Engineering Educators Bringing the World Together
2025 ASEE Annual Conference & Exposition
Palais des congrès de Montréal, Montréal, QC • June 22–25, 2025
Paper ID #45681

# Advancing a Multi-year Longitudinal Assessment Approach for an Engineering Leadership Program: A Work in Progress

**Dr. James N Magarian, Massachusetts Institute of Technology**

James Magarian is a Sr. Lecturer with the Gordon-MIT Engineering Leadership (GEL) Program and serves as Academic Director for the MIT School of Engineering Programs in Technical Leadership and Communication. He joined MIT and GEL after nearly a decade in industry as a mechanical engineer and engineering manager in aerospace/defense. His research focuses on engineering workforce formation and the education-careers transition.

**Alexander Rokosz, Massachusetts Institute of Technology**

Alex Rokosz is a Program Analyst with the Gordon-MIT Engineering Leadership (GEL) Program and MIT School of Engineering Programs in Technical Leadership and Communication. He joined MIT and GEL in 2023.

# Advancing a Multi-year Longitudinal Assessment Approach for an Engineering Leadership Program: A Work in Progress


## Abstract

This paper presents the design and evaluation of a longitudinal assessment survey for a university engineering leadership program. We review the self-efficacy assessment approach employed in the survey, including its methodological basis and its alignment with the program learning goals and curriculum. We also review similar assessment approaches employed by other engineering leadership programs, discussing areas of commonality with the present approach as well as rationales for customization. We present findings from our evaluation of the present survey instrument from its initial deployments to program participants (n = 420), including confirmatory factor analysis and internal consistency checks. Evaluating goodness of fit using a nested model comparison approach, we find that an eight-factor scheme aligned with engineering leadership capability categories from the program's curriculum exhibits good fit and demonstrates acceptable (or better) internal consistency. Further, we present an example longitudinal analysis using self-efficacy data collected via the survey instrument, assessing student development between pre- and post- survey instances for one segment of the program. From this example analysis, we demonstrate an approach for displaying consolidated results graphically. We discuss drawing insights from these findings about comparative strengths in cohorts' developmental outcomes, as well as areas of intended learning to target for improvement. Finally, we discuss next steps in the deployment of this longitudinal assessment survey, which include extending its use across a larger range of longitudinal time points (spanning program segments and into the alumni years), and its expansion to serve adjacent engineering leadership and professional skills programs at the same university.


## Introduction

With university Engineering Leadership (EL) programs continuing to launch and grow in recent years [1, 2], several contemporary studies have discussed longitudinal assessment as a means for these programs to evaluate their students' development of engineering leadership capabilities over time [3-5]. Other studies, meanwhile, have introduced alumni assessments as a way to examine EL programs through alums' career achievements and career preparedness [6-8]. Yet, most EL programs have been operating for relatively short durations. With fewer than 10 of today's active programs in North America existing prior to 2010 [1], there have been few opportunities for programs to conjoin these two approaches into long-term longitudinal studies where same-participant assessments from the university years are coupled with assessments at time points substantially into the career years. This Work in Progress Paper presents intermediate observations from such a study underway at the Gordon-MIT Engineering Leadership Program (GEL) at Massachusetts Institute of Technology. Here we share findings from survey instrument evaluation and from the instrument's initial use in longitudinal (pre- and post- program segment) assessment. We then outline plans for its expansion to additional times in the longitudinal sequence, summarize the approach's limitations, and review plans to further expand it for use in adjacent leadership and professional skills development programs at the same university.

### *Program overview and curricular origin*

GEL, established in 2007, is an undergraduate certificate program for juniors and seniors at MIT. Students elect to take the program either as a one-year or two-year course of study, with the two-year "advanced" track distinguished by additional peer leadership opportunities and coursework. The program is co-curricular for most participants, though some departments have begun recognizing the coursework for engineering elective credit. The structure of the program's first year, described by [9], consists of three core components: 1) a weekly Engineering Leadership Lab (ELL), where students work in small teams to face leadership challenges rooted in capabilities from the program's curriculum [10], 2) an Engineering Leadership seminar-style class, synchronized with the ELL, where students study the academic background of leadership capabilities prior to a given ELL and discuss lessons-learned from the previous week's ELL, and, 3) one from several elective courses that fulfill a Design and Innovation Leadership Requirement focused on the engineering design process and its inherent teamwork and leadership components. The total student workload for those in the program's first year, typically undergraduate juniors, is approximately that of 1.5 full credit MIT courses. The program's second year, typically undertaken by undergraduate seniors, constitutes an additional workload approximately equal to two more full credit courses and is also described by [9]; here, students serve as "team coaches" for the first-year students' ELL teams. These second-year students take turns facilitating the ELL activities, a responsibility for which they receive coaching and instruction from the GEL teaching staff. The second-year students also undertake a short course in project management and select an additional leadership-related elective course to take.

The foundation of GEL's curriculum, *Capabilities of Effective Engineering Leaders* [10], was developed soon after the program's launch as a consensus report from workshops involving engineering and leadership educators, leadership specialists from the military, and practicing engineering leaders. This report was also motivated by then-recent engineering accreditation criteria revisions that integrated learning outcomes in nontechnical areas [11], while the program's experiential learning-based structure drew from a benchmarking report on global EL education best-practices [12]. As this curriculum was operationalized, efforts were concurrently undertaken to design learning assessments that aligned with the engineering leadership capabilities (e.g., [13]). This initial approach did not enable long-term tracking of outcomes and its assessment scope was subject to adjustment in the years that followed as the capabilities prioritized in the program's required core courses settled out (i.e., versus elective courses); nonetheless, GEL's present method of self-efficacy-based assessment [14], aligned with components from the *Capabilities*, traces its origins to the program's early days.

*Present status of the longitudinal assessment initiative*

In recent years, GEL has advanced its assessment instrument to include multiple self-efficacy assessment scale items grouped into engineering leadership capability categories based upon the *Capabilities*. We report in this paper on an evaluation of this instrument in accord with published practices for self-assessment survey development [15, 16], including factor analysis and internal consistency checks. Following from GEL's multi-year longitudinal assessment strategy described previously [17], we present current in-process findings from the application of this assessment instrument to the first of several planned longitudinal segments: pre-/post- assessment for year one of the program. Since growth of students' self-assessed leadership capabilities can be non-linear, with dips and recoveries over time with acquired experiences [4], we introduce a method incorporated in this assessment system for participant tracking across multiple follow-on assessment instances: an anonymous self-generated participant code [18]. This anonymous code enables participants to connect future assessment surveys to past ones without the need to input identifying information or to retain a log-in credential. This method allows researchers to stitch together assessments from several time points and to distinguish short-term effects from longer-term trends. While single-segment assessments therefore have limitations, we present such findings here as a case example to illustrate a data visualization approach for comparatively assessing change across multiple assessed capability categories simultaneously. This approach enables a comparative program strengths analysis that can also be applied to a wider time range. Further, in *Future work*, we discuss the intended sequence of follow-on student and alumni assessment instances that will complete GEL's long-term recurring plans for program assessment connecting student development and career outcomes.

## Background

### *EL programs' approaches to assessment*

Given the challenges of assessment in EL relative to traditional subjects that can more readily assess learning through exams and student deliverables, recent research suggests that EL programs' assessment approaches are often rooted in one of three measurements types: self-efficacy measurement (e.g., [13, 19, 20]) skill or competency self-assessment (e.g., [5, 21-23]), or career-related outcomes assessment (e.g., [6-8]). The self-efficacy approach involves measuring students' beliefs about their abilities to carry out designated types of performances, sometimes referred to as "task-specific self-confidence" [24, p.1]. Self-efficacy measures typically employ survey items with 0%-100% confidence scales in 10% increments [14]. Items are customized to measure confidence in context- and domain-specific activity definitions that are aligned to components of a curriculum [24]. Other researchers, meanwhile, have developed engineering leadership competency items that employ Likert-type scales rather than confidence scales (e.g., [5, 22, 23]). Similarly, these items assess participants' level of agreement with statements that are descriptions of personal competence (e.g., "I can organize and structure a group to accomplish a common goal" [5, p.126]). In examples spanning both the self-efficacy and competency approaches, developers of assessment instruments have employed analytical categorization methods to organize larger quantities of individual survey items into a more succinct number of multi-item factors (e.g., four [5], six [20], etc.) that each assesses a relevant capability area. Beyond these approaches rooted in self-assessment, others have deployed assessment instruments to measure downstream outcomes related to career achievement, typically during the early alumni years (e.g., [6-8]). For instance, programs have measured the extent to which alumni have attained positions and responsibilities related to program aims [6, 8], or, through qualitative analysis, the extent to which leadership themes are apparent in alums' work and careers [7]. Examples from among these career-focused assessments have also measured the extent to which respondents feel, retrospectively, that their time in the EL program supported their career achievement [6, 8].

### *Considerations for self-efficacy assessment in EL programs*

The assessment instrument we examine in this paper follows the self-efficacy approach for pre-, mid-, and post- program learning evaluation. Several qualities of this method led to its adoption at GEL. First, though methods based upon student self-evaluation have raised validity concerns [25,26], researchers have found the self-efficacy approach can be valid and reliable when applied to appropriate measurement scenarios (e.g., [20, 27-30]); in particular, when it is employed to assess confidence in specific abilities situated in present-day, familiar performance contexts for the participant [14]. Further, validity of self-efficacy assessment instruments may be strengthened by the incorporation of "practice items" [14, p. 313] that precede measurement items and that familiarize respondents with the self-evaluation process by prompting them to first assess their confidence in highly recognizable and basic tasks (e.g., [29, 30]). Lastly, given self-efficacy's emphasis on item design that incorporates context- and domain-specificity, this assessment approach can be readily tailored to align with learning objectives in an experiential learning-based curriculum [14, 24]. At GEL, we leveraged the concurrency of curriculum and assessment design to develop self-efficacy scale items that align with capability definitions from the *Capabilities*.

*Conducting EL program assessments longitudinally*

Complex growth trajectories in students' self-assessed leadership capabilities – for instance, trajectories where mid-program assessments stagnate or dip below pre-program assessments before rising in post-program assessments – are discussed as another assessment challenge in EL programs [4, also citing: 26, 31, 32]. Whether such growth patterns exist appears to relate to the magnitude of students' incoming assessments and to the measures being examined [4]. Nonetheless, these findings have led researchers to observe that "the most accurate change in [self-assessment] ratings may be between the mid- and post-assessments, after raters have an opportunity to calibrate their views based on class experience" [4, p. 17]. Similarly, scholars of self-efficacy assessment emphasize the role of learners' "calibration" experiences in enabling accurate assessments, finding that "individuals must have realistic or accurate perceptions of their ability for a given task" [24, p. 82]. Given the likelihood that students enter EL programs at differing levels of pre-existing abilities and with differing prior calibration experiences, it follows that researchers have suggested longitudinal approaches to EL assessment [3-5]. Employing a longitudinal approach allows learning trajectories established over multiple checkpoints to form the primary basis of assessment and comparison, avoiding a reliance on potentially fragile single time point or single time segment learning outcomes measurements.

*Development of self-efficacy scale items for GEL*

Following from the program's capabilities-based curriculum [10], GEL utilized self-efficacy scale design guidelines [14] to develop assessment items that align with capability action descriptions. In the assessment model examined in this paper, these items, 29 in total (shown later in Table 2), are organized into a hypothesized set of eight capability categories. The category scheme follows that of the *Capabilities* [10], yet with the two largest original categories (by item count) divided into smaller component categories that reflect how learning in related areas in the program is organized in practice. This results in eight capability categories: *Initiative and accountability*, *Team building*, *Self-awareness*, *Sensemaking*, *Communicating across differences*, *Communicating to build trust and to influence*, *Visioning*, and *Implementing and delivering*. Since this basis for these groupings of related items comes from an established curricular structure, itself based on consensus workshops, our evaluation of the assessment model follows a confirmatory factor analysis (CFA) approach that tests this hypothesized grouping of the items [16]. Our overall examination of the assessment instrument, as discussed in *Methods*, includes both this evaluation of the structure of the model (i.e., CFA) and internal consistency checks for each factor in the model based on empirical data.

## Methods

*Participant recruitment and longitudinal response matching*

All GEL students and alumni are periodically invited to participate in survey-based assessment as part of the longitudinal assessment plan described in an earlier stage of this project [17] and in compliance with a determination from MIT's Institutional Review Board. Findings examined in this present paper are from pre- and post- GEL Year-1 survey sessions conducted between September 2021 and May 2024. All sections of survey questions composing the longitudinal assessment are hosted via Qualtrics XM survey software. Any participants accessing this assessment system, ranging from incoming undergraduate students through mid-career alumni, begin at the same survey welcome/consent screen, and are then routed to appropriate subsets of questions based upon class year and academic status information they provide. For the pre- and post- Year-1 survey instances examined here, students were provided 15-minutes of in-class time to complete the survey. The welcome/consent screen informs participants that the survey is voluntary and anonymous, and that it is part of a sequence of surveys where same-participant responses will be aggregated. If consent is provided, the survey asks participants to generate an anonymous Self-Generated Identification Code (SGIC) in accord with methods described in [18]. This SGIC approach prompts respondents with a series of questions that are each designed to produce one character of a six-digit code based on inquiring about information that is likely to be enduring and remembered (e.g., "enter the number of older siblings you have"), resulting in an anonymous tracking code that becomes part of each survey record. The pre-/post- Year-1 self-efficacy change analysis presented in this paper utilized these SGICs to match same-student pre- and post- year responses.

*Self-efficacy measurement and reporting*

Students provided self-efficacy ratings as part of both the pre- and post- Year-1 segments of the longitudinal survey. Just prior to entering their ratings, respondents were shown a practice item (i.e., "How confident are you in your current ability to always arrive at your meetings on time?") to familiarize them with the rating system and confidence scale (i.e., a 10-point confidence scale, spanning 0-100% in 10% increments) [14]. Responses were then recorded in Qualtrics for the 29 items composing the assessment instrument.

Tracking changes in student self-efficacy ratings (pre-/post- program or program segment) enables comparisons across ratings to reveal which capability areas exhibit the largest and smallest changes over time. However, a simple comparison of change magnitudes does not account for differences in average starting values among ratings. Incoming student ratings that are initially high (relatively speaking) in certain areas may be less likely to exhibit as high a magnitude of further positive change compared to areas that are initially low. We have therefore developed an approach for visualizing results so that both overall magnitudes of ratings and the magnitudes of pre-/post- change can be viewed simultaneously. To accomplish this, we present self-efficacy results in scatter plots where the x-axis represents the mean self-efficacy score of each item post-hoc and the y-axis represents the mean change in self-efficacy scores between the instances of the assessment being compared. Y-axis values are determined by computing averages of same-student comparisons: here, we subtract the initial instance self-efficacy survey score from the subsequent score for each item, and for each respondent, before computing mean change across respondents. Statistical inference to assess the significance of changes in self-efficacy is also possible by the Wilcoxon signed-rank test, accounting for paired ratings coming from the same respondent. We present an example set of self-efficacy change analyses with significance levels in Appendix Table A1; the significance levels themselves are not a focus of this paper's discussion because they are limited to a single segment of a longer intended longitudinal sequence that will be examined in future work.

*Confirmatory factor analysis and consistency checks*

To perform confirmatory factor analysis, we utilized the *SEM* (Structural Equation Modeling) command in Stata. We mapped the 29 items to their respective eight factor groupings and computed several conventional measures of goodness-of-fit to examine the factor groupings in our model: root mean square error of approximation (RMSEA), comparative fit index (CFI), Tucker-Lewis Index (TLI) and the standardized root mean square residual (SRMR) [33, 34]. We assessed the factors to determine whether all items in each exhibit at least a minimum factor loading of 0.4 [35, 36]. We also assessed the hypothesized model form by comparing our proposed eight-factor scheme to a single-factor scheme through a $\chi^2$ difference test. By calculating the difference between the $\chi^2$ of the single-factor and eight-factor schemes and assessing its significance, it is possible to determine whether the eight-factor scheme exhibits the expected better model fit relative to the single-factor scheme. Finally, we assessed internal consistency for each factor by examining whether each exhibited a Cronbach's alpha value exceeding a minimum of 0.7 [37, 38].

*Response sample and data handling*

To achieve the minimum recommended sample size for CFA, $\geq$ 300 respondents [39, 40], it was necessary to aggregate assessment survey data from multiple academic years. Since GEL has a typical Year-1 class size of 130-150 students, and with some attrition expected, we employed a three-year aggregated sample for CFA from academic years AY2021-22, AY2022-23, and AY2023-24. Forming this multiyear sample for CFA was accomplished algorithmically by extracting the first instance of self-efficacy ratings data from each individual in the overall dataset from that span (yielding n=420); thus, no repeat ratings from the same individual are contained within this sample employed for CFA. Forming the sample employed for this paper's second purpose, the example pre-/post- program segment self-efficacy analysis, however, required a different method. Here, matched pairs of same-individual pre- and post- ratings needed to be aggregated. We employed the SGIC to identify the same-individual pairs of ratings in the dataset from the same time span, yielding a smaller sample of n=132 matched pairs. A discrepancy in the sizes of these two subsamples drawn from the AY2021-22 - AY2023-24 data is expected for several reasons related to conditions requisite for achieving perfect survey pair matches. First, the data from any given survey instance are incomplete (relative to the enrolled student population) due to both voluntary response rate and imperfect attendance at the in-person survey session, resulting in an average rate of missing responses of 20% at a given survey instance. Secondly, the program typically experiences student attrition, averaging 33% student loss between survey instances (i.e., spanning the start to end of an academic year) across our multiyear sampling period. Lastly, given that it is voluntary and unenforceable, the use of the anonymous SGIC for longitudinal matching comes with a known trade-off of an expected data loss of 10%-20% [18] from a given survey instance due to SGIC user input errors or omissions, despite a checkbox prompt for all respondents to verify their SGIC. These issues compound to reduce the number of same-student perfect matches across two survey instances relative to the total number of individuals completing at least one survey instance.

Among the superset of n=420 unique individuals in this study, 76.2% provided voluntary demographics information at the incoming survey instance (notably, the longitudinal survey sequence does not inquire again about demographics after the first instance). Table 1 presents a summary of participant demographics. The GEL program cohort has tended to be diverse in its representation of historically underrepresented demographics in engineering. Here, for instance, we find that women compose 54.4% of respondents, Black or African American students compose 9.6%, and Hispanic or Latinx students compose 17.6%. Recently reported (Fall 2024) representations of these groups in the undergraduate population at MIT are 48.2%, 7.7%, 14.1%, respectively [41]. An area where the GEL cohort appears to under-represent individuals relative to the MIT population is in international student status, where we found 8.8% of respondents to be international students (compared to 11.7% of undergraduates university-wide).

**Table 1.** Participant demographics

| | $n^1$ | % | | $n^1$ | % |
|---|---|---|---|---|---|
| Year[2] | | | Race/Ethnicity[3] | | |
| Junior | 369 | 87.9 | Asian or Asian American | 128 | 40.9 |
| Senior | 51 | 12.2 | Black or African American | 30 | 9.6 |
| | | | Hispanic or Latino/a | 55 | 17.6 |
| Gender | | | Native American or Alaska Native | 3 | 1.0 |
| Woman | 174 | 54.4 | Native Hawaiian or Pacific Islander | 2 | 0.6 |
| Man | 141 | 44.1 | White | 136 | 43.5 |
| Non-binary | 5 | 1.6 | Other | 5 | 1.6 |
| International Student (non-U.S.) | | | First Generation College Student | | |
| No | 290 | 91.2 | No | 243 | 76.7 |
| Yes | 28 | 8.8 | Yes | 74 | 23.3 |

Notes: 1. Total observation counts differ across demographic categories due to missing responses to voluntary questions
2. Year is based on data provided in the incoming survey
3. Respondents can select multiple races and ethnicities

## Results

### *Assessment instrument evaluation*

Table 2 presents components of the assessment instrument, with its 29 items grouped by factor. Factor loadings are shown for each item, with each found to exceed the minimum acceptable factor loading value of 0.4 [35, 36], indicating sufficient correlation between items and their associated factors. We also note acceptable (or better) measures of internal consistency for each factor, as evidenced by Cronbach's alpha values exceeding 0.7 for each [37, 38]. Further, Table 2 introduces abbreviated labels for the factors and items that are referenced later in this *Results* section.

We next examine the model-level goodness-of-fit indices for our eight-factor model, comparing them to those for a baseline model that maps all 29 items to a single factor (Table 3). Both models have a high and significant $\chi^2$, indicating we cannot presume the impossibility of a better-fitting model configuration in either case. However, a test comparing $\chi^2$ between these nested models, as well as an examination of other fit indices, allows us to determine the acceptability of fit of the proposed eight factor model [42, 43]. These alternate fit evaluations are appropriate because the $\chi^2$ test is highly sensitive to large sample sizes; it tends to imply a mis-fit model for large samples, even if the mis-fit is trivial, and is therefore conventionally augmented by other tests [42, 44]. In the case of evaluating our eight-factor model against the baseline model, a nested model comparison $\chi^2$ test indicates the eight-factor model is a significantly better fit. Here, lower $\chi^2$ indicates better fit [42], and the nested model comparison finds that $\chi^2$ is significantly reduced by 771.0 ($p < 0.01$) in the eight-factor model relative to the single-factor model. Next, we observe that other conventional fit measures applied to the eight-factor model also indicate good fit: acceptably low values of RMSEA (0.07) and SRMR (0.06) [33, 45], and acceptably high values of CFI (0.91) and TLI (0.90) [34, 45], though we note that the latter is at the conventional minimum. Further, we find these fit index values to be comparable to those from a recent evaluation of a similarly configured assessment instrument in another engineering education context [44]. In contrast, the RMSEA, SRMR, CFI, and TLI fit indices for our single-factor baseline model are poorer in all cases, indicating unacceptable fit of that model configuration (Table 3). From these findings, we infer that the eight-factor model has fit characteristics acceptable for its use as an engineering leadership capabilities assessment.

**Table 2**: Items and factors in engineering leadership assessment

| Factor | Item Label | Item | Cronbach's α | Factor Loading |
|---|---|---|---|---|
| | | How confident are you in your current ability to do the following?: | | |
| Initiative & Accountability (IA) | | | 0.76 | |
| | a | To step forward and take responsibility for an activity that's critical to your project's mission when others fail to get started on it. | | 0.73 |
| | b | To make the best possible decision in a situation where you have incomplete or uncertain information. | | 0.71 |
| | c | To uphold your commitments to your team, or to promptly coordinate mutually acceptable changes to commitments, when the conditions you face become more difficult than you expected. | | 0.67 |
| | d | To always contribute your best efforts as a team player, based on your personal strengths and skills in teamwork. | | 0.57 |
| Team Building (TB) | | | 0.81 | |
| | e | To contribute toward creating a team environment where all team members feel a sense of belonging, including among those who are different from you. | | 0.66 |
| | f | To recruit and select new team players who bring complementary skills that are needed by the team. | | 0.73 |
| | g | To lead the establishment of shared norms that enhance the performance of your team. | | 0.80 |
| | h | To lead post-project reviews so that your team can learn from experiences and improve performance on future projects. | | 0.66 |
| Self-Awareness (SA) | | | 0.85 | |
| | i | To identify and understand your own personal values. | | 0.68 |
| | j | To make effective decisions about your career and about changes in career direction. | | 0.83 |
| | k | To make life choices and changes that are aligned with a personal vision and meaningful sense of purpose. | | 0.91 |
| | l | To consistently act upon your personal values in situations where they are challenged. | | 0.68 |
| Sensemaking (SM) | | | 0.79 | |
| | m | To recognize when a situation's complexity warrants creating a system model to aid understanding. | | 0.64 |
| | n | To translate customer needs into technical requirements that engineers will interpret effectively. | | 0.78 |
| | o | To navigate the internal organizational politics of your organization to create a project plan that is acceptable to internal and external decision-makers. | | 0.77 |
| | p | To seek out multiple perspectives when defining problems. | | 0.60 |
| Communicating across Differences (CD) | | | 0.71 | |
| | q | To listen very carefully to a team member with whom you disagree, rather than solely thinking about what you are going to say next. | | 0.54 |
| | r | To communicate a technical concept to senior leaders in your organization who do not share your technical background. | | 0.70 |
| | s | To help two of your fellow team members with strong differences of opinion reach an agreement. | | 0.78 |
| Communicating to Build Trust and to Influence (CT) | | | 0.74 | |
| | t | To provide a team member with constructive criticism that improves their performance. | | 0.68 |
| | u | To advocate for and receive additional resources for your team when only limited resources are available. | | 0.72 |
| | v | To build positive working relationships with your supervisor and fellow team members. | | 0.69 |
| Visioning (VN) | | | 0.89 | |
| | w | To lead your team's formulation of a shared mission and vision for a project. | | 0.90 |
| | x | To launch your team in ways that motivate team members to achieve the goals, scope, and deadlines of your project at the project's start. | | 0.89 |
| | y | To identify the critical questions that need to be answered before the best solutions can be selected on a project. | | 0.78 |
| Implementing & Delivering (ID) | | | 0.87 | |
| | z | To determine in the first week of a project if your team is lacking in competencies or resources that will be needed. | | 0.73 |
| | aa | To convince your team to discard a really novel idea and instead use a satisfactory idea that can more assuredly be delivered on time. | | 0.75 |
| | ab | To detect and resolve gaps in your team's preparedness for a major project milestone, such as a product demonstration or test. | | 0.86 |
| | ac | To convince your team to agree that a major change is needed part-way through a project without negatively impacting team motivation. | | 0.81 |

**Table 3.** Model fit statistics

| | 1-Factor Model | 8-Factor Model |
|---|---|---|
| $\chi^2$ | 1679.97*** | 908.97*** |
| RMSEA | 0.10 | 0.07 |
| SRMR | 0.06 | 0.06 |
| TLI | 0.78 | 0.90 |
| CFI | 0.79 | 0.91 |
| Degrees of Freedom | 377 | 349 |

Note: ***p< 0.001

*Self-efficacy change analysis and visualization*

For illustrative purposes, we next present a self-efficacy change analysis using data collected via the eight-factor instrument. Here, we examine self-efficacy over the GEL Year-1 segment from the same-student pre-/post- segment matched pair ratings (n = 132). Figure 1 presents a scatter plot of the mean rating outcomes for items composing each of the eight factors, employing the x- and y-axis scheme described in *Methods*. A linear trend line applied to these data is downward sloping (slope = -0.34). This negative slope indicates a trend where the higher a self-efficacy score is initially, the less it is expected to positively change over the course of a program segment. We also find, when plotted individually, that the points for all 29 of the individual items follow a similar left-to-right downward trend in this axis scheme (Appendix Figure A1). The layout of the axes in Figure 1, meanwhile, assists with results interpretation: a hypothetical point toward the bottom left of the graph indicates both a small outgoing self-efficacy rating and a small amount of growth (which, together, also imply a low incoming rating); comparatively, a point toward the upper right would indicate a relatively large amount of self-efficacy growth and a comparatively high outgoing rating. Thus, the upper right of the graph suggests comparatively strong outcomes. The upper left of the graph, however, is also an area indicative of promising outcomes: here, a relatively large amount of self-efficacy growth is observed, yet the incoming ratings were likely quite low to begin with, suggesting both substantive development and room for further growth.

Per Figure 1, we observe a positive change in mean self-efficacy scores over the course of GEL Year-1 for all factors, with these increases ranging from five to 11 points (y-axis). We also observe that the final mean self-efficacy scores for the eight factors range from 67 to 80 (x-axis). Additional descriptive statistics characterizing the distributions for all points plotted in Figure 1 are provided in Appendix Table A1. While we find that most mean self-efficacy ratings are within 3 points of each other on both the x- and y-axes of Figure 1, we observe that the "Implementing and Delivering" factor has a final mean rating that is more than six points lower than any other factor. This spotlights a factor characterized by a low mean incoming self-efficacy score, and, though there was substantial growth in the rating over the course of GEL Year-1 (i.e., > 10 points), the post-segment mean rating for this factor remained substantively below the other categories. "Initiative and Accountability," on the other hand, is a factor exhibiting the opposite progression: its post-segment mean is the highest self-efficacy score of any of the factors (79.48), while it simultaneously exhibits the smallest pre-/post- change in self-efficacy score of any of the factors (5.09). We observe "Sensemaking" to be the factor that improved the most with respect to its final self-efficacy score, placing it highest above the linear trend line. Though we do not aim to draw conclusions about self-efficacy changes based on the single program segment used for illustrative purposes in this Work in Progress paper, we provide statistical tests of the significance of the changes for reference in Appendix Table A1. As shown, all of the pre-/post- segment mean changes in the eight factors are found to be statistically significant ($p < 0.001$).

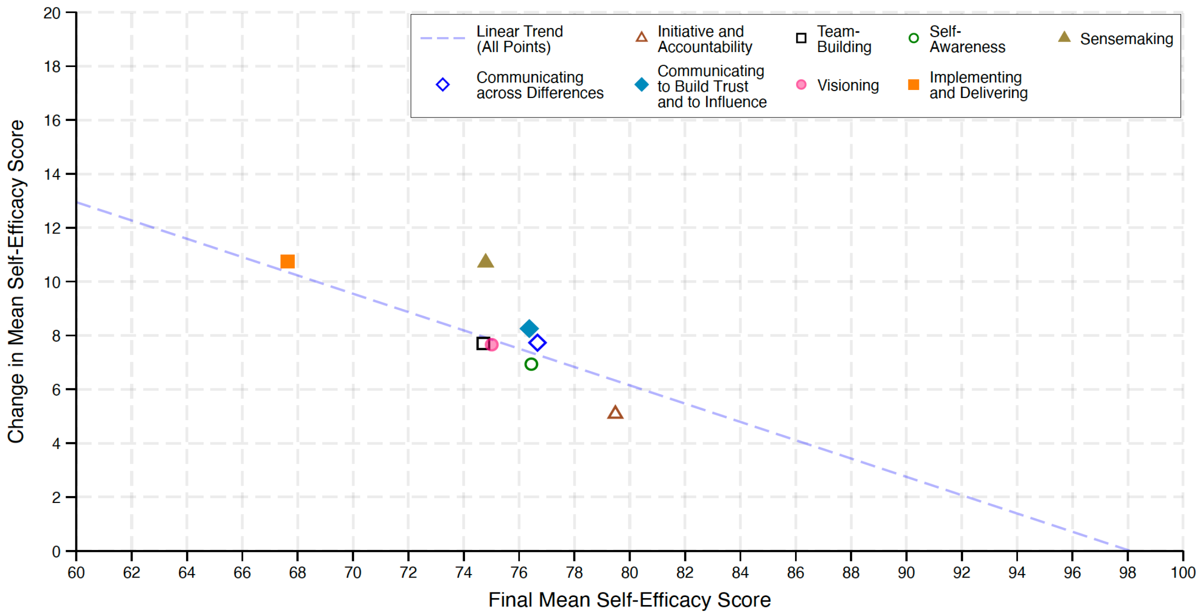


**Figure 1.** Mean self-efficacy outcomes for all factors

Without the availability of an external comparison group at this stage of the study, we must infer meaning from these findings through internal relative comparison – for instance, identifying comparative strengths and areas for improvement in learning across the curriculum at present. To enable this comparison, we establish a thresholding convention shown on the graphs in Figure 2. While Figure 1 presents the factor (category) means, Figure 2's graphs instead plot the means for all 29 individual items. The data in each of Figure 2's eight graphs are identical, yet each separately highlights the items composing a different one from among the factors. All of Figure 2's graphs also contain an identical triangular shaded region. With its hypotenuse matching the slope of the linear best-fit line for the set of 29 mean self-efficacy scores, this shaded region is set to encompass the bottom 20% of items in terms of growth (relative to outgoing self-efficacy). This "bottom 20%" threshold therefore reveals weaker areas of learning that may be improvement opportunities for the program. The 20% threshold value is a strategic choice and is not statistically derived; however, we intend to use it consistently over many years to allow us to detect whether the same weak areas underperform consistently. Beyond examining the composition of the threshold region, we also plan to examine the position of the diagonal threshold line over time; if the line rises along the vertical axis in successive cohorts, a year-over-year net growth in self-efficacy ratings is detected.

As shown in Figure 2, clusters of points composing each factor can be discerned relative to the trendline and threshold, illustrating that growth in some capability areas appears to be stronger than other areas. For instance, all four items in the "Sensemaking" group are above the trendline, reflecting a category that exhibits above-average growth over GEL Year-1 (relative to other factors). The items composing "Initiative and Accountability," meanwhile, are all below the trendline, with two in the threshold region, suggesting possibilities for program improvement in this area. In total, six items are in the shaded 20% threshold area: items a, c, k, v, x, and z. These graphs in Figure 2 offer more granularity than the factor averages for identifying improvement opportunities in the program. Meanwhile, Appendix Table A1 provides more detailed statistics for the self-efficacy ratings for all 29 items, including tests of statistical significance of the pre-/post- program changes. At the $p < 0.05$ significance level (or better), we find that 28 of the 29 items exhibit a statistically significant improvement in self-efficacy between the beginning and end of GEL Year-1, with Item c from "Initiative and Accountability" being the one item exhibiting non-significant change.

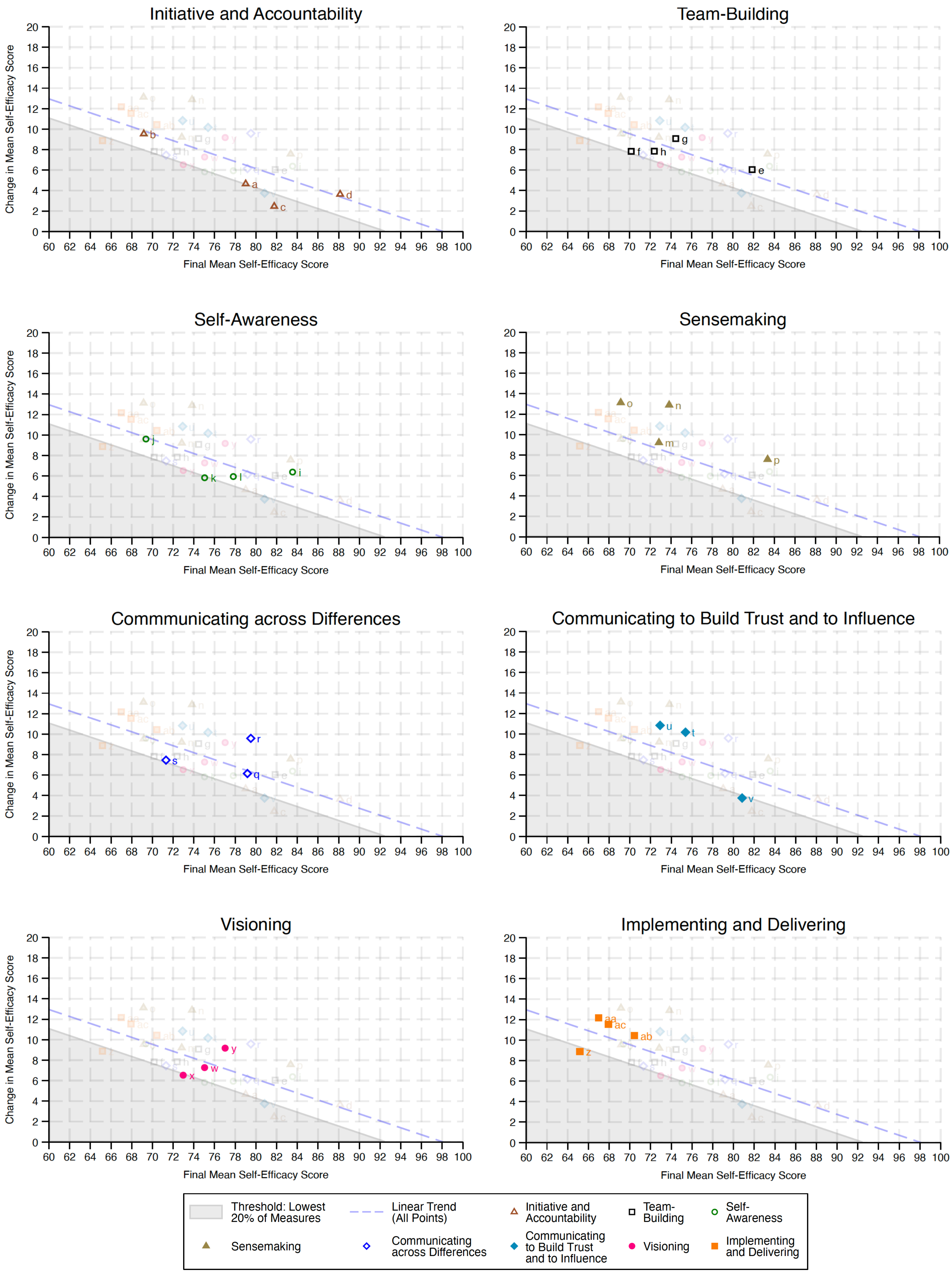


**Figure 2.** Mean self-efficacy scores for all items (grouped by factor)

*Limitations of findings*

Limitations should be considered when interpreting findings from this Work in Progress paper. First, changes in self-assessed leadership skills are known to be sensitive to measurement timing and to variation in incoming students' capabilities; therefore, a series of multiple measurements taken over time may be necessary to accurately identify growth in these capabilities among an EL program cohort [4, 26, 31]. The example self-efficacy outcomes presented here are inferred from only two measurement points (i.e., before and after one segment of GEL). Formal characterization of engineering leadership capability growth trends associated with GEL is best accomplished with additional longitudinal data, such as from follow-on self-efficacy measurement and career outcomes measurement (both of which are presently in process, as discussed in *Future Work*). This paper's self-efficacy outcomes data are therefore intended for methodological illustration purposes. Further, though our findings indicate a successful evaluation of GEL's assessment survey instrument in terms of model fit (factor structure) and internal consistency, some of the evaluation indices were at or near acceptability thresholds. The instrument will therefore benefit from further checks upon model fit and internal consistency from additional acquired data as the instrument is used in the future, or if additional measurement items are proposed for inclusion into the instrument. For instance, while the present 8-factor model demonstrated significantly better model fit compared to an ungrouped model (i.e., single-factor model) and exhibited strong RMSEA and SRMR error and residual metrics, respectively, its TLI and CFI model fit indices were near the low-end threshold of 0.9. Meanwhile, the "Communicating across Differences" factor was found to have a Cronbach's alpha near the low-end threshold of 0.7, suggesting internal consistency should be monitored for this factor as additional data are acquired.

**Discussion and conclusions**

*Longitudinal self-efficacy measurement as a tool for EL program evaluation*

Similar to prior works, we find self-efficacy measurement to be a promising means for assessment in an EL program based on conventional measures of model fit and internal consistency. Yet, since prior research highlights that this type of measurement can be sensitive to differences in incoming student capability levels (across measures) and to the extent of students' prior practice experiences, we find it appropriate to develop mitigations for such sensitivities as components of an overall assessment plan. Our mitigation approach is threefold, one facet of which is a focus of this Work in Progress study, and two of which are planned for future work.

In the present study, to accommodate variation in incoming self-efficacy levels across measures, we piloted a novel way of visualizing and interpreting self-efficacy change. The approach simultaneously presents both magnitudes of self-efficacy ratings and their pre-/post- program segment changes, while incorporating a reference trendline (Figure 2). This method takes known measurement sensitivity into account by encouraging a focus on a given measure's change across a program segment relative to its level at a time point of interest, rather than on an absolute measurement of self-efficacy. Since the empirically-derived trendline suggests that, on average, self-efficacy ratings that are already high among incoming students will be less likely to further grow to the same extent as those measures typified by low incoming ratings, a strategic goal-setting approach for continuous program review and improvement is implied: those capabilities that are core to a program's intended learning, yet that are shown to typically be low among incoming students and are exhibiting low growth (relatively speaking) may be the best targets for educational improvements in the program. How many capabilities to target for improvement in a given assessment cycle is a program's choice; however, a percentile thresholding approach, similar to that described here (Figure 2), offers a means for consistent evaluation over time.

*Future work*

Future work is planned to continue integrating this assessment system into the ongoing operation of GEL, as well as to expand it for use in peer programs at MIT that also focus on professional and leadership skills development. Both such activities are in-process as of this writing.

In GEL, this assessment system's full roll-out will further strengthen our understanding of participant outcomes through two additional attributes that help mitigate measurement sensitivities: an additional longitudinal time point for self-efficacy measurement (i.e., such that a three-point incoming, midpoint, and outgoing sequence is achieved), and a coupling of same-participant in-program measurements with alumni career-related outcomes measurements. To achieve the three-point self-efficacy sequence, the assessment described here will be recurrently deployed at the start of junior year, end of junior year, and end of senior year. It will therefore encompass a baseline measurement, a measurement at the end of GEL Year-1, and a measurement at the end of GEL Year-2. Participants who do not opt to participate in Year-2 will be invited to take the assessment at the end of their senior year for comparative purposes. Meanwhile, all student participants in GEL will later be invited, with incentives, to take a recurring alumni outcomes survey during their career years, as described in a prior publication from this research project [17]. This alumni survey assesses career readiness, occupational outcome types, career advancement, leadership roles undertaken, and perceived program support toward such roles. A prior report on this GEL alumni assessment [17] presents findings from a sample of program alumni; yet, given the state of the assessment system roll-out at that time, could not connect those alumni outcomes to in-program student assessments. We plan to examine and report on same-student data spanning the full intended sequence of longitudinal in-program and alumni surveys soon.

Lastly, efforts are ongoing to expand the use of this assessment system into a sophomore year career skills development program, the Undergraduate Practice Opportunity Program (UPOP), and a graduate-level EL program, the Riccio-MIT Graduate Engineering Leadership Program (GradEL), both of which are peer programs to GEL at MIT. GEL and GradEL share the same curricular foundation [10], while UPOP utilizes its own similar capabilities-based approach. Nonetheless, in the cases of both peer programs, some customization of the self-efficacy instrument has been necessary to align the assessment's focus with core intended learning, resulting in some unique items and factors in both cases; as such, both customized versions of the self-efficacy instrument will undergo similar factor analyses and consistency evaluations, as described in this paper, in the near future. The same alumni component of this assessment system [17], however, will be shared across all three peer programs. Further, all three programs' assessments share the same SGIC-based participant log-in, allowing analysis of same-student participation across multiple programs, as well as the development of comparison groups for downstream analyses. As an example of the latter case, sophomore year participants who do not proceed into other program components will be invited and incentivized to voluntarily participate in junior and senior year surveys for comparison purposes. Future reports from this assessment initiative therefore aim to include outcomes analyses from the three programs, including comparisons across subsamples representing differing extents of program participation, where possible.

*Conclusions*

In this Work in Progress paper, we have presented intermediate results from the ongoing roll-out of a multi-year longitudinal assessment initiative for an Engineering Leadership program. Findings suggest that a 29-item, eight-factor self-efficacy survey instrument for EL capabilities assessment meets conventional standards for model fit (factor structure) and internal consistency. Stemming from these findings, this instrument will be used as part of a survey sequence for program participants that also couples, via an anonymous Self-Generated ID Code (SGIC), with an alumni survey component. This full system will enable same-participant longitudinal assessment extending from undergraduate years into the career years. In presenting this approach and its inherent tradeoffs and limitations, as well as its relationships to existing approaches, we hope to assist the EL programs community by expanding the body of knowledge on programmatic assessment tailored for such programs' unique needs related to student outcomes evaluation.

## Acknowledgements

Development of the assessment approach described in this paper was funded in substantial part by the Bernard M. Gordon Charitable Trust through its funding of the Gordon-MIT Engineering Leadership Program. We are grateful for the funding that enabled this work. The authors are thankful for feedback on the draft manuscript from colleagues Leo McGonagle, Eileen Milligan, and Reza Rahaman at GEL. We are also grateful for guidance on data visualization from Jacqueline Goldstein at MIT's School of Engineering Communication Lab. The authors declare no conflicts of interest.

## References

[1] M. Klassen, D. Reeve, C. Rottmann, R. Sacks, A. Simpson, and A. Huynh, "Charting the landscape of engineering leadership education in North American universities," in *2016 ASEE Annual Conference and Exposition*, June 26 - 29, 2016. Available: https://peer.asee.org/26486

[2] M. Handley, D. Lang, P. Mittan, and A. Ragonese, "The history of engineering leadership development in academia: Influences, influencers, and a general roadmap," *New Directions for Student Leadership*, vol. 173, pp. 23-31, Spring 2022. https://doi.org/10.1002/yd.20476

[3] J. D. Stevens, D. Lang, M. Handley, J. J. Park, and P. Mittan, "Evaluating the effectiveness of an undergraduate engineering leadership development minor on graduates," in *2021 ASEE Virtual Annual Conference,* July 26 - 29, 2021. Available: https://peer.asee.org/37107

[4] R. Komarek, A. Bielefeldt, and D. Knight, "Self-assessment of leadership behaviors over time among students in a mechanical engineering capstone design course," in *2022 ASEE Annual Conference and Exposition,* June 26 - 29, 2022. Available: https://peer.asee.org/40937

[5] B. Ahn, M. F. Cox, J. London, O. Cekic, and J. Zhu, "Creating an instrument to measure leadership, change, and synthesis in engineering undergraduates," *Journal of Engineering Education*, vol. 103, no. 1, pp. 115-136, January 2014. https://doi.org/10.1002/jee.20036

[6] D. Lang, T. Gehr, M. Handley, J. J. Park, and A. M. Erdman, "An evaluation of an engineering leadership development program on alumni job placement and career progression," in *2020 ASEE Virtual Annual Conference,* June 22-26, 2020. Available: https://peer.asee.org/34119

[7] R. Paul and L. C. Falls, "Alumni perspectives on their undergraduate engineering leadership experience and important career skills," In *Proceedings of the Canadian Engineering Education Association (CEEA) Conference,* June 3-6, 2018, Vancouver, BC, Canada. Available: https://doi.org/10.24908/pceea.v0i0.13045

[8] R. J. Bennett, E. J. Audette, E. R. Millam, A. K. Moravetz, and S. Niebuhr, "A study of alumni of the 'Leveraging Leadership for a Lifetime' leadership development course," in *2021 ASEE Virtual Annual Conference*, July 26 - 29, 2021. Available: https://peer.asee.org/36613

[9] O. de Weck, R. Rahaman, and J. Schindall, "Integrating technical leadership and communications programs at MIT: Challenges and opportunities," In *2022 ASEE Annual Conference & Exposition*, June 26-29, 2022. Available: https://peer.asee.org/41856

[10] Bernard M. Gordon-MIT Engineering Leadership Program, *Capabilities of Effective Engineering Leaders, Version 5.1*. [Online]. Available: https://gelp.mit.edu/sites/default/files/images/Capabilities_v5.1.pdf [Accessed: March 24, 2025]

[11] L. R. Lattuca, P. T. Terenzini, and J. F. Volkwein, "Engineering Change: A study of the impact of EC2000," Baltimore, MD: ABET, Inc., 2006. Available: https://www.abet.org/wp-content/uploads/2015/04/EngineeringChange-executive-summary.pdf

[12] R. Graham, E. Crawley, and B. R. Mendelsohn, "Engineering leadership education: A snapshot of international good practice," *White paper sponsored by the Gordon-MIT Engineering Leadership Program*, 2009. [Online]. Available: https://www.rhgraham.org/resources/ELE-White-Paper-102109.pdf [Accessed: March 24, 2025]

[13] W. A. Lucas, "Using the CDIO syllabus 2.0 to assess leadership self-efficacy," in *Proceedings of the 10th International CDIO Conference,* June 16 – 19, 2014. Available: https://cdio.org/files/document/cdio2014/85/85_Paper.pdf

[14] A. Bandura, "Guide for creating self-efficacy scales," in *Self-efficacy beliefs of adolescents,* F. Pajares and T. Urdan, Eds., Greenwich, CT: IAP, 2006, ch. 14., pp. 303 – 337.

[15] M. A. Robinson, "Using multi-item psychometric scales for research and practice in human resource management," *Human Resource Management*, vol. 57, no. 3, pp. 739 - 750, 2018. https://doi.org/10.1002/hrm.21852

[16] E. Knekta, C. Runyon, and S. Eddy, "One size doesn't fit all: Using factor analysis to gather validity evidence when using surveys in your research," *CBE-Life Sciences Education*, vol. 18, no. 1, Mar. 2019. https://doi.org/10.1187/cbe.18-04-0064

[17] J. N. Magarian and R. S. Rahaman, "What engineering leaders lead: The career outcomes of an engineering leadership program's alumni community," in *2023 ASEE Annual Conference & Exposition*, June 25-28, 2023. Available: https://peer.asee.org/44614

[18] L. M. Audette, M. S. Hammond, and N. K. Rochester, "Methodological issues with coding participants in anonymous psychological longitudinal studies," *Educational and Psychological Measurement*, vol. 80, no. 1, pp. 163-185, Feb. 2020. https://doi.org/10.1177/0013164419843576

[19] J. J. Park, M. Handley, D. Lang, M.A. Erdman, "Engineering leadership development: Contribution of professional skills to engineering undergraduate students' leadership self-efficacy," *International Journal of Educational Methodology*, vol. 8, no. 1, pp. 69-80, Feb. 2022. https://doi.org/10.12973/ijem.8.1.69

[20] S. Y. Yoon, P. K. Imbrie, and T. K. Reed, "Development of the leadership self-efficacy scale for engineering students," in *2016 ASEE Annual Conference and Exposition*, June 26 - 29, 2016. Available: https://peer.asee.org/26832

[21] K. Duliba and W. N. Felder, "Strengthening systems engineering leadership curricula using competency-based assessment," in *Disciplinary Convergence in Systems Engineering Research,* A. M. Madni, B. Boehm, R. G. Ghanem, D. Erwin, M. J. Wheaton, Eds., Cham, Switzerland: Springer, 2018, pp. 1121 – 1133.

[22] D. Bayless, "Assessing the effectiveness of leadership education for engineering students," in *2014 ASEE Annual Conference and Exposition*, June 15 - 18, 2014. Available: https://peer.asee.org/20098

[23] R. Paul and L. C. Falls, "Engineering student leaders' self-assessments of leadership skills," in *Proceedings of the Canadian Engineering Education Association (CEEA)*, May 31 - June 3, 2015. Available: https://doi.org/10.24908/pceea.v0i0.5798

[24] A. Artino, "Academic self-efficacy: From educational theory to instructional practice," *Perspectives on Medical Education*, Vol. 1, pp. 76 – 85, 2012. https://doi.org/10.1007/s40037-012-0012-5

[25] J. A. Ross, "The reliability, validity, and utility of self-assessment," *Practical Assessment, Research, and Evaluation*, Vol. 11, No. 11, pp. 1 – 13, 2006. https://www.proquest.com/docview/2366824030

[26] D. Dunning, C. Heath, and J. M. Suls, "Flawed self-assessment: Implications for health, education, and the workplace," *Psychological Science in the Public Interest*, vol. 5, no. 2, pp. 69-106, 2004. https://doi.org/10.1111/j.1529-1006.2004.00018.x

[27] M. Zhao, M. A. Rodriguez, B. Wang, E. J. Santa Ana, J. Friedberg, Y. Fang, ... and S. Natarajan, "Validity and reliability of a short self-efficacy instrument for hypertension treatment adherence among adults with uncontrolled hypertension," *Patient Education and Counseling,* vol. 104, no. 7, pp. 1781 – 1788, 2021. https://doi.org/10.1016/j.pec.2020.12.029

[28] C. M. Dougherty, S. K. Johnston and E. A. Thompson (2007), "Reliability and validity of the self-efficacy expectations and outcome expectations after implantable cardioverter defibrillator implantation scales," *Applied Nursing Research*, vol. 20, no. 3, pp. 116 – 124, 2007. https://doi.org/10.1016/j.apnr.2007.04.004

[29] J. A. Baldwin, D. Ebert-May, and D. J. Burns, "The development of a college biology self-efficacy instrument for nonmajors," *Science education*, vol. 83, no. 4, pp. 397 - 408, 1999.

[30] B. Farrell, L. Richardson, L. Raman-Wilms, D. de Launay, M. W. Alsabbagh, and J. Conklin, "Self-efficacy for deprescribing: a survey for health care professionals using evidence-based deprescribing guidelines" *Research in Social and Administrative Pharmacy*, vol. 14, no. 1, pp. 18-25, 2018. https://doi.org/10.1016/j.sapharm.2017.01.003

[31] D. M. Rosch and K. L. Priest, "Strategies in assessment of leadership competencies," *New Directions for Student Leadership*, vol. 17, no. 156, pp. 87-99, 2017. https://doi.org/10.1002/yd.20273

[32] R. Komarek, D. Knight, and A. R. Bielefeldt, "Evolution of leadership behaviors during two-semester capstone design course in mechanical engineering," in *2018 ASEE Annual Conference & Exposition*, June 24 - 27, 2018. Available: https://peer.asee.org/30460

[33] L. Hu and P. M. Bentler, "Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives," *Structural Equation Modeling,* vol. 6, no. 1, pp. 1-55, 1999. https://doi.org/10.1080/10705519909540118

[34] W. H. Finch, "Using fit statistic differences to determine the optimal number of factors to retain in an exploratory factor analysis," *Educational and Psychological Measurement*, vol. 80, no. 2, pp. 217-241, 2019. https://doi.org/10.1177/0013164419865769

[35] R. A. Peterson, "A Meta-analysis of variance accounted for and factor loadings in exploratory factor analysis," *Marketing Letters*, vol. 11, no. 3, pp. 261-275, 2000.

[36] M. Tavakol and A. Wetzel, "Factor analysis: A means for theory and instrument development in support of construct validity," *International Journal of Medical Education,* vol. 11, pp. 245-247, 2020. https://doi.org/10.5116/ijme.5f96.0f4a

[37] J. M. Cortina, "What is coefficient alpha? An examination of theory and applications," *Journal of Applied Psychology,* vol. 78, no. 1, pp. 98 – 104, 1993. https://doi.org/10.1037/0021-9010.78.1.98

[38] M. Tavakol and R. Dennick, "Making sense of Cronbach's alpha," *International Journal of Medical Education,* vol. 2, pp. 53-55, 2011. https://doi.org/10.5116/ijme.4dfb.8dfd

[39] R. C. MacCallum, K. F. Wildaman, S. Zhang and S. Hong, "Sample size in factor analysis," *Psychological Methods*, vol. 4, no. 1, 1999. https://doi.org/10.1037/1082-989X.4.1.84

[40] T. A. Kyriazos, "Applied psychometrics: Sample size and sample power considerations in factor analysis (EFA, CFA) and SEM in general," *Psychology*, vol. 9, no. 8, pp. 2207-2230, 2018. https://doi.org/10.4236/psych.2018.98126

[41] Massachusetts Institute of Technology. *MIT Facts – Enrollment Statistics.* [Online]. Available: https://facts.mit.edu/enrollment-statistics/ [Accessed: March 24, 2025]

[42] M. Alavi, D. Visentin, D. Thapa, G. Hunt, R. Watson, M. Cleary, "Chi-square for model fit in confirmatory factor analysis," *Journal of Advances Nursing*, vol. 76, no. 6, pp. 2209-2211, 2020. https://doi.org/10.1111/jan.14399

[43] D. Santor, J. Haggerty, J. Levesque, F. Burge, M. Beaulieu, D. Gass, R. Pineault, "An overview of confirmatory factor analysis and item response analysis applied to instruments to evaluate primary healthcare," *Healthcare Policy*, vol. 7, pp. 79-92, 2011. https://doi.org/10.12927/hcpol.2011.22694

[44] M. Wilson-Fetrow, V. Svihla, and A. Olewnik, "Confirmatory factor analysis of the framing agency survey," in *2023 ASEE Annual Conference & Exhibition,* June 25-28, 2023. Available at: https://peer.asee.org/42737

[45] H. W. Marsh, K. Hua, and D. Grayson, "Goodness of fit in structural equation modeling," in *Contemporary Psychometrics,* A. Maydeu-Olivares and J. J. McArdle, Eds., Routledge, NY: 2005.

# Appendix

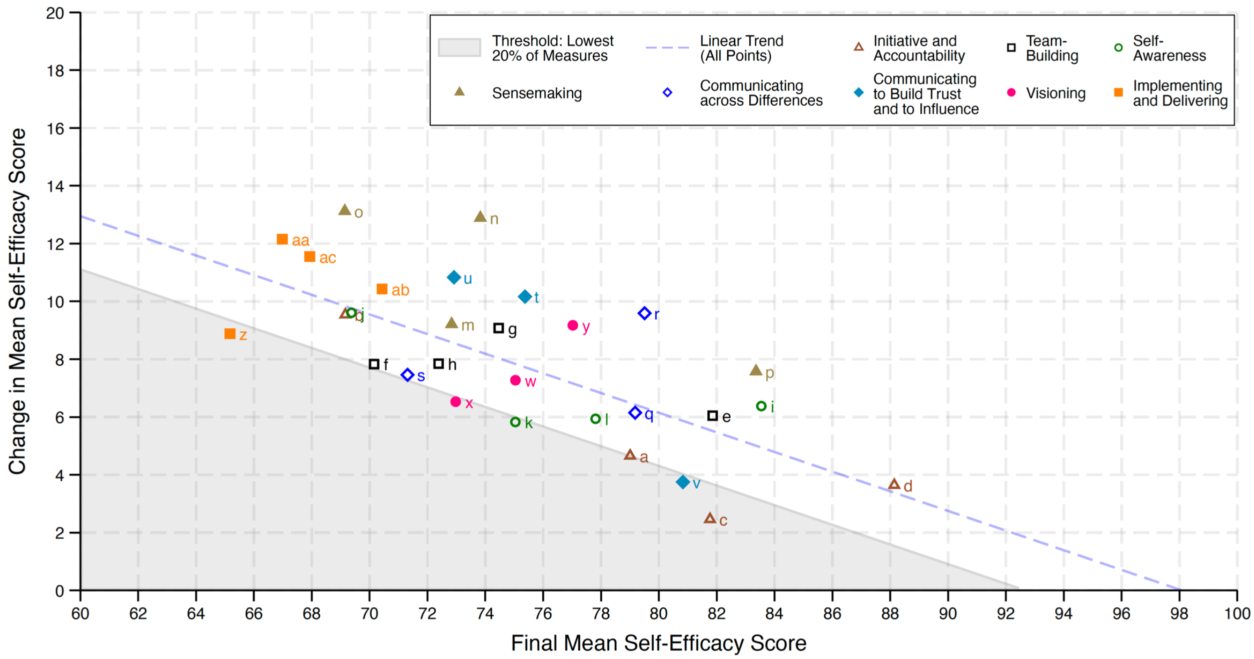


**Figure A1:** Mean self-efficacy outcomes for all items

**Table A1:** Pre- and post-program segment self-efficacy outcomes by item and factor (page 1 of 2)

| Item Label | Factors (in bold) and Items | Incoming Survey Statistics[2] n[1] | median | mean | sd | Outgoing Survey Statistics[2] n[1] | median | mean | sd | Diff. in mean[2] | Test Stat.[3,4] |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | **Initiative & Accountability (IA)** | **527** | **80** | **74.40** | **18.23** | **522** | **80** | **79.48** | **14.73** | **5.09** | **5.10** *** |
| a | To step forward and take responsibility for an activity that's critical to your project's mission when others fail. | 132 | 80 | 74.35 | 16.08 | 131 | 80 | 79.01 | 12.46 | 4.66 | 3.19 ** |
| b | To make the best possible decision in a situation where you have incomplete or uncertain information. | 132 | 60 | 59.62 | 17.03 | 131 | 70 | 69.16 | 14.57 | 9.54 | 5.50 *** |
| c | To uphold your commitments to your team, or to promptly coordinate mutually acceptable changes to commitments, when the conditions you face become more difficult than you expected. | 132 | 80 | 79.31 | 15.36 | 130 | 80 | 81.77 | 12.91 | 2.46 | 1.83 |
| d | To always contribute your best efforts as a team player, based on your personal strengths and skills in teamwork. | 131 | 90 | 84.50 | 14.30 | 130 | 90 | 88.14 | 12.23 | 3.64 | 3.03 ** |
| | **Team Building (TB)** | **526** | **70** | **67.01** | **20.73** | **520** | **80** | **74.71** | **17.47** | **7.70** | **5.44** *** |
| e | To contribute toward creating a team environment where all team members feel a sense of belonging, including among those who are different from you. | 131 | 80 | 75.81 | 19.59 | 130 | 90 | 81.86 | 16.05 | 6.05 | 3.61 *** |
| f | To recruit and select new team players who bring complementary skills that are needed by the team. | 131 | 70 | 62.33 | 21.52 | 130 | 70 | 70.16 | 18.07 | 7.83 | 4.13 *** |
| g | To lead the establishment of shared norms that enhance the performance of your team. | 132 | 70 | 65.38 | 20.24 | 130 | 80 | 74.46 | 16.33 | 9.08 | 4.99 *** |
| h | To lead post-project reviews so that your team can learn from experiences and improve performance on future projects. | 132 | 70 | 64.54 | 19.09 | 130 | 70 | 72.38 | 17.29 | 7.85 | 4.45 *** |
| | **Self-Awareness (SA)** | **513** | **70** | **69.51** | **21.63** | **524** | **80** | **76.44** | **16.53** | **6.94** | **4.93** *** |
| i | To identify and understand your own personal values. | 128 | 80 | 77.17 | 18.34 | 131 | 90 | 83.54 | 14.06 | 6.38 | 3.92 *** |
| j | To make effective decisions about your career and about changes in career direction. | 128 | 60 | 59.76 | 23.42 | 131 | 70 | 69.37 | 16.75 | 9.61 | 4.16 *** |
| k | To make life choices and changes that are aligned with a personal vision and meaningful sense of purpose. | 128 | 70 | 69.21 | 21.92 | 131 | 80 | 75.04 | 16.42 | 5.83 | 3.22 ** |
| l | To consistently act upon your personal values in situations where they are challenged. | 129 | 70 | 71.88 | 18.94 | 131 | 80 | 77.81 | 15.72 | 5.94 | 4.25 *** |
| | **Sensemaking (SM)** | **515** | **70** | **64.09** | **21.22** | **524** | **80** | **74.79** | **15.16** | **10.70** | **6.53** *** |
| m | To recognize when a situation's complexity warrants creating a system model to aid understanding. | 128 | 60 | 63.62 | 19.63 | 131 | 70 | 72.83 | 13.91 | 9.21 | 4.61 *** |
| n | To translate customer needs into technical requirements that engineers will interpret effectively. | 129 | 60 | 60.94 | 20.45 | 131 | 80 | 73.83 | 14.09 | 12.89 | 5.89 *** |
| o | To navigate the internal organizational politics of your organization to create a project plan that is acceptable to internal and external decision-makers. | 129 | 60 | 56.02 | 22.11 | 131 | 70 | 69.14 | 16.60 | 13.13 | 5.55 *** |
| p | To seek out multiple perspectives when defining problems. | 129 | 80 | 75.78 | 17.51 | 131 | 80 | 83.36 | 12.05 | 7.58 | 4.74 *** |
| | **Communicating Across Differences (CD)** | **369** | **70** | **68.93** | **18.18** | **393** | **80** | **76.67** | **14.57** | **7.73** | **5.97** *** |
| q | To listen very carefully to a team member with whom you disagree, rather than solely thinking about what you are going to say next. | 123 | 80 | 73.03 | 18.76 | 131 | 80 | 79.18 | 14.47 | 6.15 | 3.77 *** |
| r | To communicate a technical concept to senior leaders in your organization who do not share your technical background. | 123 | 70 | 69.92 | 16.54 | 131 | 80 | 79.51 | 13.10 | 9.59 | 5.91 *** |
| s | To help two of your fellow team members with strong differences of opinion reach an agreement. | 123 | 70 | 63.85 | 18.11 | 131 | 70 | 71.31 | 14.71 | 7.46 | 3.71 *** |
| | **Communicating to Build Trust and to Influence (CT)** | **367** | **70** | **68.12** | **18.97** | **390** | **80** | **76.37** | **15.29** | **8.25** | **5.65** *** |
| t | To provide a team member with constructive criticism that improves their performance. | 123 | 70 | 65.21 | 17.75 | 130 | 80 | 75.37 | 14.83 | 10.17 | 5.31 *** |
| u | To advocate for and receive additional resources for your team when only limited resources are available. | 122 | 60 | 62.08 | 18.78 | 130 | 70 | 72.92 | 14.63 | 10.83 | 5.43 *** |
| v | To build positive working relationships with your supervisor and fellow team members. | 122 | 80 | 77.08 | 17.07 | 130 | 80 | 80.83 | 15.43 | 3.75 | 2.23 * |

**Table A1:** Pre- and post-program segment self-efficacy outcomes by item and factor (page 2 of 2)

| Item Label | Factors (in bold) and Items | Incoming Survey Statistics[2] n[1] | median | mean | sd | Outgoing Survey Statistics[2] n[1] | median | mean | sd | Diff. in mean[2] | Test Stat.[3,4] |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | **Visioning (VN)** | **369** | **70** | **67.36** | **18.39** | **390** | **80** | **75.01** | **15.31** | **7.66** | **4.67** *** |
| w | To lead your team's formulation of a shared mission and vision for a project. | 123 | 70 | 67.77 | 18.23 | 130 | 80 | 75.04 | 15.34 | 7.27 | 4.30 *** |
| x | To launch your team in ways that motivate members to achieve the goals, scope, and deadlines of your project. | 123 | 70 | 66.45 | 18.84 | 130 | 80 | 72.98 | 15.63 | 6.53 | 3.47 *** |
| y | To identify the critical questions that need to be answered before the best solutions can be selected on a project. | 123 | 70 | 67.85 | 18.22 | 130 | 80 | 77.02 | 14.81 | 9.17 | 4.88 *** |
| | **Implementing & Delivering (ID)** | **472** | **60** | **56.88** | **19.70** | **512** | **70** | **67.63** | **17.81** | **10.75** | **6.50** *** |
| z | To determine in the first week of a project if your team is lacking in competencies or resources that will be needed. | 118 | 60 | 56.29 | 19.18 | 128 | 70 | 65.17 | 18.49 | 8.88 | 4.91 *** |
| aa | To convince your team to discard a really novel idea and instead use a satisfactory idea that can more assuredly be delivered on time. | 118 | 60 | 54.83 | 20.11 | 128 | 70 | 66.98 | 18.19 | 12.16 | 5.69 *** |
| ab | To detect and resolve gaps in your team's preparedness for a major project milestone, such as a product demonstration or test. | 118 | 60 | 60.00 | 18.97 | 128 | 70 | 70.43 | 16.23 | 10.43 | 5.59 *** |
| ac | To convince your team to agree that a major change is needed part-way through a project without negatively impacting team motivation. | 118 | 60 | 56.38 | 20.40 | 128 | 70 | 67.93 | 18.06 | 11.55 | 5.49 *** |

Notes:
1. Observation count varies among factors and among items due to missing voluntary responses on some surveys
2. Data reflected in summary statistics is based on all observations for which a matched pair of incoming and outgoing surveys exists for a given participant
3. Test statistics are calculated from the Wilcoxon Signed-Rank Test
4. Test statisitcs for factors are calculated by averaging each respondent's incoming and outgoing scores across the factor and using a Wilcoxon Signed-Rank Test

*p< 0.05; **p< 0.01; ***p< 0.001