ORGANIZATIONAL ASSESSMENT

Holding Higher Education to Account: Measuring What Matters in the Development of Students as Leaders

Ryan P. Brown, The Ann & John Doerr Institute for New Leaders, Rice University

Lebena Varghese, The Ann & John Doerr Institute for New Leaders, Rice University

ABSTRACT

A common claim among institutions of higher education is that "we make leaders." Although the vast majority of leaders will certainly pass through the hallowed halls of colleges and universities, whether and to what extent educational institutions have any actual impact on leader development is an open question with little supporting evidence. What would such evidence look like if we were to search for it? And how might we use an evidence-based approach to increase the effectiveness of leader development initiatives in higher education? This article describes a new initiative at Rice University, the Ann and John Doerr Institute for New Leaders, the purpose of which is to increase the leadership capacity of students across the entire university. This institute takes an evidence-based approach to leader development, uses only professional leader developers in its work, and self-skeptically determines success and failure through rigorous measurement of outcomes. We present preliminary evidence of progress in student leader development, along with a call for a more scientific approach to leader development throughout institutions of higher education.

Universities have long espoused the goal of developing the next generation of leaders as being central to their educational missions. Indeed, a common claim made by institutions of higher education, at least in the U.S., is that "we make leaders." Clearly defining what such a developmental process might entail, however, remains an ongoing challenge for universities, and measuring the extent to which they are succeeding in reaching this noble goal is both difficult and rare. Without a firm commitment to honest and rigorous measurement, no institution can hope to make consistent progress in developing students as leaders. Indeed, even if they managed to make progress, how would they know they had been successful? Evaluating successes and failures empirically is the only way to discern which efforts are yielding the desired results, and which efforts should be abandoned.

This article describes the approach of one leader development program to take measurement seriously and describes some of the preliminary findings that have derived from its work. We begin by describing what we see as some of the most prominent issues plaguing leader development initiatives within higher education. Subsequently, we detail our attempt at addressing these issues and the steps we have taken at the Doerr Institute for New Leaders to create a leader development program that is both impactful and sustainable. Finally, we present examples of the types of data we have gathered to evaluate program effectiveness, ending with an exhortation for those willing to take leader development seriously.

What Is the Problem?

The challenges of defining what success looks like and measuring the (potential) benefits of leader development initiatives are not limited to higher education. The world beyond the ivory tower fares only a little better when it comes to determining whether leader development programs are worth the hefty price paid by many corporations (Lacerenza, Reyes, Marlow, Joseph, & Salas, 2017). Claims about benefits are commonplace, but the quality of and evidence supporting many leader development initiatives vary widely across organizations (Harvard Business Publishing, 2016). Indeed, the landscape of social interventions in general, whether leadership-related or otherwise, is riddled with the refuse of good intentions. All too often, social interventions fail to produce any measureable benefits that stand the test of time and attempts at replication. With greater frequency than many people might expect, such interventions even do more harm than good, despite the grand intentions of those who implement them (Wilson, 2011).

Ryan P. Brown, Ph.D., is the Managing Director for Measurement at Rice University's Doerr Institute for New Leaders. He has authored or co-authored over 50 peer-reviewed articles and book chapters on topics ranging from interpersonal forgiveness to the role of culture in mental health. In 2016, he published *Honor Bound: How a Cultural Ideal Has Shaped the American Psyche* (Oxford University Press). Before joining the Doerr Institute, he was the L. J. Semrod Presidential Professor of Psychology at the University of Oklahoma, where he also served as the associate chair of the Department of Psychology and as a co-founder of the Institute for the Study of Human Flourishing.

Lebena Varghese, Ph.D., is a postdoctoral research fellow at Rice University's Doerr Institute for New Leaders, where she helps examine the impact of the Institute's initiatives on students' growth and development as leaders. Prior to joining the Doerr Institute, she worked in the survey research team at ACT, examining ways to increase survey response rates. Her previous research examined the impact of workplace mentoring on career advancement and stress reduction.

When this occurs, we might do well to ask how it is that our good intentions failed to produce the changes we wished to see. What went wrong? Were there important moderators that we failed to consider? Was the central failure at the level of our ideas or our execution of those ideas? Without careful empirical analysis along the way, we cannot hope to answer these questions reliably. Whether in the realm of leader development or any other domain of social intervention, we should take an evidence-based, empirical approach if we want to learn from our failures and accurately identify our successes.

Of course, recognizing that we ought to take such an approach to evaluating our intervention efforts and actually implementing a rigorous measurement system are two very different things. In the realm of leader development, we might get derailed from measuring outcomes by a failure to define what leadership is, or to decide whether we want to be focused on leadership education (teaching people about leadership and theories of leadership) or leader development (helping people grow in their leadership capacities; Day, Fleenor, Atwater, Strum, & Mckee, 2014). If we manage to overcome these obstacles, we still have to define what success would look like in order to begin the process of empirically evaluating whether and to what extent our efforts are producing the fruits that we intended.

...measuring outcomes is risky. What if our efforts produce little evidence of meaningful, lasting change?

A room filled with leadership experts might find it difficult to reach consensus on what these fruits should look like. And even if they do manage to reach a consensus, someone is likely to point out that measuring outcomes is risky. What if our efforts produce little evidence of meaningful, lasting change?

Would such a failure threaten our sources of funding? Once our funding is lost and our reputation damaged, could we realistically hope to secure new funds again in order to start over? Might it be better, in the end, to try to convince people that we must be successful because (1) we are the leadership experts, after all, and (2) just look at how many students have passed through our programs! If no one pauses long enough to consider these two assertions carefully, they just might be enough to satisfy the casual critic and allow us to continue simply doing what we have always done.

As long as assertions of expertise satisfy the call for evidence-based practices and body counts are our central index of success, leader development at institutions of higher education will be little more than empty promises. We will not know if we are being successful, nor will we know which of our programs or initiatives is responsible for any successes we might achieve. Consequently, we might pour our resources into many programs that fail to produce any benefits, while failing to fund those that might truly advance our goals at levels that could make a real difference in developing students as leaders.

An Evidence-Based Solution

The Doerr Institute for New Leaders began in 2015 with a strategic gift from Ann and John Doerr to Rice University. This gift was given to elevate the leadership capacity of Rice students across the entire university, and by doing so, to inspire other universities around the country to develop students as leaders in a similar, evidence-based fashion. Since its inception, the Doerr Institute has operated according to four "First Principles": (1) leader development should be considered a core function of a college or university (consistent with the claims so often made by those in higher education); (2) leader development initiatives should use evidence-based approaches, rather than simply following the latest fad or long-beloved

method (3) leader development initiatives should employ professional leader developers, not just well-intentioned but untrained volunteers; and finally, (4) rigorous measurement of desired outcomes (not just body counts) should preside over any serious leader development enterprise.

Although each of these principles is equally important, in this article we focus on principal 4, the careful measurement of outcomes that informs the continuation or retirement of programs within the Institute. The optimal way to ensure that our programs are making an impact is to evaluate them empirically. Evaluation fosters program designers' decision-making processes (e.g., continue/discontinue a program, make a change in the program content or approach) and enables them to ensure that programs are delivering on objectives (Ely, Boyce, Nelson, Zaccaro, Hernez-Broom, & Whyman, 2010; Grinnel, Gabor, & Unrau, 2015). Evaluating a program is a complex process that involves multiple phases, such as working closely with stakeholders to articulate objectives; pilot-testing instruments (e.g., surveys, behavioral exercises, observation rubrics) that will be used for data collection; disseminating results to the key stakeholders at strategic times (e.g., mid-program, end-of-program); and closing the evaluation loop — integrating findings from current evaluation efforts to amend existing plans and program objectives (if necessary).

At the Doerr Institute, the measurement team follows the process described above to help the implementation team measure its outcomes. However, to measure outcomes we need to establish a set of criteria. We borrow from Kirkpatrick's (2009) taxonomy (reaction, learning, behavior, and results) to identify types of evaluation criteria. Measuring outcomes at the results level refers to linking the impact of leader development programs to organizational metrics, but because we expect our student "clients" to graduate in a relatively short time span, we are less interested in assessing institutional outcomes at this level of the taxonomy than most businesses tend to be. The mission of our institute is to increase the leadership capacity of all students within our university across all colleges, disciplines, and school levels (i.e., undergraduate and graduate). Thus, we are ultimately interested in the degree of personal transformation experienced by students who participate in our programs, rather than changes to the university itself. Therefore, we focus on evaluating our programs against criteria that fall within the first three levels of the taxonomy. In addition to summative assessments, we also carry out formative assessments of our programs. Summative assessments provide evidence for the ultimate effectiveness of a training program. In contrast, formative assessments focus on internal processes and help to identify process-related features that could lead to improvements in the quality of training and the ultimate impact that we want to achieve (Ely et al., 2010).

Our implementation team has created initiatives that fall into three broad types — namely, *Activation* (one-on-one coaching), *Synthesis* (group coaching around a common theme), and *Catalyst* (more narrowly-focused skills-based training). The measurement team at the Doerr Institute does not play a role in content development for any of these programs. Nonetheless, the measurement team engages with the implementation team to identify the objectives of training initiatives and determine how best to measure these objectives in a scientific manner. This process is fundamental and is analogous to a scientist determining how to operationalize his or her hypotheses in an experiment. One could easily conclude that a program has little to no impact if the construct being measured is not actually the intended outcome. Similarly, a common pitfall of leader development initiatives is the creation of overly idealistic objectives (e.g., turning average, 18-year-old students into transformational leaders over lunch). It is important to be realistic about the impact potential of a two-hour workshop, as opposed to the

impact of an intensive, long-term, immersive training opportunity. In the former instance, the outcome is likely to be witnessed at the reaction and awareness level rather than at the behavioral level.

The measurement team also determines the design of the overall evaluation strategy. As leadership scholars have observed, cross-sectional designs that examine simple, bi-variate associations diminish the strength of the inferences we can draw from our evaluation efforts (Avolio, Reichard, Hannah, Walumba, & Chan, 2009; Feldman & Lankau, 2005). Therefore, at the very least, we adopt a pre/post comparison design to examine the impact of our programs, sometimes including longitudinal assessments that span a year or more in time. In the case of certain programs, we go a step further and adopt a quasi-experimental or experimental design with one or more comparison groups comprising students who are not exposed to our training program and who are matched on key demographic and motivational variables. Such evaluation designs enhance our confidence in the conclusions we make about the impact we are having on students and allow us to determine how long a program's effects last. A program with a small but long-lasting impact might be deemed more valuable than a program with a large but more temporary effect.

Preliminary Evidence of Developmental Impact

The Doerr Institute currently has three core programs designed to help students develop as leaders (all programs are free of charge and provide no academic credits). The Activation program provides students with an individual, professional leadership coach for a full semester. The oldest and largest of the Institute's initiatives, this program combines structured feedback and goal setting with individually tailored coaching and informal social accountability. The Synthesis program moves this one-on-one coaching into a peer group setting under a common theme, such as leading with confidence or overcoming perfectionism, and employs professional coaches as group facilitators. Finally, the Catalyst program narrows the scope of training even more to focus on specific leader competencies, such as how to give effective feedback, or how to launch a team. Each of these initiatives has its own, unique objectives. Consequently, outcome measures for each initiative are unique as well. Here, we will focus our discussion of program impacts on the Activation program, as we currently have more data associated with this program than with any other.

> *All students, regardless of major or background, are eligible to participate in the Institute's leader development initiatives.*

From the inception of the Doerr Institute, professional leadership coaching has been a cornerstone of the Institute's developmental portfolio. Any student at Rice who wants to develop his or her leadership abilities can receive professional, certified leadership coaching for a semester. These coaches are experienced professionals who work with executives in the business community and other leaders, and they receive ongoing, specialized training from the Doerr Institute on working with college students and on the specifics of the Rice University culture. Students do not have to compete to receive a coach, nor do they have to pay for this service. Thus, the Institute does not create any direct or indirect filters on the populations it serves. All students, regardless of major or background, are eligible to participate in the Institute's leader development initiatives.

Consequently, the Institute has managed to attract an almost perfect cross-section of the student body across every demographic or personality characteristic that we have measured (more on this later).

All coaching through the institute begins with an assessment of a student's emotional intelligence, using a validated tool called the EQi-2.0 (Stein & Book, 2011). There are many conceptual models and approaches to the measurement of emotional intelligence in the research literature, and the Doerr Institute does not take a hard line on which model is best. Rather, we use the EQi-2.0 (a so-called "mixed model" measure of emotional intelligence; Ashkanasy & Daus, 2005) because its predictive validity is supported by research (e.g., O'Boyle Jr., Humphrey, Pollack, Hawver, & Story, 2011), is efficient to administer, provides easily understood feedback to students, and facilitates discussions of the types of trainable "soft skills" that help to distinguish great leaders from poor ones, beyond cognitive skills or basic personality. Students debrief this assessment with their coaches and then complete a standardized leader development plan, in which they engage in a process of self-reflection on what the concept of leadership means to them and on their own ideals and values in the leadership domain. Students reflect on and articulate what they believe the best version of themselves as a leader might look like, and then they create a focused plan for how to grow toward this ideal. The leader development plan follows a research-based format for effective goal setting, and following its co-creation between the student and coach, goal progress is defined and monitored throughout the remainder of the coaching engagement. Examples of some of the most common leadership goals set by students are self-confidence, interpersonal skills, self-regulation, self-awareness, effective communication, and empathic engagement.

To evaluate whether professional coaching is effective in enhancing students' capacity to lead, the Institute has created a multi-dimensional, multi-method evaluation process that includes (but is not limited to) the following types of data:

1. **Reaction-Level Data.** At the most basic level, the Institute gathers data from students on every interaction between them and their coach. Students report on the perceived value of each coaching session, articulate their goal-related action steps, and evaluate their goal progress throughout the semester. Although it is critical to examine such reaction data to identify process-related opportunities for improvement, we will not discuss these low-level outcomes further in this article.

2. **Pre-Post Developmental Change Data.** Students complete a multi-item *Authentic Leader Identity Scale* (see Appendix) before and after a semester-long coaching engagement, which typically spans 4 to 5, hour-long sessions. This pre-post assessment allows the Institute to determine whether any growth in leader identity has occurred over time across all students engaged in the coaching process. Authentic leader identity comprises self-categorization as a leader, self-confidence as a leader, value-behavior consistency, and self-awareness of leadership strengths and weaknesses. Research shows that developing a strong leader identity is a fundamental part of motivation and skill development as a leader (Day & Harrison, 2007; Day, Harrison, & Halpin, 2009). Additional pre-post measures are included each semester, including measures of well-being (Diener, Emmons, Larsen, & Griffin, 1985), sense of purpose (Steger, Frazier, Oishi, & Kaler, 2006), and self-concept clarity (Campbell, Trapnell, Henie, & Katz, 1996).

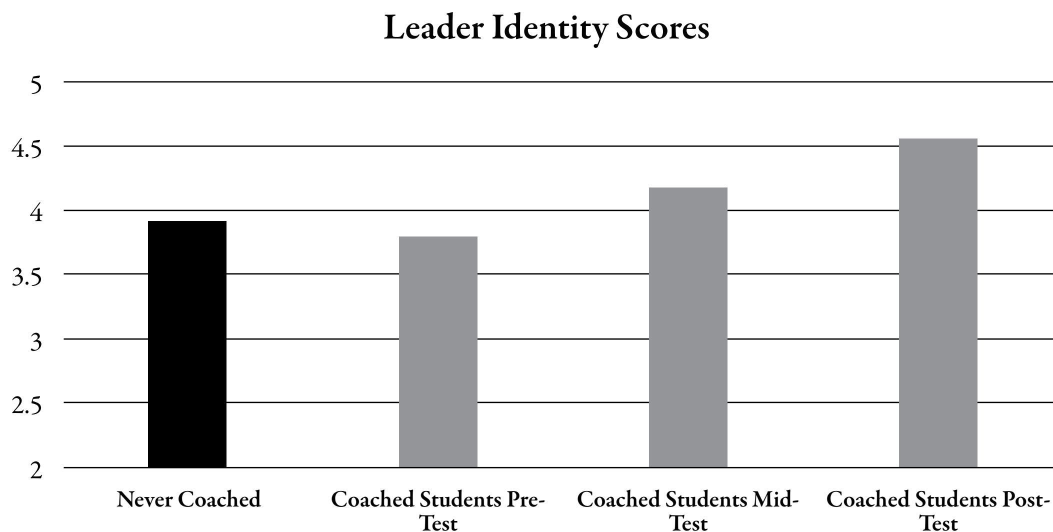
3. **Comparative Data.** Data from a campus-wide student survey, on which we have included the Authentic Leader Identity Scale, allow us to compare the leader identity scores of coached students to those of students who have never been coached and to examine the extent to which leader identity changes over the course of a student's college education.

4. **Independent Observational Data.** During one semester, students participating in the coaching program were solicited to recruit a friend, roommate, or teammate who knew them well enough to provide some observations about them. The Institute subsequently asked these acquaintances to evaluate how much growth they had observed in the student who had been coached across a range of variables that reflected common coaching goals, as well as two "foils" that were not expected to be student goals or to reflect secondary effects of leadership coaching (in other words, these foils provided discriminant validity as "non-dependent variables").

5. **Behavioral Data.** Finally, we have obtained behavioral impact data through a campus-wide, senior exit survey administered by the university. This survey asks graduating students to indicate which leadership roles they have held in the past year. The survey includes every such role available at the university, so seniors simply have to select the roles they have held. We have coded these leadership roles for the levels of leadership responsibility that they involve, using a coding system validated by a set of subject matter experts at the university. This coding system allows us to calculate an *emergent leadership experience* (or ELE) score for every senior. This ELE score gives us a behavioral index of formal leadership engagement through which we can evaluate one type of impact the Doerr Institute might have on students.

Pre-Post Developmental Change Data: Over multiple, large samples, we have found that students who worked with a leadership coach exhibit substantial changes in their leader identity scores over the course of a semester. We measure these students' leader identity scores at the beginning of the semester when they sign up to receive coaching. These scores are measured again in the middle of the semester on a campus-wide survey (we return to this survey's results shortly). These leader identity scores are measured a final time at the

Figure 1



Note: Scores on the Authentic Leader Identity scale range from 1 to 5 and reflect the average response across 9 items.

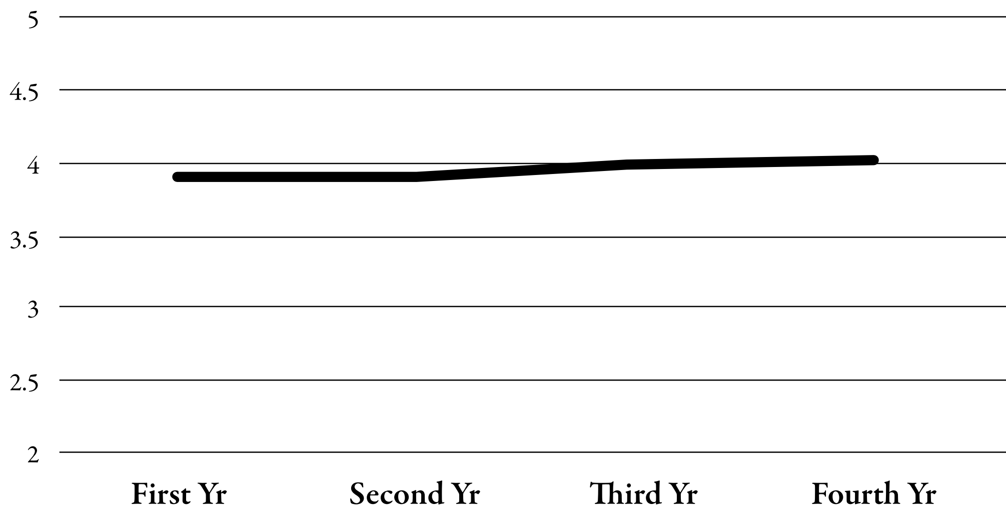
end of the semester as part of the final evaluation of the coaching experience, among a set of other measures. Mean leader identity scores for coached students at these three time points (before, during, and after coaching) are shown in Figure 1, alongside their non-coached peers (left-most bar). The growth in leader identity among coached students from pre to post is statistically significant ($t[183] = 20.04, p < .001$), substantial in effect size (Cohen's $d > 1.3$), and replicable over multiple semesters.

Comparative Data. These changes in leader identity parallel changes we have measured in psychological well-being, sense of purpose, and self-concept clarity. Specifically, satisfaction with life (Diener, et al., 1985) increased significantly from pretest to posttest[1], as did sense of purpose[2] (Steger, et al., 2006) and self-concept clarity[3] (Campbell, et al., 1996). These changes in primary and secondary outcomes are consistent with evidence from experimental studies on the effects of coaching outside of higher education, although our effects are somewhat larger than the results of some prior studies using older study participants (Burt & Talati, 2017).

In contrast to these results among coached students,

1 Pretest $M = 3.42$, $SD = 0.80$; Posttest $M = 3.88$, $SD = 0.76$; $t(179) = 9.59$, $p < .001$, $d = 0.72$.
2 Pretest $M = 3.43$, $SD = 0.93$; Posttest $M = 3.91$, $SD = 0.84$; $t(176) = 8.14$, $p < .001$, $d = 0.61$.
3 Pretest $M = 3.12$, $SD = 0.71$; Posttest $M = 3.62$, $SD = 0.75$; $t(178) = 10.19$, $p < .001$, $d = 0.76$.

Figure 2



Note: Sample includes over 2200 students who have not participated in one-on-one leader development coaching through the Doerr Institute.

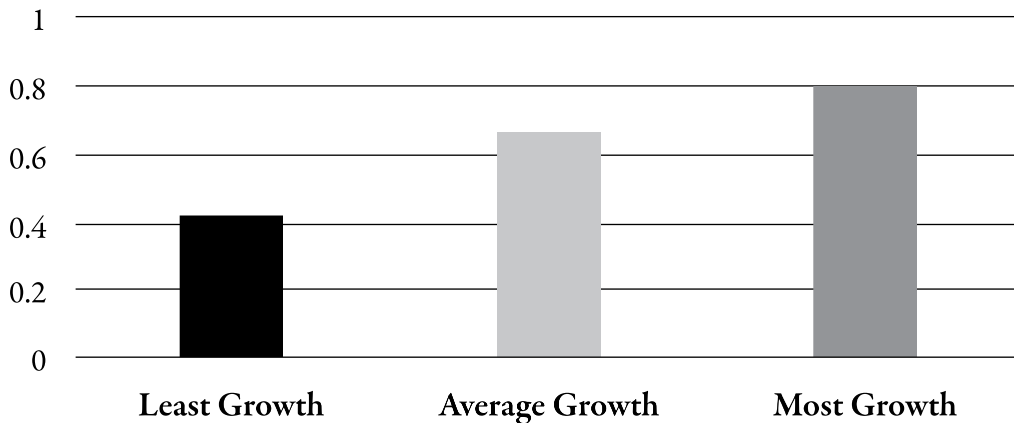
when we examined leader identity scores in a large sample of non-coached students, we found that leader identity does not change appreciably over time without intervention (see Figure 2). In a campus-wide sample of over 2,200 students who had *not* worked with a leadership coach, average leader identity scores of first-year students were barely distinguishable from those of sophomores, juniors, or seniors (although with these large sample sizes the overall ANOVA was statistically significant)[4]. What small differences emerged across year in school might even be attributable to selection bias in survey engagement or to selective institutional attrition. The implications of this rather flat line are worth pausing to ponder. Essentially, these data indicate that after four years of an elite college education, without purposeful intervention, students graduate with little more in terms of leader identity than they had as seniors in high school, expressing almost the same degree of self-confidence, self-awareness, and self-categorization as leaders that they had when they began their college education.

The problem of self-selection bias represents a threat to validity in our assessment work. Because participation in our programs is not mandatory for students, perhaps we simply take students who are already on a growth trajectory (for reasons that have nothing to do with us) and claim that their growth is the result of our intervention efforts. Although self-selection bias remains an ever-present concern for the measurement team, we have found that students who come to the Institute to work on their leadership abilities are an almost perfect representation of the overall student body in terms of basic demographics (e.g., gender, major, international vs. domestic origin, first-generation status, ethnicity) and personality (measured by the Big Five), as well as their tendency to be working on personal development goals (whether leadership-related or otherwise) on their own. In fact, the only two significant differences we have found between students who participate in leadership coaching and the general student population concern their leader identity scores and their desire to develop their leadership skills. Specifically, students who come to work with us have slightly lower leader identity scores compared to the broader student body, but they also have a slightly greater desire to develop as leaders (a motivation difference that is measured by a single item on a campus-wide survey). Importantly, we have also seen that this desire to develop as a leader was negatively related to increases in leader identity over a semester in a sample of over 100 students engaged in professional coaching. Thus, one of the only differences we have found between students who sign up to work with a leadership coach and the rest of the study body is actually predictive of less growth on our focal outcome measure. We are currently designing a randomized, wait-list-controlled study to examine the roles of motivation and formal intervention on changes in leader identity.

Independent Observational Data. Recently, we asked all of our professional coaches working with students to nominate the students who they believed grew the most over the course of the semester, as well as the students who grew the least (all other students received no nomination, so they are placed in the "average growth" category in Figure 3 below). The data below reflect changes in leader identity from before to after coaching from over 260 students. These data show that students who coaches believed grew the most also exhibited the largest changes in leader identity, followed by students in the average growth group, and then students who coaches believed grew the least. It is noteworthy that even students in the least-growth group still increased significantly in leader identity, although their growth was significantly less than that of students in the most-growth group. Note also that coaches did not have access to student self-report data, so their growth nominations were made independently of student self-reports.

4 Freshman ($M = 3.91$, $n = 845$); Sophomores ($M = 3.91$, $n = 638$); Juniors ($M = 3.99$, $n = 423$); Seniors ($M = 4.02$, $n = 371$); $F(3, 2273) = 4.00$, $p < .01$

Figure 3



Note: Change scores (post-test leader identity minus pretest leader identity) are statistically significant (i.e., different from 0) for all three groups.

Similarly, in another study we surveyed roommates, teammates, and friends (hereafter, "acquaintances") of students who were being coached at the beginning and end of the semester. These acquaintances made a series of observations at the end of the semester about the levels and types of growth that they had observed in their coached friends. Twenty-five complete friend pairs (coached and non-coached students) were sampled.

Examination of the observational ratings provided by acquaintances at the end of the semester provides additional validating evidence for the pre-post changes in leader identity we have measured. Acquaintances rated observed growth in coached students along dimensions that prior data indicated would likely represent common coaching goals within the sample, as well as two "foils" (enthusiasm for university athletics, and concern for the environment). Ratings range from 1 (*none at all*) to 7 (*a great deal*). Goal-related growth was, on average, significantly higher[5] than was growth on the foils[6], [7]. Except in the case of self-control and self-confidence, this was particularly true among students who actually identified the domain as being one of their goals (the left bars in Figure 4). It is noteworthy that all coached students were rated by their acquaintances as having grown in self-confidence, regardless of whether or not self-confidence was one of their goals.

Behavioral Data. Finally, we obtained behavioral data on the impact of leadership coaching via a campus-wide senior exit survey administered by Rice at the end of the spring semester for graduating

5 ($M = 4.6$)
6 ($M = 3.0$)
7 $t(24) = 4.98, p < .001$

students. This survey asks students to indicate which campus leadership roles they have held in the past year, and we coded these roles for the levels of leadership responsibility that they involve, using a coding system validated by a set of subject matter experts at the university. Using these codes, we can calculate an Emergent Leadership Experience (or ELE) score for every senior, which gives us a behavioral index of formal leadership engagement across the entire senior class.

To be clear, the Institute does not define leadership in a positional manner, nor does it equate "success" in any leader development program with the number of people who become presidents or CEOs. The Doerr Institute also does not encourage students to run for campus-wide offices or insist that they take on formal roles within student clubs or businesses. Nonetheless, if students are truly being developed as leaders, and if the coaching-related psychological changes we have described here are more than just internal shifts in students' personal narratives, then we should expect to see some students evidencing a greater willingness to step into formal leadership roles after working with the Doerr Institute. If their peers agree that they are ready to lead, then their greater personal willingness ought to translate into greater success in stepping into leadership roles with higher levels of responsibility.

For comparison purposes, we created a matched sample of graduating seniors to compare the ELE scores of students who had worked with a professional leadership coach with those of students who had not. Coached students were matched at a 1:2 ratio with non-coached students on gender, ethnicity, and major. The GPAs at graduation of these groups were incidentally identical (3.60 for both groups). Emergent leadership

Figure 4

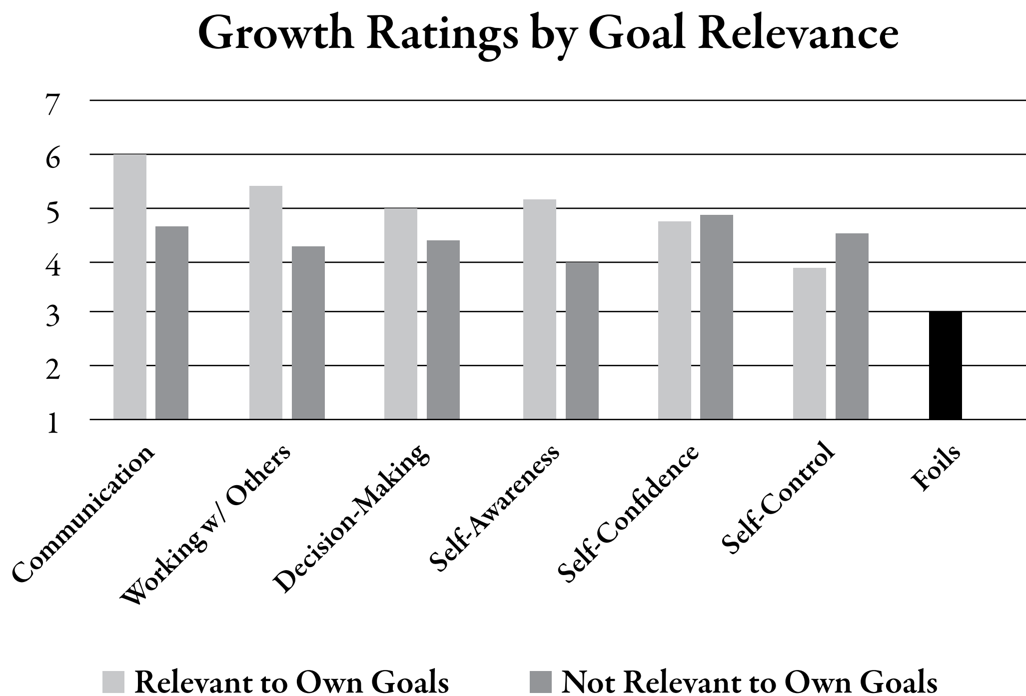
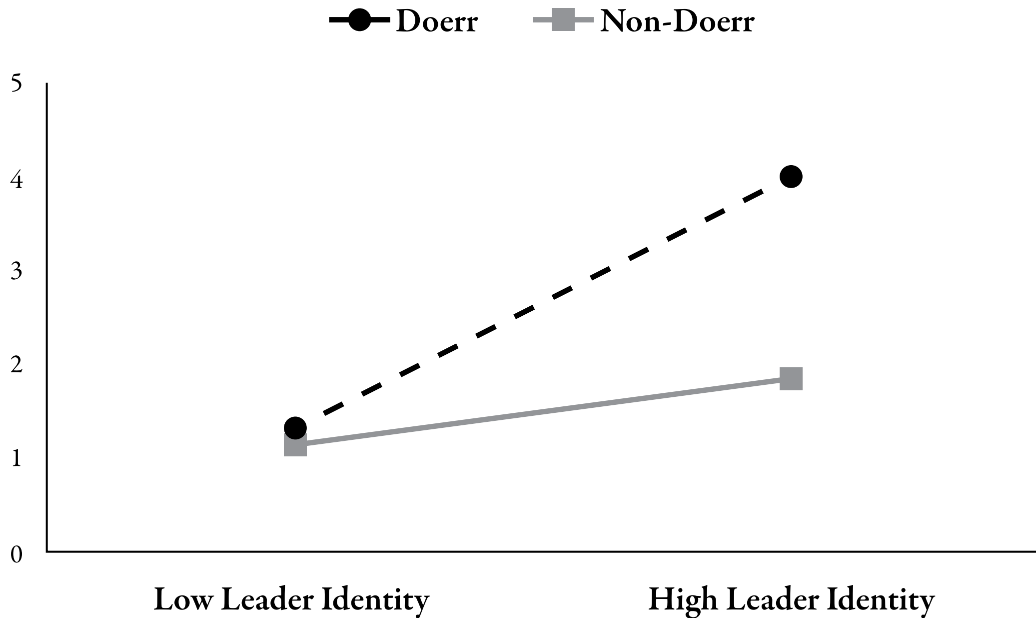


Figure 5



experience scores ranged from 0 to 17. An ANCOVA on ELE scores as a function of coaching status (coached vs. non-coached), controlling for gender, GPA, and international vs. domestic status, revealed a significant difference between students who were coached[8] and students who were not[9,10].

In data from a previous cohort, we had observed that less than half the graduating seniors (47%) had earned any ELE points — thus, all of the formal leading across the university is done by less than half the student body. Analysis of whether or not students earned any ELE points as seniors revealed that whereas 42% of non-coached students earned 1 or more ELE points, 61% of coached students did so (not including any students who were coached as seniors) — a substantial increase in levels of leadership responsibility, despite the lower starting levels of leader identity among students who come to the Doerr Institute seeking development.

Beyond this simple association between ELE scores and engagement with the Doerr Institute, we have found that leader identity scores (measured in the fall of students' senior year) and their ELE scores (measured at the end of the spring) are significantly associated with one another, but only among students who had previously engaged with the Doerr Institute (we did not include data from students who engaged with us during their senior year, as there would be no way for the Doerr Institute to have an impact on the

8 ($M = 2.45$, n = 174)
9 ($M = 1.49$, n = 384)
10 $F(1, 553) = 14.61, p < .001$

leadership roles held by those students as seniors). As Figure 5 below shows, students with stronger leader identity scores also held greater levels of leadership responsibility (as evidenced by their ELE scores)[11], but this association was weaker and not statistically significant among students who had not engaged with the Doerr Institute[12], (in other words, leader identity interacted with Doerr Institute engagement in the prediction of students' ELE scores[13]). Importantly, students with weak leader identities were very unlikely to have earned many ELE points, whether they worked with the Doerr Institute or not. Thus, defining oneself as a leader, having confidence to lead, and being self-aware of one's strengths and weaknesses as a leader (all of which are captured by leader identity) seems to be a prerequisite, though not sufficient on its own, to actually serving in high-level, formal leadership roles.

Conclusions, Future Directions, and an Exhortation

The Doerr Institute's mission is to enhance the leadership capacity of Rice students across the entire university. Central to this mission is the rigorous measurement of outcomes, which enables the Institute to avoid falling prey to many of the pitfalls that are endemic to such endeavors—including such problems as group think, the confirmation bias, and the Good Samaritan bias (assuming you are having the effects that you intend to have simply because you mean well).

As the preliminary evidence shows, the impact of just one of the Institute's programs, one-on-one leadership coaching, appears to be quite meaningful and crosses the domains of cognition, emotion, and behavior. Current and future projects will continue to explore and test the limits of these preliminary findings. One such ongoing study investigates some of the secondary benefits of one-on-one professional coaching described here. This study includes appropriate comparison groups that also take our measures of well-being and authentic leader identity, and it also includes a measure of a potential mediator of the apparent coaching benefits (changes in self-concept clarity). One of the comparison groups completes the same leader development plan that "Doerr students" work on with the help of their professional coach, so the inclusion of this element of the study will allow us to determine whether simply a little guided self-reflection and goal setting might be sufficient to produce at least some of the benefits that we have documented within our coaching program. The secondary benefits of leadership coaching that we have found also suggest the possibility of a variety of other, tertiary benefits, including benefits to academic performance, retention, and perhaps even athletic performance (for student athletes) that we plan to examine in the coming years.

Although this has not been the focus of our measurement efforts, the university itself might also experience some important benefits as a consequence of the Doerr Institute's programs, especially in the area of student recruitment. Given the competition among elite, selective universities for the highest caliber students, schools that offer such intensive leader development programs ought to realize a competitive advantage over those that do not, once the existence and merits of such programs become known to prospective students (and their parents). Anecdotally, we are beginning to see some evidence that this is the case at Rice (e.g., a 20% increase in applications to Rice in the last year alone, with explicit references to the Doerr Institute in student application essays), but more rigorous investigation is warranted.

Beyond the practical, competitive advantages that might accrue to schools that decide to take leader development more seriously (treating it as a core function of the institution and thinking carefully and systematically about desired outcomes that are then rigorously measured), we believe there is a moral dimension to doing so that should not be overlooked. Yes, many (if not most) schools claim to be developing

11 $\beta = 0.47, p < .001$
12 $\beta = 0.12, p = .095$
13 $\beta = 0.20, p < .01$

the next generation of leaders. And it is clear that colleges and universities house the next generation of leaders in campus dorm rooms, feed them in campus dining halls, and teach them in campus classrooms. But whether schools actually develop them as leaders is an entirely different question. If data from our own university are representative of higher education more generally, then this common claim about the development of students as leaders should be called into question. Without direct, empirical evidence to support their claims, universities should be held to account in the same way that we would hold a drug company accountable for claims about the effectiveness of its pharmaceutical products. Empirical claims should always be backed up by data. Without real outcome data, the claims made by universities about leader development are little more than empty promises (Kaiser & Curphy, 2013).

But the moral dimension of leader development concerns not just whether we are making fraudulent claims, but whether we are failing to do what we ought to be doing in higher education. The need for great leaders has never been greater than it is today. Nations have the capacity to destroy the world 10 times over, at the same time as international cooperation and alliance give way to creeping nationalism and protectionism. Even if humanity manages to avoid nuclear or biological self-destruction, climate change threatens to step in and destroy the planet more slowly, but just as surely. Solving such complex problems will require strong leadership from many quarters, not just within a single nation, and even if we were to solve all of the major problems facing humanity today, the next generation would be certain to face its own set of new problems tomorrow. Surely we can do better than we are doing now to prepare this next generation of leaders in our institutions of higher education, so they are truly ready to take the helm as leaders when their time comes. Raising our game to match the rising stakes of poor leadership will require a commitment to specifying and measuring objectives, but we already know how to do this. We simply have to commit to treating leader development like we treat other types of training and development and make leadership a priority.

◆ ◆ ◆

References

Ashkanasy, N. M., & Daus, C. S. (2005). Rumors of the death of emotional intelligence in organizational behavior are vastly exaggerated. *Journal of Organizational Behavior, 26*, 441-452.

Avolio, B., Reichard, R., Hannah, S., Walumbwa, F., & Chan, A. (2009). A meta-analytic review of leadership impact research: Experimental and quasi-experimental studies. *The Leadership Quarterly, 20*(5), 764-784. doi: 10.1016/j.leaqua.2009.06.006

Burt, D., & Talati, Z. (2017). The unsolved value of executive coaching: A meta-analysis of outcomes using randomised control trials. *International Journal of Evidence Based Coaching and Mentoring, 15*, 17-24.

Campbell, J. D., Trapnell, P. D., Heine, S. J., & Katz, I. M. (1996). Self-concept clarity: Measurement, personality correlates, and cultural boundaries. *Journal of Personality and Social Psychology, 70*, 1, 141-156.

Chan, K. Y., & Drasgow, F. (2001). Toward a theory of individual differences and leadership: Understanding motivation to lead. *Journal of Applied Psychology, 86*, 481-498.

Day, D. V., Fleenor, J. W., Atwater, L. E., Strum, R. E., & Mckee, R. A. (2014). Advances in leader and leadership development: A review of 25 years of research and theory. *The Leadership Quarterly, 25*, 63-82.

Day, D. V., & Harrison, M. M. (2007). A multilevel, identity-based approach to leadership development. *Human Resource Management Review, 17*, 360–373.

Day, D. V., Harrison, M. M., & Halpin, S. M. (2009). *An integrative approach to leader development: Connecting adult development, identity, and expertise*. New York, NY: Taylor & Francis.

DeRue, D. S., & Wellman, N. (2009). Developing leaders via experience: The role of developmental challenge, learning orientation, and feedback. *Journal of Applied Psychology, 94*, 859-875.

Diener, E., Emmons, R. A., Larsen, R. J., & Griffin, S. (1985). The satisfaction with life scale. *Journal of Personality Assessment, 49*, 71-75.

Ely, K., Boyce, L., Nelson, J., Zaccaro, S., Hernez-Broome, G., & Whyman, W. (2010). Evaluating leadership coaching: A review and integrated framework. *The Leadership Quarterly, 21*(4), 585-599. doi: 10.1016/j.leaqua.2010.06.003

Feldman, D., & Lankau, M. (2005). Executive coaching: A review and agenda for future research. *Journal of Management, 31*(6), 829-848. doi: 10.1177/0149206305279599

Grinnell, R., Gabor, P., & Unrau, Y. (2015). *Program evaluation for social workers* (7th ed.). New York, NY: Oxford University Press. Harvard Business Publishing (2016). *Research report: The state of leadership development*. Harvard Business Publishing Corporate Learning: Brighton, MA.

Hiller, N. J. (2005). *An examination of leadership beliefs and leadership self-identity: Constructs, correlates, and outcomes*. Unpublished doctoral dissertation. University Park, PA: Pennsylvania State University.

Harvard Business Publishing (2016). Research report: The state of leadership development. Harvard Business Publishing Corporate Learning: Brighton, MA.

Kaiser, R. B., & Curphy, G. (2013). Leadership development: The failure of an industry and the opportunity for consulting psychologists. *Consulting Psychology Journal: Practice and Research, 65*, 294-302.

Kirkpatrick, D. (2009). *Evaluating training programs*. San Francisco, CA: Berrett-Koehler.

Lacerenza, C. N., Reyes, D. L., Marlow, S. L., Joseph, D. L., & Salas, E. (2017). Leadership training design, delivery, and implementation: A meta-analysis. *Journal of Applied Psychology, 102*, 1686-1718.

O'Boyle Jr., E. H., Humphrey, R. H., Pollack, J. M., Hawver, T. H., & Story, P. A. (2011). The relation between emotional intelligence and job performance: A meta-analysis. *Journal of Organizational Behavior, 32*, 788-818.

Steger, M. F., Frazier, P., Oishi, S., & Kaler, M. (2006). The meaning in life questionnaire: Assessing the presence of and search for meaning in life. *Journal of Counseling Psychology, 53*, 80-93.

Stein, S. J., & Book, H. E. (2011). *The EQ edge: Emotional intelligence and your success* (3rd ed.). Ontario: John Wiley & Sons.

Wilson, T. D. (2011). *Redirect: The surprising new science of psychological change*. New York: Little, Brown, and Co.

Appendix: The Authentic Leader Identity Scale

The Authentic Leader Identity scale is a brief integration of multiple facets of a person's leadership self-construal, including self-categorization as a leader, leadership self-efficacy, motivation to lead, authenticity, and self-awareness of one's strengths and weaknesses as a leader. In a sample of over 2,800 students, these 9 items exhibited a largely unidimensional structure in a principal axis factor analysis, with weaker secondary factors distinguishing items 1-4 from items 5-9. For examples of related measures of leader identity, see Chan and Drasgow (2001), Hiller (2005), and Day et al. (2009).

Response scale: 1 (*Disagree strongly*) to 5 (*Agree strongly*)

Items (α = **.89**)**:**

1. I see myself as a leader.
2. I feel confident to lead when opportunities arise.
3. I have a desire to pursue roles in which I can be a leader.
4. I have a clear understanding of my strengths as a leader.
5. I feel confident enough in my personal convictions that I would assert them even if it meant disagreeing with friends, teammates, or colleagues.
6. I am comfortable expressing an unpopular position when I feel it is appropriate.
7. I act in ways that are consistent with my values.
8. I understand the ways that my weaknesses as a leader can affect others.
9. I have a clear sense of my values and core beliefs.