Individual capacity growth over time in leadership courses: an intra-individual multilevel model approach

Robert C. Klein
Department of Psychology, Cornell University, Ithaca, New York, USA, and
David Michael Rosch
Department of Agricultural Leadership, Education and Communications, University of Illinois at Urbana-Champaign, Champaign, Illinois, USA

Abstract

Purpose – Our study was designed to investigate the longitudinal trajectories of student leader development capacities in a sample of students enrolled in multiple leadership-focused courses across several semesters. Our goal was to assess the degree to which course enrollment was associated with growth over the time that students engage as undergraduates in academic leadership programs, and if so, to assess the shape and speed of capacity change.

Design/methodology/approach – We utilized a multilevel intra-individual modeling approach assessing students’ motivation to lead, leader self-efficacy, and leadership skills across multiple data collection points for students in a campus major or minor focused on leadership studies. We compared an unconditional model, a fixed effect model, a random intercept model, a random slope model, and a random slope and intercept model to determine the shape of score trajectories. Our approach was not to collect traditional pre-test and post-test data – choosing to collect data only at the beginning of each semester – to reduce time cues typically inherent within pre-test and post-test collections.

Findings – Our results strongly suggested that individual students differ greatly in the degree to which they report the capacity to lead when initially enrolling in their first class. Surprisingly, the various models were unable to predict a pattern of longitudinal leader development through repeated course enrollment in our sample.

Originality/value – Our investigation employed statistical methods that are not often utilized in leadership education quantitative research, and also included a data collection effort designed to avoid a linear pre-test/post-test score comparison.

Keywords Multilevel modeling, Longitudinal assessment, Leadership courses

Paper type Research paper

Much of the world appears to be facing a leadership crisis as trust in corporate and government leaders has declined significantly in recent years. The United Nations reported that trust in the United States government has decreased from 73% in 1958 to 24% in 2021 (Perry, 2021). This loss of trust could be due to contemporary challenges like the COVID-19 Pandemic and increasing economic inequality that exposed divisions that leaders have not adequately addressed. Furthermore, only 43% of Americans have some trust or confidence in the ethical behavior of large businesses, and 46% report that they do not have much trust or confidence (Horsley, 2018). As trust in our leaders wanes, leaders must be better trained to...
lead in complex and adaptable ways. In the context of this highlighted nationwide leadership crisis, our study takes on added significance. We aimed to investigate whether the current educational curricula, specifically leadership courses, have a tangible impact on developing leader capacity. We suggest a novel method for evaluating how leader development is taught, as the prevailing methods do not adequately address the current leadership void. Research on leadership education evaluation has primarily focused on interventions that measure the impact of singular leadership learning experiences. The weakness in this approach is that interventions train leaders to respond to a challenge with a specific skill instead of building their leadership toolkit. With the increasing complexity of leadership, leaders need to be trained to handle challenges in context-free, generalizable ways.

Our study presents a potential paradigm shift from evaluating leadership interventions to evaluating change within students over time frames much longer than single-semester leadership courses. Such a design, often labeled “intra-individual multilevel modeling,” follows students over several repeated data collection waves to map the trajectory of their development (intra-individual change) within the context of a population of students enrolled in leadership courses (modeling a layer of collective change). Our approach involved collecting panel data on students taking leadership courses and modeling their development to better understand how a leadership curriculum impacts their leader capacity development. By identifying patterns, we can develop a more comprehensive understanding of who leadership courses benefit the most. By addressing the concerns raised by contemporary researchers who evaluate leadership programming (Dugan, 2006; Rosch & Jenkins, 2020), our data collection method can refocus leadership education to evaluate how individuals develop leader capacity.

Leadership development in education and organizations
In order to understand leadership development, it is crucial to recognize the roles colleges and universities play in nurturing future leaders, given that leadership education has long been a significant aspect of their mission (Komives, Lucas, & McMahon, 2007; Morse, 2009). Universities have invested in leadership education by developing degree programs and establishing leadership centers (Astin & Astin, 2000) and scholars have developed theories of leadership for students (e.g. Higher Education Research Institute [HERI], 1996). The college years are a critical time for developing leadership skills (Pascarella & Terenzini, 2005), which lead to numerous benefits for the individual and their academic learning, employers, and communities. In studies of students participating in leadership development programs, participants increased their ability to lead (Dugan & Komives, 2010), multicultural awareness (Cress, Astin, Zimmerman-Oster, & Burkhardt, 2001), motivation to lead (Rosch, Collier, & Thompson, 2015), and clarified their values (Polleys, 2002). Compared to students not enrolled in leadership courses, Students who completed a leadership course reported higher levels of self-efficacy and leader capacity (Arendt, 2004; Endress, 2000; Posner, 2009).

Academically, when students participated in collaborative learning environments, those who reported higher levels of leader capacity also had higher grades (Dunbar, Dingel, Dame, Winchip, & Petzold, 2018). In organizational settings, 80% of participants in leadership training reported that their training helped them improve their business skills and bring fresh perspectives to their place of employment (Black & Earnest, 2009). Finally, participants in leadership programs increased their understanding of cultural diversity within their community, and their leadership training influenced the organizations they volunteered with (Black & Earnest, 2009).

While previous research suggests leadership education has numerous benefits for individuals, organizations, and communities in the short term, the long-term impact of leadership development programs has been inconclusive. In one study, Garza (2000) found
that ten years after participating in a leadership development program, leadership training played a significant role in participants’ professional development and decisions to attend graduate school. In a separate study, Felser (2005) found no strong positive impact on participants’ self-reported leadership capabilities ten years after participating in a university-sponsored leadership development program. These conflicting findings suggest that there is a need for stronger tools and methods to assess the impact of longitudinal leadership development programs (e.g. Dugan, 2006; Rosch & Jenkins, 2020) and to conduct longitudinal research that can provide a better understanding of the developmental impact of leadership courses (e.g. Bass & Bass, 2009; Burns, 2012). We hope to better address these issues using an intra-individual approach, which follows individuals over time across various engagements. Since the discussion of leader capacity is nuanced, we define terms and concepts in the subsequent section.

**Defining leader capacity**

Building on the discussion of leadership development programs and their lasting impact, this section focused on defining the specific leader capacities addressed in the study. Leader “capacity” is an integrated combination of knowledge, behaviors, and skills that determine an individual’s effectiveness in a leadership role (Day *et al.*, 2009; Hannah, Avolio, Luthans, & Harms, 2008). For students, coursework can be significant in developing leader capacity. Students clarify definitions, have discussions, and connect theories and concepts to their leadership practice through courses. Students can also develop their leader capacity through extra- and co-curricular experiences such as on-campus jobs, student organizations, and volunteer service. Given the potential impact of the COVID-19 pandemic on student extracurricular involvement and the unique circumstances commuter students face, leadership courses offer a unique and well-positioned space for leadership development.

Rosch and Collins (2020) found that in order to practice leadership effectively, a leader needs to be “ready,” “willing,” and “able”; if one part of the foundation is missing, the potential for effective action is drastically reduced. Our study used the Ready, Willing, and Able Leadership (RWAL) Scale to measure student leader capacity (Rosch & Collins, 2020).

“Ready” leaders possess a mindset conducive to leading, known as leader self-efficacy (LSE). Beginning with Bandura’s social cognitive theory of self-efficacy, self-efficacy is the internal belief that an individual can successfully accomplish a given task (Bandura, 1997). Efficacy is often applied to various professional activities. In leadership, leader self-efficacy refers to whether an individual believes they would be successful in serving in a leadership position and managing a leader’s responsibilities (Hannah *et al.*, 2008). “Willing” leaders possess the motivation to pursue leadership experiences, as described by the Motivation to Lead (MTL) construct (Chan & Drasgow, 2001). MTL is divided into three parts: affective identity MTL, social normative MTL, and non-calculative MTL (Chan & Drasgow, 2001). Affective identity MTL refers to the degree to which an individual enjoys participating in the leadership process as it aligns with their self-concept as a leader. Social normative MTL stems from the duty, responsibility, or obligation to lead from one’s peers or organizations they are part of. Non-calculative MTL refers to how individuals pursue leadership positions without regard for the benefits (social or material) that come with leadership positions. Having higher MTL is predictive of why students assume leadership positions, how they persevere and display resilience, and what opportunities they pursue (Rosch & Villanueva, 2016).

When students are both “ready” and “willing,” they possess the efficacy and motivation to lead and are “able” to practice leadership. Leadership skill includes communicating a compelling vision, empowering followers to reach their full potential, clarifying expectations...
and providing rewards or punishments based on performance. Originally utilizing the transactional and transformational leadership framework (Burns, 2012), leadership skill includes transactional and transformational leadership behavior. Recent research suggests a lack of clear distinction between the two among university students (Rosch & Collins, 2020). Educators attempt to evaluate how well students develop their capacity to lead but often undermine their scholarship by utilizing data collection and analysis methods that fail to adequately investigate certain presumptions about leader development program participation results.

**Potentially incorrect presumptions about program participation**

*Every leadership program for every student*

Many leadership courses and experiences are designed to be open to students regardless of their background characteristics (Dugan, 2011). This structure presupposes that regardless of the diversity of students, most programs are designed without such diversity in mind. However, this is problematic as it oversimplifies the complexity of leadership practice. In reality, there are multiple paths towards effective leadership practice and most leadership situations do not have clear right and wrong answers. When researchers assess students without regard to their backgrounds, incoming capacity, and personality differences, a crucial part of the story is missing in their analysis. Instructors need to adopt a more intra-individual approach that considers their students’ diversity (Avolio & Hannah, 2008) rather than one that lumps students together in their course evaluation by examining their mean scores.

*All leadership development is linear*

Current statistical techniques, such as regression, are conducted under the implicit assumption that leadership development is linear. However, leader capacity development is rarely linear and is marked by twists, turns, and delays (Boyatzis, 2009; Dugan, 2011). Leadership development is not a simple process where one attends a learning experience and walks away with a heightened ability to lead. Instead, development is incremental and maintained through continuous participation in leadership learning experiences (Komives, Owen, Longerbeam, Mainella, & Osteen, 2005). The leadership development process could be more accurately described as quadratic, a sine wave, or some other complex function. Currently, our latest scholarship does not provide enough evidence in any direction to make a specific assumption about the shape of leadership developmental trajectories.

*Participation results in broad-based progress*

Leadership researchers implicitly presume that the leadership learning experiences work, that it is beneficial to attend training regularly, and that increasing the frequency of training sessions leads to greater benefits. A preponderance of published scholarship on the impact of leadership education suggests that students attending programs increase their leader capacity (Harvard University, 2016; Posner, 2009; Rosch et al., 2015). However, research suggests that more complex person-level differences determine whether students see their leader capacity increase. If students lack the requisite motivation or efficacy, they might not report positive leadership development and instead report no development or negative development (Keating, Rosch, & Burgoon, 2014; Rosch & Villanueva, 2016). Therefore, researchers hoping to predict leadership development should consider a holistic perspective of leader capacity that accounts for person-level differences instead of focusing on a subset of variables.
A common measurement issue: The soft tyranny of the pre-test and post-test design

Pre- and post-test designs are a common way to evaluate the effectiveness of leadership interventions (Dimitrov & Rumrill, 2003). However, they have become dominant in leadership education, leading to a narrow focus on demonstrating an increase in leader capacity (Arthur & Hardy, 2014; Posner, 2009; Rosch, Ogolsky, & Stephens, 2017). This often involves giving participants an assessment of their leader capacity at the beginning of a leadership learning experience and then again at the end to measure the difference. While participants often report an increase in leader capacity after an intervention, it is important to consider what happens after the intervention ends. Research has shown that some aspects of leader capacity remain elevated three months after a learning experience, while others decrease significantly from their post-test high (Rosch, Stephens, & Collins, 2016). Moreover, the developmental cue of a post-course survey can encourage participants to report a higher level of development. Therefore, it is worth exploring alternative evaluation methods that do not rely on pre- and post-tests.

One reason to be cautious about pre- and post-tests is that they suffer from notable biases. For instance, participants may mature, change their behavior and attitudes, and seek different experiences between the pre- and post-tests. Without controlling these confounding factors, evaluations may reflect maturity rather than the intervention’s impact. For example, Posner (2009) found significant changes in leadership behaviors between first-year college students and seniors, suggesting that students may develop leader capacity through their involvement in extracurricular activities. Relatedly, response shift bias threatens validity because the intervention often causes a shift in perception that leads participants to overestimate the change between the pre- and post-tests (Drennan & Hyde, 2008; Howard & Dailey, 1979). This bias might explain why many leadership interventions yield positive development.

In consideration of an intra-individual approach to assessment

When evaluating the effectiveness of leadership programs, the field has predominantly focused on a program-centered approach, which emphasizes the leadership learning experience as the primary driver of capacity development. Typically, this approach involves comparing the mean scores of groups of program participants before and after the program.

In contrast, our study takes an intra-individual approach, examining how individuals develop their leader capacity across multiple learning experiences and semesters on campus. To capture intra-individual change accurately, sophisticated multilevel model statistical techniques are necessary. Additionally, leadership development programs in higher education aim to prepare students for real-world leadership beyond the academic setting. Evaluating students solely within the confines of a single programmatic intervention limits the ability of evaluators to examine how programs result in behavior change in the “real world.” Collecting data from students engaged in various programs over multiple years, we adopt an intra-individual longitudinal perspective that goes further than a pre-post assessment structure. Therefore, our comprehensive approach enables a more comprehensive understanding of the long-term effects of leadership education, considering the broader context of students’ development and the cultivation of their leader capacity.

Present study design

The cues that students receive when data is collected at the very beginning and end of courses might be stronger than scholars realize, subtly - or not so subtly - nudging students down or up the range of Likert-scale responses on quantitative measures of their capacity. One of our
primary goals was to eliminate these cues while still collecting longitudinal data. Since these interventions are limited to a controlled space – the classroom – they do not account for how students interact with their coursework over time, such as returning to multiple courses and assimilating new experiences and information. We introduced a novel form of evaluation to mitigate the above biases and issues.

The LEAD course research study is an ongoing effort that collects data from students enrolled in leadership courses at a large Midwestern university. The study distinguishes itself from cross-sectional and longitudinal studies (single group or quasi-experimental designs) in several notable ways. Firstly, it is event-contingent; students are invited to assess their leader capacity only at the beginning of courses, where a later course loosely serves as a post-test to an earlier course – thus removing implicit cues that scores should be higher when data is collected at the end of an experience. Secondly, it employs a panel design, where participants’ data are matched across several leadership courses over time to map a developmental trajectory for each participant. Thirdly, the study includes unbalanced cases since not all students take the same number of leadership courses. Participants have between two and seven cases included for analysis, and multilevel modeling (discussed in the methods section) accounts for unbalanced cases and weights participants appropriately (Frees, 2004). Fourthly, there is no post-course evaluation. Instead of a post-course evaluation given at the end of the semester, students complete the same survey at the beginning of the following semester. The departure from the traditional post-course evaluation was motivated by the finding that leader capacity reliably increases between pre- and post-course evaluations (Arthur & Hardy, 2014; Posner, 2009; Rosch et al., 2017). By removing the post-course assessment, we sought to understand the longer-term effects of leadership course enrollment on leader capacity development. This new approach allowed us to capture changes in leader capacity over a more extended period beyond the immediate effects observed at the end of a course. By tracking students’ progress through subsequent semesters, we aimed to gain deeper insights into leadership education’s sustainability and lasting impact on leadership development. The present study uses longitudinal multilevel models to examine student leader capacity development.

Methods

Participants

Our study included undergraduate students enrolled in semester-long leadership courses within the [Department] at a large, research-extensive university in the Midwestern United States. The data were collected from various courses such as “Foundations of Leadership,” “Leading Teams,” “Leadership Communications,” and “Leadership Ethics,” among others, spanning from the Fall 2019 to Fall 2022 semesters.

Out of the participants in this study, the majority, accounting for 98%, fell within the age range of 18 to 24. Regarding gender identification, approximately 69% identified as women, while 29% identified as men. Regarding racial and ethnic backgrounds, approximately 72% of the participants identified as White, 11% as Asian/Asian-American, 6% as Black/African-American, 6% as Latinx, 1.4% as Middle Eastern, and 3.5% as multi-racial. It is worth noting that while the sample’s demographics mostly align with the overall undergraduate student population at the university, there was a slight overrepresentation of women and white students in our sample.

Procedures

Data were collected via Qualtrics online surveys and in-person courses through the [Department]. A research team member invited students to complete the survey, provide their
consent, and complete the instrument within the first two weeks of each semester. Data were entered into SPSS v. 27 for initial data screening and descriptive statistics. The overall survey included more items than just those related to our study. We removed participants from the study who did not have at least two time points, resulting in 318 total participants spread across 767 observations. By emphasizing the exploratory nature of our study, we recognize the inherent constraints of including participants with only two time points. As many students had three points, our multilevel modeling approach addresses this issue by assigning a higher weight to participants with more time points. Our study can serve as a stepping stone for further research and contribute valuable insights to the existing knowledge base on leadership development.

Measures
Our study used the Ready, Willing, and Able Leadership (RWAL) Scale developed by Rosch and Collins (2020) to assess student leader capacity. This scale comprises five subscales that measure different aspects of leadership: (1) affective identity motivation to lead (AI-MTL), (2) social normative motivation to lead (SN-MTL), (4) leader self-efficacy (LSE), and (5) leadership skill. Participants provided item responses on a 7-point Likert scale, ranging from 1 = Strongly Disagree to 7 = Strongly Agree. The reliability of the entire scale was found to be very good ($\alpha = 0.84$). The RWAL Scale served as a comprehensive measurement tool to capture various dimensions of student leader capacity in our study.

Motivation to Lead. Motivation to lead consists of three subscales. The affective identity motivation to lead (AI-MTL) subscale assessed the extent to which individuals enjoy engaging in leadership activities that align with their self-concept as a leader. This subscale comprised items such as “I am the type of person who likes to be in charge of others” and “I usually want to be the leader in the groups that I work in.” The reliability of this subscale was excellent ($\alpha = 0.90$).

Social normative motivation to lead (SN-MTL) measures the perceived duty, responsibility, or obligation to assume leadership roles based on influences from peers or the organizations to which individuals belong. Example items for this subscale include “I feel I have a duty to lead others if I am asked” and “I agree to lead whenever I am asked or nominated by other group members.” The reliability of this subscale was good ($\alpha = 0.66$).

Non-calculative motivation to lead (NC-MTL) examined individuals’ inclination to pursue leadership positions without considering the social or material benefits of such roles. Reverse-scored items were included in this subscale, such as “I will never agree to lead if I cannot see any benefits from accepting the role” and “I will only agree to be the leader if I know I can benefit from that role.” The reliability of this subscale was found to be very good ($\alpha = 0.86$).

Leader self-efficacy. Leader self-efficacy (LSE) measures an individual’s beliefs regarding their ability to effectively serve in a leadership position and manage the associated responsibilities. This construct captures their confidence in influencing and guiding a group they lead. An example item from the subscale is “I am confident in my ability to influence a group that I lead,” and a reverse-scored item is “I have no idea what it takes to keep a group running smoothly.” Reliability was good ($\alpha = 0.76$).

Leadership skill. Leadership skill encompasses items that capture an individual’s ability to engage in leadership behaviors effectively. This construct focuses on actions and practices that demonstrate consideration for others and the recognition of their contributions. Example items include “I give special recognition when the work of other group members is very good” and “I behave in a way that is thoughtful to the needs of other group members.” Reliability was good ($\alpha = 0.76$).
Analytic design

Our study employed multilevel modeling, a technique that can be used when collecting nested data with different “levels” of analysis (such as students in classrooms across schools) or longitudinal intra-individual data (such as following an individual over several sets of responses). Our study employed an intra-individual multilevel model where we followed several hundred students over the course of years, assessing the changes in their capacity to lead. In this sense, one “level” of data might be represented by how an individual student’s scores change over time. Layered on top, another “level” would be the collective sample of all students in our study, whose scores change over time. This design is more appropriate for examining longitudinal data than regression, where observations must be independent and data are collected at one time point (Singer & Willett, 2003).

To further facilitate an understanding of our findings, we briefly explain some key terminology related to multilevel modeling. Multilevel models, also known as mixed models, incorporate fixed and random effects. Fixed effects remain constant across all participants, like in ordinary least squares (OLS) regression. In the context of our study, a fixed effect would be the overall impact of the curriculum on leadership skill, regardless of individual variation. On the other hand, random effects capture the unique variations in an individual’s development that are not explained by fixed effects. For example, individual students might have varied levels of prior leadership experience or motivation (their “intercept”), which would introduce variability in their responses. Furthermore, individuals might also progress at different rates (their “slope”). Random effects are further categorized into random slopes and random intercepts. Random slopes account for varying development rates among individuals, whereas random intercepts accommodate different initial capacity levels.

Researchers determine statistical significance using the t-statistic and p-value for a fixed effect. However, estimating the significance of random effects is slightly more complex. In our analysis, we employed a 95% confidence interval, as it provides information regarding both statistical significance and the direction and magnitude of the effect and allows us to contextualize the amount of slope and intercept variability. Confidence intervals work under the assumption of long-run, repeated sampling. In a 95% confidence interval, 95 out of 100 constructed intervals would contain the parameter of interest. If the confidence interval encompasses the value zero, we accept the null hypothesis, indicating the absence of an effect. Additionally, the proximity of the confidence interval to zero indicates a smaller effect size.

Having outlined these essential terms, we describe our approach to model building in our study.

In our approach to model building, we began with an (1) unconditional model that only incorporated the intercept or the mean value of leader capacity (i.e. leader self-efficacy) across the entire sample. This model is practically equivalent to simply calculating the mean score of all measures across the sample, lumping all scores together. Unconditional models serve as the foundation for multilevel modeling analyses. From the unconditional model, we computed the intra-class correlation (ICC), a statistical measure that assesses the suitability of employing multilevel models. The ICC examines the proportion of explained and unexplained variance. Notably, each ICC (for each scale of leadership capacity) exceeded 0.5, indicating an appropriate level of variation in leader capacities among individuals, suggesting an unconditional model was inappropriately simple to investigate student capacity change (Liljequist, Elfving, & Roaldsen, 2019).

Subsequently, we developed a (2) fixed effect model, introducing a single predictor variable: enrollment in leadership courses. Fixed effect models assume that all individuals start at the same point and develop at the same rate; contrasted with the unconditional model, we now incorporated our “time” variable (enrollment in subsequent leadership courses over time) to understand the longitudinal relationship.
Since it might be unrealistic to presume that all students start with the same capacities and develop at the same rate, our next models included random effects to account for potential individual differences. First, we constructed a (3) random slope model that estimated both the fixed effect of enrollment and the random slope associated with enrollment. Here, our primary interest was understanding if different students grow at different rates in their capacity when taking the same number of leadership courses. Moving forward, we constructed a (4) random intercept model that estimated enrollment’s fixed effect alongside the intercept’s random effect. In this model, our primary interest was understanding if starting leader capacity explains additional variance beyond the number of leadership courses participants enrolled in and the rate at which they grow in their capacity to lead.

Finally, we developed a (5) random slope and intercept model, which estimated three essential parameters: the fixed effect of enrollment (the number of semesters enrolled in leadership courses), the random slope of enrollment (the rate at which their capacity changes as a result of those courses), and the random effect of the intercept (the fact that students come to their first leadership course at differing incoming capacities to lead). Through conducting these successively more complex models, we sought to explore the intricate dynamics between enrollment in leadership courses and individuals’ leader capacities, somewhat practically analogous to hierarchical multiple regression, where each successive step allows a researcher to better understand how much variance is predicted in score changes.

We first analyzed means and dispersion statistics using the previously described longitudinal dataset to examine normality. We then employed a multilevel modeling approach to examine leader capacity development. Before analysis, we carefully matched the data to ensure the correct identification of individual participants across time points. Using maximum likelihood estimation, we analyzed the data with R version 4.2.3 and the NLME package. We conducted the following analyses with each of the five variables: (1) affective identity MTL, (2) social normative MTL, (3) non-calculative MTL, (4) LSE, and (5) leadership skill. We began by fitting an unconditional model, or null model, without any predictor variables included. In multilevel longitudinal analysis, the unconditional model estimates the variation in the outcome variable that can be attributed to differences across individuals. We estimated the intra-class correlation (ICC) from the unconditional model to determine the appropriateness of multilevel modeling. The higher the ICC, the more “nested” the data and the higher the necessity of multilevel models. Each ICC was above 0.5, indicating an appropriate amount of variation between individuals’ leader capacities (Liljequist et al., 2019). Subsequently, we added our time variable as a predictor and examined the fixed effects to examine the overall trend in the outcome variables over time.

Next, we fit three multilevel growth curve models to assess whether time engaged in leadership courses predicted leader capacity development. Specifically, we tested a random intercept model to account for individual differences in the level of leader capacity that participants started with. We then tested a random slope model to account for individual differences in the rate at which students might grow in their leader capacity. Finally, we tested a random intercept and slope model to account for both differences simultaneously. We then compared these growth curve models against the unconditional model to investigate if the growth curve models fit the data better than the unconditional models.

**Results**
Descriptive statistics of the analytic variables are included in Table 1. Participants demonstrated a moderate-to-high level of LSE ($M = 5.38$, $SD = 0.81$, Range: 2–7), AI-MTL ($M = 5.33$, $SD = 1.13$, Range: 1–7), SN-MTL ($M = 5.69$, $SD = 0.87$, Range: 1.67–7), NC-MTL ($M = 5.17$, $SD = 1.23$, Range: 1–7), and leadership skill ($M = 5.75$, $SD = 0.73$, Range: 2.8–7).
We analyzed our data for non-normality at the item level. All skewness scores fell within ±1.0. Most kurtosis scores also fell around or below 1.0, while one score was well above 1.0.

Our analysis explored five different models for each of the analytic variables: (1) the unconditional model, (2) the fixed effects model, (3) the random slope model, (4) the random intercept model, and (5) the random slope and intercept model. We aimed to incrementally build to a random slope and intercept model, which is the most descriptive of the five. The reason for this incremental approach is to examine how the estimation of additional parameters impacts the fixed and random effects. In the unconditional model, only the intercept is estimated. The unconditional model allowed us to calculate intra-individual and inter-individual variance and justify a multilevel modeling technique. The fixed effects model estimated a regression coefficient that fits a model for the dependent variable, the multilevel model equivalence of ordinary least squares (OLS) regression. Since OLS regression does not account for person-level differences, we used additional multilevel models to estimate person-level parameters such as incoming leader capacity (the random intercept) or the rate of change of development (the random slope). We first estimated the random intercept and the random slope parameters in separate models. After, we combined them into a random slope and intercept model, the most integrated model that allowed us to simultaneously estimate the fixed effect of enrollment in leadership courses, and a random slope and intercept measuring leader capacity.

Among the unconditional model analyses, each of the five measures of leader capacity had a statistically significant intercept that was higher than zero \( p < 0.05 \) and each ICC was over 0.5, indicating an appropriate amount of variation between individuals’ leader capacities (Liljequist et al., 2019). The fixed effects models focused on capacity growth over time and across multiple leadership courses, presuming that the population of students does not vary in their entering capacity to lead, nor in the rate at which they develop increased capacity (presumptions that many readers may treat skeptically). These analyses suggested that only growth in leader self-efficacy (not other capacities) were associated with course enrollment over time. For example, the fixed effect coefficient for non-calculative motivation to lead \( (\beta_1 = -0.002, p = 0.93) \) suggests that the impact of multiple course enrollment over time was practically 0. Still, this modeling structure is based on the presumption that all students enter into their course experience at the same level of leader capacity and grow at identical rates, so it makes sense to explore more complex models to investigate the degree that enrollment over time is associated with growth.

The random slope analyses (which presume that students all enter at the same capacity level but learn at different rates) suggested, like the fixed effects model, that repeated course enrollment predicted only growth in leader self-efficacy. For example, the fixed effect for non-calculative motivation to lead \( (\beta_1 = -0.002, p = 0.95) \) and the random slope for enrollment over time \( (\mu_1 = 0.14, 95\% CI = [0.06, 0.35]) \) suggests that the impact of enrollment was practically zero and students’ rate of growth varied by 0.14 units on a 7-point Likert scale, a minuscule difference when measured over the course of months and years.

<table>
<thead>
<tr>
<th></th>
<th>N</th>
<th>Mean(SE)</th>
<th>SD</th>
<th>Skewness</th>
<th>Kurtosis</th>
</tr>
</thead>
<tbody>
<tr>
<td>AI-MTL</td>
<td>767</td>
<td>5.33 (0.04)</td>
<td>1.13</td>
<td>−0.51</td>
<td>−0.20</td>
</tr>
<tr>
<td>NC-MTL</td>
<td>766</td>
<td>5.17 (0.04)</td>
<td>1.23</td>
<td>−0.71</td>
<td>0.15</td>
</tr>
<tr>
<td>SN-MTL</td>
<td>767</td>
<td>5.69 (0.03)</td>
<td>0.87</td>
<td>−0.92</td>
<td>1.70</td>
</tr>
<tr>
<td>LSE</td>
<td>766</td>
<td>5.38 (0.03)</td>
<td>0.81</td>
<td>−0.63</td>
<td>1.03</td>
</tr>
<tr>
<td>Leadership skill</td>
<td>765</td>
<td>5.75 (0.03)</td>
<td>0.73</td>
<td>−0.53</td>
<td>0.41</td>
</tr>
</tbody>
</table>

Table 1. Descriptive statistics

Source(s): Created by authors
Our random intercept analyses made the presumption that the level of incoming leader capacity among students varied, even though their presumed growth rate did not. These analyses suggested that repeated course enrollment was only significant to growth in leader self-efficacy. For example, in the random intercept model for non-calculative motivation to lead, the fixed effect of course enrollment over time ($\beta_1 = -0.002, p = 0.93$) suggests that the impact of enrollment was practically zero. The random intercept for enrollment ($u_0 = 0.92, 95\% CI = [0.82, 1.03]$) suggests students differed in their non-calculative motivation to lead by 0.92 units on a 7-point Likert scale before they even stepped foot in a classroom to learn about leadership; practically speaking, not all students enter with the same level of non-calculative motivation to lead.

The random slope and intercept analyses, the most complex in our series, presumed individual students could differ in their incoming capacity level and growth rate over time. Still, similar to other analyses, this version similarly suggested that repeated course enrollment was only significant to growth in leader self-efficacy. For example, in the non-calculative motivation to lead random slope and intercept model, the fixed effect of enrollment ($\beta_1 = 0.002, p = 0.95$), the random slope of enrollment ($u_1 = 0.15, 95\% CI = [0.06, 0.35]$) and the random intercept ($u_0 = 0.97, 95\% CI = [0.87, 1.09]$) suggest that the impact of enrollment was practically zero, as students’ rate of growth varied by only 0.15 units on a 7-point Likert scale, and students differed in their non-calculative motivation to lead by 0.97 units, on a 7-point Likert scale. Like the random intercept model, students differ in their incoming non-calculative motivation to lead by almost an entire unit, affirming that students enter leadership courses with different capacity levels.

As enrollment in leadership courses emerged as a statistically significant predictor of change in leader self-efficacy, we will highlight each specific finding within this analysis (see Table 2). The unconditional model suggested there was appropriate variation to utilize a multilevel modeling approach ($ICC = 0.57$). In the fixed effect model, repeated course enrollment revealed a statistically significant ($p < 0.001$) but marginal ($\beta_1 = 0.06$) impact. In the random slope model, repeated course enrollment showed a statistically significant ($p < 0.001$) but marginal ($\beta_1 = 0.06$) impact. The random slope also showed a statistically

<table>
<thead>
<tr>
<th>Fixed effects</th>
<th>Unconditional model</th>
<th>Fixed effects model</th>
<th>Random slope model</th>
<th>Random intercept model</th>
<th>Random slope and intercept model</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Coef</td>
<td>$p$</td>
<td>Coef</td>
<td>$p$</td>
<td>Coef</td>
</tr>
<tr>
<td>Intercept</td>
<td>5.37</td>
<td>0.000</td>
<td>5.30</td>
<td>0.000</td>
<td>5.30</td>
</tr>
<tr>
<td>Time</td>
<td>0.06</td>
<td>0.000</td>
<td>0.06</td>
<td>0.000</td>
<td>0.06</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Random effects</th>
<th>SD</th>
<th>95% CI</th>
<th>SD</th>
<th>95% CI</th>
<th>SD</th>
<th>95% CI</th>
<th>SD</th>
<th>95% CI</th>
<th>SD</th>
<th>95% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>0.61</td>
<td>[0.55, 0.68]</td>
<td>0.65</td>
<td>[0.58, 0.73]</td>
<td>0.09</td>
<td>[0.04, 0.18]</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Time $^b$</td>
<td>0.09</td>
<td>[0.04, 0.18]</td>
<td>0.09</td>
<td>[0.04, 0.18]</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Note(s):** Significance for fixed effects is calculated with $p$-values. Significance for random effects is calculated with confidence intervals

$^a$95% CI refers to 95% confidence intervals. The first number is the lower limit and the second is the upper limit

$^b$Time refers to the change in leader capacity score in subsequent leadership course enrollment

**Source(s):** Created by authors

Table 2. Longitudinal model building for LSE Journal of Leadership Education
significant (95% CI = [0.04, 0.18]) but marginal impact ($u_1 = 0.09$). In the random intercept model, repeated course enrollment showed a statistically significant ($p < 0.001$) but marginal ($\beta_1 = 0.06$) impact. The random intercept showed a statistically significant (95% CI = [0.55, 0.68]) and moderate ($u_0 = 0.61$) impact. The statistically significant and consistent impact of repeated course enrollment on leader self-efficacy highlights that students marginally differ in their rate of developing leader self-efficacy but do enter leadership courses with differing levels of leader self-efficacy.

Finally, in the random slope and intercept model, repeated course enrollment continued to show a statistically significant ($p < 0.001$) but marginal ($\beta_1 = 0.06$) impact. The random slope showed a statistically significant (95% CI = [0.04, 0.18]) but marginal impact ($u_1 = 0.09$). In comparison, the random intercept showed a statistically significant (95% CI = [0.58, 0.73]) and moderate impact ($u_0 = 0.65$). The model revealed that the fixed effect for the entire sample suggests that a student’s leader self-efficacy increases by 0.06 on a 7-point scale for every subsequent leadership course enrolled in. The random slope effect was close to zero, $u_1 = 0.09$, which implies that participants only marginally vary in their growth rate in leader self-efficacy. There was more variation in the random intercept, $u_0 = 0.65$, suggesting that participants enroll in leadership courses with a fair amount of variation in their leader self-efficacy scores.

**Discussion**

Our study aimed to explore the link between enrolling in leadership courses and developing student leader capacity. Taking an intra-individual approach that follows individual students over several semesters of course enrollment and utilizing multilevel modeling techniques, we sought to uncover the intricate nuances of this relationship. While we did not find a statistically significant impact of enrollment on motivation to lead or leadership skill, we discovered a noteworthy finding: enrollment in leadership courses did not seem to predict much leader capacity growth at all.

Our results were unexpected; all growth curves were not statistically or practically significant, suggesting that students as a group did not develop a greater capacity to lead as they spent more time in leadership-focused classes. These findings suggest that enrollment in leadership courses may not influence motivation to lead or leadership skill. Moreover, while statistically significant, leader self-efficacy (LSE) growth was practically meaningless; each subsequent course increased LSE by 0.06 units on a 7-point Likert scale, and we did not find any meaningful difference in the rate at which students developed LSE. However, an optimist (or gallows humorist) might state that at least students do not report lower levels of capacity after enrolling in leadership courses.

The incremental approach of building a random slope and intercept model means significant overlap between the fixed and random effects of the previous models. However, these models serve different purposes and make distinct presumptions about the underlying processes. The fixed effects model assumes all students start at the same capacity level and develop at the same rate. To add further complexity, two models estimated either the random intercept or the random slope. Estimating the random intercept assumes students enter leadership courses with differing capacity levels, while estimating the random slope assumes students develop at different rates. The most descriptive model, the random slope and intercept model, combines the assumptions from the previous models into a single model. From the random slope and intercept models, the random slopes were statistically or practically 0.0, implying that students might develop leader capacity at similar rates. Examining the random intercepts, students enter the classroom with different levels of leader capacity, with some capacities differing more than others.
Implications

The findings of our study raise important questions that warrant further examination of how individuals develop leader capacity. Despite a plethora of previous research indicating growth after leadership learning experiences (e.g. Arthur & Hardy, 2014; Posner, 2009; Rosch et al., 2017), our study, which followed hundreds of students over several years who enrolled in the types of leadership courses offered at many universities, did not observe any significant changes in leader capacity over the course of several years. To be clear, there were individual students whose capacity rose over time, but on the whole, gains were averaged out by losses. This disparity between our findings and several published reports prompts a critical evaluation of the differences.

The design of our study is common in longitudinal panel studies in other disciplines (e.g. Dobson & Ogolsky, 2022; McCoach & Kaniskan, 2010). Our results cause us to question the potential influence of study designs, data collection methods, and contextual factors in leadership education studies that may influence outcomes. From one perspective, our lack of statistically definitive findings could be the fault of our work. One plausible explanation for the disparity in our results could be our study design and data collection methods. It is possible that the measurement instruments employed, such as the RWAL Scale (Rosch & Collins, 2020), may have limitations in capturing the subtle changes in leader capacity over time. Furthermore, the time frame in which we collected data may have influenced the results. From another perspective, our study might serve as a window to future studies that examine capacity gains in ways that are not yet common in leadership education research. For example, future research could explore alternative designs and measurement approaches that capture leader capacity development over different periods or incorporate additional assessment methods to provide a more comprehensive and nuanced understanding of growth. They could also focus on specific sub-groups to investigate why some students might gain while others lose capacity over their years as a university student.

Additionally, contextual factors surrounding college students might have influenced our results. Our study focused on a specific university setting and student population, which may have unique characteristics and dynamics that differ from previous research contexts. Factors such as institutional culture, student demographics, and program structure might influence how individuals develop leader capacity. Exploring contextual influences and understanding how they shape an individual’s leader capacity is crucial for understanding leader capacity development comprehensively.

Our study explored the degree to which enrolling in multiple leadership courses over time was associated with increased student leader capacity. Since the results do not suggest that students develop significant leader capacity through leadership courses, we offer suggestions for improving the leadership learning experience. As there seems to be a disconnect between what students learn in class and how they practice leadership, educators can help students align their capacity and their leadership skill by meeting students where they are at and providing practical examples to practice what they are learning in class (Dugan & Komives, 2007).

Future research

Since our study raised more questions than answers, there are many directions for future research. At the individual level, we observed a diverse pattern of outcomes. Some students reported an increase in their leader capacity, while others remained at their initial level or even experienced a decrease. Future research endeavors could delve into these unexplored variables that may play a crucial role in shaping the trajectory of student leader capacity development. Understanding why certain individuals experience growth in their leader capacity while others do not would provide valuable insights for designing more targeted and
Effective leadership development programs (Kjellström, Stälne, & Törnblom, 2020). Potential factors to explore include personal characteristics such as social identity or past leadership experiences, environmental factors, individual motivations, and specific contextual experiences (Dugan & Komives, 2007). Suppose our findings might be representative of a larger picture of the developmental trajectories of university students who enroll in formal leadership courses. In that case, educators need to know more about what leads students to grow and what is not as helpful – and what types of students might be succeeding while others are not.

For researchers who use surveys or pre- and post-test designs, it is crucial to explore alternative approaches for assessing leader capacity to determine the degree to which this design might be associated with patterns of response that bias findings and potentially influence students to report inflated levels of development (Drennan & Hyde, 2008; Howard & Dailey, 1979). To mitigate these biases, we propose adopting diverse assessment styles that provide a more comprehensive and accurate understanding of leader capacity. Incorporating innovative evaluation techniques, such as performance-based assessments, behavioral observations, 360-degree feedback, or experiential simulations, can offer a more authentic and contextualized leadership development measure. By utilizing a multifaceted approach to assessment, researchers and practitioners can gain a more nuanced perspective on students’ leadership capabilities, enabling a more effective design of leadership education programs.

If researchers plan to use surveys, it would be helpful to look at interaction effects at the individual level to gain deeper insights into the dynamics of leadership development. By examining interaction effects in their statistical models, researchers can uncover how variables like extracurricular involvement, in combination with other variables, influence the development of leadership skills and behaviors. Interaction terms are relatively simple to add to statistical models by taking the cross-product of the two variables. Expanding on the example of extracurricular involvement, an involvement mean score can be calculated and multiplied with a time variable. The new interaction term can be added to the model to explore how time and involvement influence the development of leader capacity.

The unprecedented challenges brought about by the COVID-19 pandemic have undeniably influenced various facets of education, including leadership development. Our data collection spanned pre-pandemic, during the pandemic, and post-pandemic periods, offering a unique perspective on leadership education across these distinct phases. The shift to remote learning, coupled with the broader societal impacts of the pandemic, may have affected students’ engagement with leadership courses and their subsequent perceptions of leader capacity. While our study provides insights across these different timeframes, it is crucial to acknowledge the broader pandemic context when interpreting our findings. Future research might benefit from a more detailed exploration of the nuanced impacts of pandemic living and learning on leadership development, offering insights into how educational institutions can adapt and evolve in the face of such global challenges.

Limitations

We want to highlight a few notable statistical limitations of our study. While multilevel modeling weights cases depending on their level of missing data, our study had a lot of missing data, which can still impact the effect sizes of our analyses. Our approach to studying student leadership development comes with the messy reality that students do not take courses in order in subsequent semesters. Some students take multiple courses a semester and others have not taken leadership courses in years. The amount of missing data impacts the statistical power and the ability to estimate random effects accurately.
The data itself also had some irregularities that might have impacted the analysis. Students frequently overestimate or underestimate their leader capacity and our study had many students who might be statistical outliers. It was common for students to answer the survey with a series of “strongly agree” statements without much variation in their scores. For example, students reported their capacity of leader self-efficacy beginning a course at 5.38 out of a 7-point scale. The skew toward higher levels of leader capacity suggested that many students were already at ceiling performance; if their self-reported score accurately represented their true capacity, that would suggest they could not improve and would either regress toward the mean or stay the same.

Moreover, the statistical assumption for regression is that results follow a normal distribution. Students might trend toward non-normal score distributions in a non-random sample (like those who take a leadership course). Scores toward the end of the distribution might have a higher error term than scores closer to the mean. When these students are re-tested, they tend to regress toward the mean value and bias their actual level of leader capacity. Furthermore, those with higher levels of leader capacity might demonstrate smaller gains than those with lower leader capacity. While some students reported lower levels of leader capacity, many scores stayed the same. This suggests that while there was no growth, students still ranked themselves highly after learning the content.

Conclusion
In this study, we investigated the impact of enrollment in leadership courses on student leader capacity, utilizing the Ready, Willing, and Able Leadership (RWAL) Scale. Our approach, employing multilevel modeling, revealed intriguing insights. While we did not find a significant influence of enrollment in leadership courses on motivation to lead or leadership skill, enrollment in leadership courses emerged as a statistically significant predictor of change in leader self-efficacy. However, the actual increase in leader self-efficacy from enrolling in subsequent classes was practically meaningless. These findings shed light on recognizing individual differences in developing leader capacity. Our study contributes valuable insights to the leadership education literature, underscoring the potential of enrollment in leadership courses to bolster students’ confidence in their leadership abilities. Future research endeavors should delve into additional variables and refine measurement approaches. Overall, our findings hold implications for curriculum design and pedagogical strategies in leadership education, seeking to empower upcoming leaders to address contemporary societal challenges.

References


**Corresponding author**

Robert C. Klein can be contacted at: rk673@cornell.edu