions we measure (Rosch & Collier, 2013). Ideally, we would pair our analysis of our students with analysis from pre-test / post-tests from a randomized control comprised of students from our campus who are not enrolled in the course to see what growth exists in our students as compared to the general student population over a semester's time. We are hopeful that our future analyses will be able to better disentangle these contextual effects to determine the true effect of the course itself.

Conclusion. The analysis here assesses students' multicultural competence and leadership development as a result of taking a course that focuses on cross-cultural leadership. Although our course succeeds in changing some students' responses in ways that align with stated learning outcomes, the results also ask us to consider if the course is accomplishing these goals to a satisfactory extent. We look at this work as an example of the assessment and practice cycle, demonstrating how such a tool may be useful in articulating to various stakeholders what such a course succeeds in accomplishing in terms of student outcomes. Our assessment regime could help others find ways to evaluate the effectiveness of cross-cultural leadership educational interventions and engage in reflection on learning goals to the end of developing culturally competent future leaders.

References

- Association of American Colleges & Universities (AAC&U). (2013). Global learning VALUE Rubric. Retrieved from <https://www.aacu.org/value/>
- Bender, R., & Lange, S. (2001). Adjusting for multiple testing – when and how? *Journal of Clinical Epidemiology*, *54*, 343-349.
- Bennett, M.J. (2004). Becoming Interculturally Competent. In J.S. Wurzel (Ed.), *Toward multiculturalism: A reader in multicultural education* (pp. 62-77). Newton, MA: Intercultural Resource Corporation.
- Berger, P. L., & Luckmann, T. (1966). *The social construction of reality*. New York: Anchor.
- Bogardus, E.S. (1947). Measurement of personal-group relations. *Sociometry*, *10*(4), 306-311.
- Braskamp, L., Braskamp, D., & Engberg, M. (2014). Global Perspective Inventory (GPI): Its purpose, construction, potential uses, & psychometric characteristics. Retrieved from <https://gpi.central.edu/supportDocs/manual.pdf>
- Brungardt, C., Greenleaf, J., Brungardt, C., & Arsendorf, J. (2006). Majoring in leadership: A review of undergraduate leadership degree programs. *Journal of Leadership Education*, *6*(1), 4-25.
- Chrislip, D.D., & O'Malley, E. (2013). *For the common good: Redefining civic leadership*. Wichita, KS: KLC Press.
- Correll, J., Judd, C.M., Park, B., & Wittenbrink, B. (2010). Measuring prejudice, stereotypes, and discrimination. In J.F. Dovidio, M. Hewstone, P. Glick, & V.M. Esses (Eds.), *The SAGE handbook of prejudice, stereotypes, and discrimination* (pp. 45-62). London, UK: SAGE.
- Cullen, M. (2008). *35 dumb things well-intentioned people say: Surprising things we say that widen the diversity gap*. Garden City, NY: Morgan James.
- Dallimore, E., Rochefort, D.A., & Simonelli, K. (2010). Community-based learning and research. *New Directions for Teaching and Learning*, *124*, 15-22.
- Davis, M.H. (1983). Measuring individual differences in empathy: Evidence for a multidimensional approach. *Journal of Personality and Social Psychology*, *44*, 113-126.
- Fine, L.E. (2015). Teaching multicultural leadership using a social constructionist approach. *Journal of Leadership Education*, *14*(2), 209-217.
- Fink, A. (2006). *How to conduct surveys: A step by step guide*. 3rd ed. Thousand Oaks, CA: Sage.

- Ford, J., Harding, N., & Learmonth, M. (2008). *Leadership as identity: Constructions and deconstructions*. Basingstroke, UK: Palgrave MacMillan.
- Friedman, T.L. (2007). *The world is flat 3.0: A brief history of the twenty-first century* (3rd ed.). New York: Picador.
- Gerstl-Pepin, C., & Aiken, J.A. (2012). *Social justice leadership for a global world*. Charlotte, NC: Information Age Publishing.
- Gleim, J.A., & Gleim, R.R. (2003). Calculating, interpreting, and reporting Cronbach's alpha reliability coefficient for likert-type scales. Retrieved from <https://scholarworks.iupui.edu/bitstream/handle/1805/344/Gleim%20&%20Gleim.pdf?s..>
- Hammer, M.R. (2012). The Intercultural Development Inventory: A new frontier in assessment and development of intercultural competence. In M. VandeBerg, R.M. Paige, & K.H. Lou (Eds.), *Student learning abroad* (pp. 115-136). Sterling, VA: Stylus.
- Hartman, E., Lough, B., Toms, C., & Reynolds, N. (2015). The beauty of global citizenship; The problem of measurement. In B. Oomen, E. Park, M. Sklad, & J. Friedman (Eds.), *Going Global: The theory, practice, evaluation, and experience of education for global citizenship*. Amsterdam, Netherlands: Drukkerij Publishing.
- Hartmann, D. (2015). Reflections on race, diversity, and the crossroads of multiculturalism. *The Sociological Quarterly*, 56, 623-639.
- Hovland, K. (2014). Global learning: defining, designing, demonstrating. *American Association of Colleges and Universities*. Retrieved from http://www.aacu.org/sites/default/files/files/Global/global_learning_2014.pdf
- IDI. (2015). The roadmap to intercultural competence using the IDI. Retrieved January 22, 2015, from (<http://idiinventory.com/>).
- IUPUI Center for Service and Learning. (2015). CMG interview protocol. Retrieved from <http://csl.iupui.edu/doc/cmng/cmng-interview-protocol.pdf>
- Jamieson, S. (2004). Likert scales: How to (ab)use them. *Medical Education*, 38, 1217-1218.
- Kane, T.D., Zaccaro, S.J., Tremble, Jr., T.R., & Masuda, A.D. (2002). An examination of the leader's regulation of groups. *Small Group Research*, 33(1), 65-120.
- Karim, A.U. (2003). A Developmental Progression Model for Intercultural Consciousness: A Leadership Imperative. *Journal of Education for Business*, 79(1), 34-39.
- Knapp, T.R. (1990). Treating ordinal scales as interval scales: An attempt to resolve the controversy. *Nursing Research*, 39(2), 121-123.

- Komives, S.R., Longerbeam, S.D., Owen, J.E., Mainella, F.C., & Osteen, L. (2006). A leadership identity development model: Applications from grounded theory. *Journal of College Student Development, 47*(4), 401-418.
- Komives, S.R., Lucas, N., & McMahon, T.R. (2013). *Exploring leadership: For college students who want to make a difference*. 3rd ed. San Francisco, CA: Jossey-Bass.
- Komives, S.R., Wagner, W., & Associates (2009). *Leadership for a better world: Understanding the social change model of leadership development*. San Francisco, CA: Jossey-Bass.
- Kuh, G.D. (2008). *High-impact educational practices: What they are, who has access to them, and why they matter*. Washington, DC: American Association of Colleges & Universities.
- Kuh, G.D., & Ewell, P.T. (2010). The state of learning outcomes assessment in the United States. *Higher Education Management and Policy, 22*(1), 9-28.
- Likert, R. (1932). A technique for the measurement of attitudes. *Archives of Psychology, 140*, 1-55.
- Lugg, C.A., & Tooms, A.K. (2010). A shadow of ourselves: Identity erasure and the politics of queer leadership. *School Leadership and Management, 30*(1), 77-91.
- Maki, P. (2002). Developing an assessment plan to learn about student learning. *Journal of Academic Librarianship, 28*(1), 8-13.
- Nieto, J. (2006). The cultural plunge: Cultural immersion as a means of promoting self-awareness and cultural sensitivity among student teachers. *Teacher Education Quarterly, 33*(1), 75-84.
- Oakleaf, M. (2009). The information literacy instruction assessment cycle: A guide for increasing student learning and improving librarian instructional skills. *Journal of Documentation, 65*(4), 539-560.
- Parrillo, V.N., & Donoghue, C. (2005). Updating the Bogardus social distance surveys: A new national survey. *The Social Science Journal, 42*, 257-271.
- Pope, R.L., Mueller, J.A., & Rynolds, A.L. (2009). Looking back and moving forward: Future directions for diversity research in student affairs. *Journal of College Student Development, 50*(6), 640-658.
- Rosch, D.M., & Schwartz, L.M. (2009). Potential pitfalls in outcomes assessment in leadership education. *Journal of Leadership Education, 8*(1), 177-194.

- Rosch, D.M., & Collier, D. (2013). Incoming leadership-oriented differences between students in a leadership studies course and a team-based project course. *Journal of Leadership Education, 12*, 103-121.
- Rost, J.C. (1991). *Leadership for the twenty-first century*. Westport, CO: Praeger.
- Seemiller, C. (2013). *The student leadership competencies guidebook: Designing intentional leadership learning and development*. San Francisco, CA: Jossey-Bass.
- Sobania, N., & Braskamp, L.A. (2009). Study abroad or study away: It's not merely semantics. *Peer Review, 11*(4), 23-26.
- Taylor, P.J., Russ-Eft., D.F., & Taylor, H. (2009). Gilding the outcome by tarnishing the past: Inflationary biases in retrospective pretests. *American Journal of Evaluation, 30*(1), 31-43.
- Upcraft, M.L., & Schuh, J.H. (1996). *Assessment in student affairs: A guide for practitioners*. San Francisco: Jossey-Bass.
- Wark, C., & Galliher, J.F. (2007). Emory Bogardus and the origins of the Social Distance Scale. *The American Sociologist, 38*(4), 383-395.
- Welch, B.L. (1947). The generalization of "student's" problem when several different population variances are involved. *Biometrika, 34*(1-2), 28-35.
- Wheatley, M. (2001). Listening as healing. Retrieved from <http://www.margaretwheatley.com/articles/listeninghealing.html>
- Wholey, J.S., Hatry, H.P., & Newcomer, K.E., Eds. (2004). *Handbook of practical program evaluation*. 2nd ed. San Francisco, CA: Jossey-Bass.
- William, D., & Black, P. (1996). Meanings and consequences: A basis for distinguishing formative and summative functions of assessment? *British Educational Research Journal, 22*(5), 537-548.
- Wren, T. (2002). Cultural identity and personal identity. In A.W. Musschenga, W. van Haaften, B. Spiecker, & M. Slors (Eds.), *Personal and moral identity* (pp. 231-258). Dordrecht, Netherlands: Springer Science+Business Media / Kluwer.

Author Biographies

Leigh E. Fine is an Assistant Professor in the Staley School of Leadership Studies at Kansas State University. He has taught graduate and undergraduate courses related to multicultural leadership, ethics in leadership, peer leader development, and foundations of

leadership. His scholarly interests include leadership pedagogy, assessment, queer theory, social construction, and intersections of gender and sexual identity. Email: fine@k-state.edu

Chance Lee is an instructor in the Staley School of Leadership Studies at Kansas State University where the mission is to develop knowledgeable, ethical, caring, inclusive leaders for a diverse and changing world. His forthcoming dissertation explores the distinction between the global and the local, specifically regarding student learning. Email: chance@k-state.edu