Leadership Growth over Multiple Semesters in Project-Based Student Teams Embedded in Faculty Research (Vertically Integrated Projects)

Julia Sonnenberg-Klein, and Edward J. Coyle, *Fellow, IEEE*

Abstract—

Contribution: **This longitudinal study modeled student leadership growth in a course sequence supporting long-term, large-scale, multidisciplinary projects embedded in faculty research. Students (half from computer science, computational media, electrical engineering, and computer engineering) participated for 1-4 semesters.**
Background: **Project-based learning is used widely in higher education. It is used in industry for leadership development, but leadership development in PBL has not been explored in higher education. A preliminary analysis implied leadership growth through the third semester of participation, but the design did not control for attrition.**
Research Questions: **At the student level, how do leadership role ratings change over multiple semesters of participation? Do first (and second) semester ratings differ by number of semesters students eventually participate?**
Methodology: **The study involved two peer evaluation questions on 1) the degree to which students coordinated the team's work and 2) served as technical/content area leaders. Analysis employed analysis of variance to examine attrition by initial ratings (N = 1,045) and multilevel growth modeling to study change over time (N = 585). A strength of using peer evaluations is the large sample size, but a weakness is that the tool was developed for student assessment and not educational research. The study did not control for participation in leadership programs outside the course.**
Findings: **On average, individual leadership role ratings increased each semester through the third semester of participation. Ratings of students who left the program after 1 or 2 semesters did not differ from ratings for those who participated longer.**

Index Terms—**leadership, multidisciplinary, interdisciplinary, project-based learning, team-based learning, undergraduate, higher education, Vertically Integrated Projects, VIP.**

I. INTRODUCTION

THERE is worldwide interest in students' development of professional skills to address workforce needs, particularly leadership skills [1], and in the United States, colleges and universities have long been expected to develop future leaders [2]. Employers place greater value on leadership than high grades, ranking only

internships and being from the desired major above leadership in attributes they seek in job applicants [3]. Leadership is critical to problem solving, community engagement, and career success, and leadership capacity affects outcomes across higher education [4]. However, institutions can pay lip-service to student leadership, claiming to create "global citizen leaders" without measuring outcomes, and often only offer leadership programming as an extracurricular activity [2, p. 13].

For decades, PBL has been employed in engineering and computer science education, from small projects embedded in traditional courses to project-based courses that span a semester, multiple semesters, or years [5], [6]. The utility of using PBL in leadership development has been explored in industry [7], [8], but the connections are not assessed in higher education [9]. A review of assessments of PBL in higher education found that teamwork and collaboration skills were assessed in 3 of 76 studies [10]. In the first, aerospace engineering undergraduates participated in a 10-week PBL experience. The study examined a variety of specific and general skills, with teamwork addressed in a single survey item that did not address leadership [11]. The second study involved a project-based learning experience in which operations management master's students worked in groups of four. The study assessed teamwork through four survey questions that did not involve leadership [12]. In the third study, students from three majors (computer science, graphic design, and hotel/restaurant management) collaborated on interdisciplinary PBL projects. Researchers studied the frequency with which students mentioned soft skills in student journals and focus groups, finding that leadership was mentioned in 8% of comments [13].

This study seeks to fill a gap in research on student leadership development and PBL. The study focuses on leadership activity over multiple semesters in large project-based teams embedded in faculty research. The study builds on a cross-sectional study in which student peer evaluations were used to examine student leadership activity by academic rank (year in school) and number of semesters on the team [14]. Results of the previous study showed no

^Corresponding author: Julia Sonnenberg-Klein.

Julia Sonnenberg Klein is with the VIP Program at the Georgia Institute of Technology, Atlanta, GA 30332 USA, and is a doctoral candidate in Education Policy Studies at Georgia State University, Atlanta, GA 30302 USA (email: jsk@gatech.edu).

Edward J. Coyle is with the School of Electrical and Computer Engineering and the VIP Program at the Georgia Institute of Technology, Atlanta, GA 30332 USA (email: ejc@gatech.edu).

difference by academic rank, and significant differences between students in their 1st, 2nd and 3rd or later semesters with medium to large effect sizes. On average, students of all academic ranks in their 1st semester on their teams provided similar levels of leadership; as did students in their 2nd semester with their teams; and in their 3rd and later semesters.

The primary limitation of the prior study was that it used data from a single semester. If students who received low ratings in their first semester did not return for a second semester, and if ratings did not change across semesters, the mean for second-semester students would be higher than for first-semester students even though their leadership roles had not changed. Attrition could similarly lead to higher means for third-semester students.

The purpose of this study is to address the limitations of the initial study, and to determine whether undergraduate leadership measured at the student level increases over multiple semesters of participation. The study seeks to answer the following research questions:

- Do leadership role ratings for undergraduates in their 1st and 2nd semesters of participation differ by the number of semesters they eventually participate in the program?
- At the student level, to what degree do leadership roles change over multiple semesters of participation?

A. Leadership Education and Development

Academic programs incorporate a variety of leadership skills into learning outcomes. Across 522 types of academic programs in the United States, Seemiller identified 60 competencies related to leadership development [15]. He grouped the them into eight clusters: learning and reasoning; self-awareness and development; interpersonal interaction; group dynamics; civic responsibility; communication; strategic planning; and personal behavior [15].

Komives and Sowcik differentiate between leadership education and leadership development. Leadership education involves structured instruction and is typically offered through campus offices of student affairs [2], [16]. In contrast, leadership development increases skills and leadership capacity in applied settings such as clubs, teams, student government, etc. [17]. A meta-analysis of studies on student leadership education found that knowledge acquisition outpaced transfer of skills, indicating that students learned about leadership through instruciton, but had fewer opportunities to use the knowledge and develop the skills in applied contexts [18].

At some institutions, leadership is incorporated into the campus and/or engineering curriculum as degree supplements, certificates, minors, or as full degree programs [19]. In a survey of programs in engineering, Paul and Gradon [19] identified 40 programs that involved leadership, of which 11 focus primarily on leadership. Seven of the 11 programs were open to undergraduates, 2 were for high achieving undergraduates, and 2 were graduate certificates. Across the programs, Paul and Gradon identified five

themes: effective leadership, independent learning, experiential learning, innovation and technology, and systems thinking. Experiential learning, which included PBL, was present in over half of the 11 programs.

B. Pedagogical Framework

The study involved student teams in Georgia Tech's Vertically Integrated Projects (VIP) Program. VIP is a model for undergraduate research in which large student teams are embedded in faculty research, scholarship, and creative endeavors. The model is used at 46 colleges and universities around the world [20]. Student leadership is a key aspect of the model because students who return for 2nd, 3rd and subsequent semesters are expected to take on additional leadership and technical responsibilities. This enables faculty to manage large teams, allowing faculty to serve more students than apprentice-style undergraduate research [21], [22].

VIP is a specific case of PBL. Krajcik & Blumenfeld identified key aspects of PBL [23]. In PBL, learning focuses on a problem that is meaningful and important to the students [23]. In VIP, the problem is based in a faculty member's research, design, or exploration efforts, and students join teams they find interesting.

The second key feature of PBL is that "students explore the driving question by participating in authentic, situated inquiry... As students explore the driving question, they learn and apply important ideas in the discipline" [23, p. 318]. In VIP, faculty establish teams because they want/need the students' expertise, which may be in the mentor's own field or another. Then within VIP, students apply knowledge and skills from their disciplines, and they seek out/learn new knowledge and skills as needed.

Krajcik & Blumfeld's third key aspect of PBL is that "students, teachers, and community members engage in collaborative activities to find solutions to the driving question. This mirrors the complex social situation of expert problem solving" [23, p. 318]. In VIP, students work alongside and in community with their instructors; students coordinate within/between subteams; and teams engage stakeholders, sponsors, and experts. As a former student explained, "These interactions have a different dynamic than the typical student-teacher relationship, as students are more like collaborators than pupils" [24].

The final element of PBL is that "students create a set of tangible products that address the driving question. These are shared artifacts, publicly accessible external representations of the class's learning" [23, p. 318]. VIP artifacts vary by team and project, but deliverables include prototypes, deployments, presentations, wikis of ongoing documentation and design work, research posters, and publications. Not all products are accessible to the public, but they are used by continuing students, faculty mentors, and stakeholders.

C. Vertically Integrated Projects

VIP is a special case of PBL because teams are large and projects are long-term. While PBL is defined as involving

teams, PBL teams can be as small as 2 students [23]. Georgia Tech's VIP teams are expected to have at least 8 students. In Spring 2023, the average team size was 23, with a median of 20. In PBL, projects do not typically outlast students who work on them. VIP is a special case because projects last at least 3-5 years, much longer than any student's participation.

To enable students to participate multiple semesters, VIP is offered for 1-2 credits/semester, with 1 credit for sophomores (2nd year students) and 1-2 for juniors and seniors (3rd and 4th-5th year students). Two semesters of participation are equivalent to a standard 3-4 credit class.

An example of a VIP team is Lighting from the Edge of Space. Led by an Electrical Engineering professor who studies lightning, the team has been running for 8 years with approximately 17 students each semester. The team designs/builds high-altitude data-collection systems, launches the systems with weather balloons, analyzes collected data, and works to expand and optimize the systems. In Spring 2022 the team enrolled students from Aerospace Engineering, CMPE, CS, Earth & Atmospheric Science, EE, Mechanical Engineering, and Physics.

A variety of research has been done on the impact of and dynamics within VIP Programs and Engineering Projects in Community Service (EPICS), a service-based learning version of and predecessor to VIP. A mixed methods study involving EPICS alumni found EPICS experiences supported development of professional skills, bridged education and practice, and provided work-relevant experience [25]. An analysis of institution exit surveys found that compared to a matched group of non-VIP participants, VIP participants more strongly agreed that their educations contributed to their ability to work in multidisciplinary teams, ability to work with people from diverse backgrounds, and their understanding of technologies related to their fields [26]. Social network analysis of peer evaluations showed that within VIP, students interacted more often with students of other races/ethnicities, and more often with students from other majors [27]. Social network analysis also showed correlation between number of semesters on the team and helping and advising teammates, with no correlation between academic rank and helping teammates [28].

Analysis of VIP enrollments across five institutions was done to assess equity in enrollments by student demographics. Results showed small effects sizes for status as historically underserved minorities, very small effect sizes for first-generation students and transfer students, and slightly higher participation among women than men [29]. An earlier analysis of VIP enrollments at a single institution found representative enrollment by race/ethnicity, and that students returned for second and subsequent semesters at the same rate, regardless of race/ethnicity [30]. Analysis of enrollments and policies show a close relationship between policies on how credits count toward degree requirements, participation rates, and enrollment in second and subsequent semesters [31]. Degree programs that incentivized multiple semesters of participation, such as Electrical Engineering (EE) and Computer Engineering (CMPE), and that allowed VIP to fulfill multi-semester requirements, such as Computer Science (CS) and Computational Media (CM), had higher participation and persistence rates than other majors at the time. Analysis also found that departments with more VIP instructors were more likely to have established policies on how credits count [32].

Commonalities Across the Program

While projects and team sizes vary, commonalities across the program are the grading framework, typical team structure, and scheduled weekly meetings. Students are graded in three equally weighted categories: teamwork, documentation, and contributions to the project. Expectations differ by major, academic rank, and number of semesters on the team. Formal feedback is given at the middle and end of each semester. Midterm feedback is meant to be advisory, to enable students to improve performance before the end of the semester.

Each team is scheduled for one 50-minute meeting each week. Regardless of team size, most teams operate with sub-teams working on related aspects of the project. At weekly meetings students/subteams report on their progress, stay abreast of others' work, and develop to-do lists for the coming week.

Differences Across the Program

Differences across teams in the program are team size, leadership styles, the nature of projects, and disciplinarity diversity. Team size is critical to ensuring continuity from semester to semester. The Georgia Tech program recommends maintaining teams of at least 8-10 students/semester to ensure enough students return the next semester to continue the work, but some teams are very large, with more than 70 students.

While all teams are based in faculty projects/interests, projects can be faculty-driven, embedded in their core research; faculty-student-stakeholder-driven, such as developing, evaluating, and deploying apps for healthcare partners; student-stakeholder-driven, such as partnering with marginalized communities to study and support equity in built spaces with guidance from instructors; and competition-driven, EcoCAR, Formula SAE, etc.

Teams vary in the diversity of majors involved. A disciplinarily narrow team is Automated Algorithm Design, which only enrolls CS, CMPE and EE students. In contrast, the Soccer, Community, Innovation, and Politics team's work involves sociology, politics, economics, and technology, and it attracts students from every college on campus.

Leadership systems and project management styles also vary by team. While VIP provides new-instructor workshops on effective practices, the program does not provide training or frameworks for leadership or project management. Some teams use project-management methods from industry (scrum, agile, etc.), while others are less formal. Some also establish multilevel hierarchies, with student managers coordinating work between multiple subteams.

Manuscript ID TE-2023-000177

D. Theoretical Framework

In this study, student leadership growth is viewed through the context of Tuckman's model of group development [33], [34]. Published in 1965 and revised in 1977 with Jensen, Tuckman's model was the first to describe group development [35]. It is still the most often referenced model in organizational development and change [36], and it is used in higher education contexts. For example, in studying a semester-long PBL course, Cresswell-Yeager used Tuckman's model to frame communication within student groups. The model is also widely used in experiential learning program design and facilitator training [37].

Tuckman's model consists of five stages: forming, storming, norming, performing and adjourning [34]. In the forming stage, members get to know each other, and interactions are polite [38]. Power structures emerge in the storming stage. Group members may resist the formation of team structures or vie for power within the emerging structure [38], [39]. In the norming stage, members develop shared mental models for how the team will function, and the group becomes cohesive [38], [39]. In the performing stage, members work productively. In the adjourning stage, the group separates.

Because VIP teams include new students each semester, teams continually cycle through Tuckman's five stages. However, new and returning students experience the stages differently. In the forming stage, group members get to know one another. Whereas new students encounter an entirely new group of people, returning students already have working relationships with each other and instructors. In the storming stage, team power structures emerge. New students unfamiliar with the project are not well positioned to lead. In contrast, returning students are expected to help orient and mentor new students, positioning returning students as likely leaders.

II. METHODS

A. Data

The study involved four semesters of midterm peer evaluations administered in 2021 and 2022, along with enrollment information to determine students' 1st, 2nd, and subsequent semesters of participation. Semesters in which students withdrew were not counted as semesters of participation. Midterm evaluations were used to capture student dynamics in the midst of team establishment, because final evaluations would reflect performance after a full semester of growth.

Because prior analyses have shown no correlation between leadership [14] or help-giving [28] in VIP by academic rank, academic rank was not included in the analysis. Only results for undergraduates were analyzed, but reviews by graduates of undergraduates were included.

The peer evaluation is administered online. Before students are asked to evaluate classmates' ability or quality of work, they are first asked about the degree to which they interact and about roles classmates take on the team. In the first question, students are presented with a list of teammates and asked how often they interacted with each on a Likert scale of 1-5. A rating of 1 corresponds with, "NEVER: I do not know this person." A rating of 5 corresponds with "VERY FREQUENTLY: More than once a week." Teammates they report interacting with infrequently (ratings of 1-2) are excluded from subsequent questions, so students only provide ratings for classmates they interact with.

The analysis involved two peer evaluation questions. Response options were on a 5-point Likert scale with response anchors at the high and low ends of the scales. The two questions and scale anchors were:

- Think about how your team is organized. Please indicate whether each student below is (or is not) a technical/ content area leader on the team. (1 = Not a technical/content area leader; 5 = Technical/content area leader)
- Think about how your team is organized. Please indicate whether each student below coordinates (or does not coordinate) the team's work. (1 = Does not coordinate the team's work; 5 = Coordinates the team's work.)

The initial dataset included evaluations of 3,536 students with 49,165 responses for serving as a technical/content area leader and 48,705 for coordinating the team's work. In some evaluator-evaluatee pairings, the evaluator indicated interacting with a classmate somewhat often (3-5), answered one of the two leadership questions, but then backed-up and indicated they did not interact with the classmate as much (1 or 2), leaving one of the leadership questions answered. These cases were excluded.

Cases were excluded when reviewers gave all reviewees ratings of 5 on the same item, because the reviewers did not provide useful comparisons among classmates. For the same reason, cases were excluded when reviewers gave all 1s on the same item (and they had likely reversed the scale). These accounted for 13% reviews.

To enable comparisons across students who began within the two-year period being studied and who could have participated for at least three semesters during the semesters of analysis, the study was limited to students who first enrolled in VIP in Spring or Fall of 2021. This reduced the number of reviewed students from 3,494 to 1,118, with one or more reviews for each student each semester. The proportion was less than half of reviewees from the two-year period, because more students entered the program in 2022 than in 2021.

Finally, 73 cases were excluded because the number of midterm ratings did not match the number of semesters of participation. In some cases, students were evaluated but withdrew from the course, yielding too many midterm evaluations. In other cases, evaluations were missing. This may have been because their reviews were excluded in previous steps, students did not work closely enough with classmates to be evaluated (a sign to the instructor of a problem), the students worked with graduate student mentors instead of other teammates, or students did not participate in the evaluations because they assisted instructors with

TABLE I
Sample

College	Total Number of Semesters in VIP				
Major	1	2	3	4	Total
Engineering	**266**	**130**	**70**	**11**	**477**
Mechanical Eng.	75	32	23	1	131
Computer Eng.	38	16	9	2	65
Industrial Eng.	40	21	1	1	63
Aerospace Eng.	25	18	9	1	53
Biomedical Eng.	28	9	13	2	52
Electrical Eng.	22	14	8	1	45
Other	38	20	7	3	68
Computing	**75**	**40**	**263**	**15**	**393**
Computer Science	73	40	260	14	387
Other	2	0	3	1	6
Sciences	**68**	**16**	**9**	**2**	**95**
Biology	30	3	2	0	35
Neuroscience	16	1	1	1	19
Physics	7	2	2	0	11
Other	15	10	4	1	30
Liberal Arts	**18**	**7**	**8**	**0**	**33**
Computational Media	3	1	5	0	9
Public Policy	5	3	1	0	9
Other	10	3	2	0	15
Design	**22**	**4**	**6**	**1**	**33**
Industrial Design	16	3	2	0	21
Architecture	4	1	4	1	10
Other	2	0	0	0	2
Business	**11**	**3**	**0**	**0**	**14**
Total for Analysis 1	460	200	356	29	1045
Total for Analysis 2		200	356	29	585

TABLE II
Combined Leadership Role Rating

Semester of Rating	N	Mean	Median	SD
First	1045	3.48	3.50	0.83
Second	585	3.74	3.83	0.82
Third	385	4.00	4.07	0.73
Fourth	29	4.07	4.33	0.81
All Semesters	2044	3.66	3.75	0.83
Two and More Semesters	999	3.85	3.94	0.79

1-5 scale (Table II). Because the investigation showed normality, and because parametric tests are more sensitive than non-parametric tests, parametric tests were used in the analysis.

B. Analysis

To answer the first research question, analysis of variance (ANOVA) was used to compare ratings students received in their 1st semester by how many semesters they eventually completed. The same was done for 2nd semester ratings. ANOVA is appropriate when subjects' scores are independent of other subjects' scores. Although students reviewed teammates, their ratings were not influenced by scores received by reviewers, so their mean scores were treated as independent. ANOVA also assumes normality and homogeneity of variances, which were checked. Dunnett's test was selected for use in post-hoc analysis because it works well with unequal group sizes (Table II).

Multilevel modeling was used to answer the second research question, how leadership roles change over multiple semesters of participation. While repeated measures ANOVA can be used to model growth, the method cannot be used when cases have varying numbers of measurements. In this sample, students participated for varying numbers of semesters, which represents missing data for semesters in which students did not participate. Multilevel modeling can handle this type of missing data. Multilevel modeling is often used to account for groupings and the resulting lack of independence between group members (students within classrooms, classrooms within schools, etc.). In these types of models, individuals usually represent level 1 (student test score, student demographics, etc.), and groupings represent level 2 (students grouping by class, average scores for the class, etc.). In growth modeling, measurements taken at different times represent level 1, and groupings of measurements by student represent level 2.

SPSS was used for both analyses. While SPSS is not the best software for multilevel modeling [40], the most recent edition of [41] includes thorough explanations and screenshots. Multilevel models are built in stages. The first model is a null model, which includes groupings but no predictors. As predictors are added, fit statistics for the previous and new models are compared to determine if the addition improved model fit. Maximum Likelihood was used as the estimation method so fit statistics could be compared [41]. The log likelihood ratio chi-square test, Akaike's Information Criterion (AIC), and Schwarz's Bayesian Criterion (BIC) were used to compare model fit. Multilevel

grading. While multilevel modeling can handle missing data, matching each rating to the correct semester of participation was difficult when the number of ratings did not match the number of semesters, and excluding them simplified data preparation. This left 1,045 cases and 2,044 pairs of mean ratings (one pair of mean ratings per student per semester, with students participating 1-4 semesters). Because student level of experience in the program was the primary focus of the analysis, the data was restructured by each student's semester of participation (1st semester, 2nd semester, etc.) instead of by time (Fall 2021, Spring 2022, etc.).

The distributions by race/ethnicity and gender in the final sample were 52% Asian, 31% white, 6% black or African American, 6% Hispanic or Lantino, 4% two or more races, and 1% unknown; and 61% male and 39% female.

To determine if the two leadership questions could be combined in the analysis, correlations between the 2,044 mean rating pairs were examined. Regression showed high but non-perfect correlation between mean ratings on the two items, with a Pearson correlation of 0.82. Because the two questions were highly correlated, and because they measured different aspects of student leadership, they were averaged to yield a single leadership role rating (Table II).

Normality of the combined ratings were examined through Q-Q plots and histograms for the full sample (for research question 1) and for students who participated multiple semesters (research question 2). The Q-Q plots both showed normality. The histograms showed relatively normal distributions with higher frequencies at means of 3, 4 and 5, which may be the result of reviewer agreement. Both distributions were shifted to the right of the midpoint of the

linear modeling assumes a linear relationship between predictors and the dependent variable. The similarity in change between first and second semester group means (0.26) and second and third semester group means (0.26) implied linearity (Table II). The smaller increase between the third and fourth semester (0.07) were investigated and are discussed in the results section. Multilevel modeling also assumes that residuals are independent and normally distributed. A histogram and Q-Q plots were used to examine normality of residuals. Scatter plots and box plots were used to assess relationships between residuals and other variables.

B. Limitations

The scope of the study is limited to enactment of two aspects of leadership as reported by peers, coordination of the team's work, and serving as a technical or content area leader. Two items do not constitute a full construct. Cases were excluded when reviewers gave all of their reviewees the highest or lowest rating, when the number of midterm reviews received in the 2-year period did not match the number of semesters of participation, and when reviewees did not interact with classmates enough to be reviewed. If missingness was due to students' lack of interaction within their team, the results are less valid.

While peer evaluations may provide more objective assessments of leadership roles than self-reported measures, peer observations do not capture activities unobserved by peers such as planning, problem-solving, and decision-making between student and instructors, mentors, and stakeholders. Input from these other stakeholders would provide a more comprehensive view of student leadership.

Another limitation is that the study did not account for other leadership education or development activities within or outside of VIP. The institution offers a minor in leadership, and some students/teams have participated in leadership education workshops offered by other units on campus, but the VIP Program does not actively promote or track participation. Participation in these programs could explain differences between students in growth over time.

While multilevel modeling can handle missing data, three or more measurements per case are recommended for the method [42]. In the sample, less than half of the cases had three or more measurements. As a result, the modeled growth is more heavily influenced by changes between the first and second semesters. The similarity in group mean changes between the 1st and 2nd semester and the 2nd and 3rd semester made this less of a concern. The smaller group mean change between the 3rd and 4th semester were investigated.

III. RESULTS AND DISCUSSION

A. Results

1) Differences in early semesters by number of semesters eventually completed

ANOVA was used to compare mean ratings in students' 1st semester of participation by the number of semesters they eventually completed. Levene's test of homogeneity of
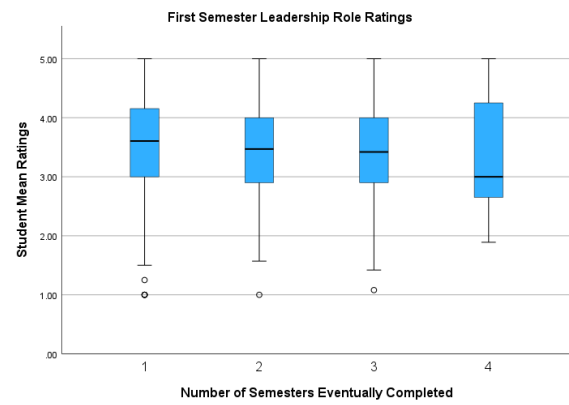

Fig. 1. Mean First Semester Leadership Role Ratings by Number of Semesters Eventually Completed

variance was not significant, indicating ANOVA would be appropriate. ANOVA showed differences between mean ratings students received in their first semester by the number of semesters eventually completed at the .05 level with a small effect size ($F(3, 1041) = 1.95$, $p = .04$, $\eta_P^2 = .01$). However, Dunnett's test showed no statistically significant differences between groups at the .05 level. The greatest observed (yet not significant) difference was between students who completed one semester ($N = 460$, $M = 3.56$, $SD = 0.80$) and four semesters ($N = 29$, $M = 3.25$, $SD = 0.83$), with a significance of $p = .08$ (Fig. 1).

Second semester ratings were also examined for differences by number of semesters eventually completed. Levene's test of homogeneity of variance was not significant, indicating ANOVA would be appropriate. ANOVA showed no difference in 2nd semester ratings by number of semesters eventually completed at the .05 level ($F(2, 582) = .093$, $p = .91$).

2) Leadership Growth

Data for students who participated for two or more semesters was used to model leadership growth over multiple semesters ($N = 585$). In the null model, mean ratings were grouped by student, and time was not included as a predictor. The null model converged when the covariance structure for repeated effects was set to diagonal. The ICC was .07, indicating that 7% of variance could be attributed to the clustering of measurements by student (AIC = 3866.29, BIC = 3898.49). The intercept was 3.72, representing the grand group mean.

In the second model, time was added as a fixed effect. The estimated change in mean ratings per semester was 0.28 (Fig. 2), with a significant t-test ($t(678) = 13.5$, $p < .001$). The ICC for the intercept increased to .09, and the log likelihood ratio test was statistically significant, confirming that addition of time as a predictor improved model fit ($\chi^2(1$, Deviance = 162.77$) < .001$, AIC = 3705.52, BIC = 3743.09). Notably, variances for the repeated measures were statistically significant for the first three time measurements, but not for the fourth. The lack of significance for the 4th semester after the addition of time as a predictor implied that the 4th semester ratings did not fit the growth curve.
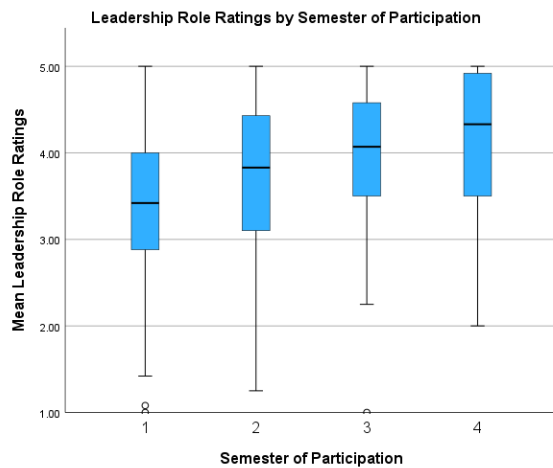
Manuscript ID TE-2023-000177

Fig. 2. Role Ratings by Semester of Participation

A new null model was run with the same number of cases by student, but with the 29 instances of 4th semester ratings excluded. This yielded an ICC of .09. The time predictor was added to the new null model. The estimated change in mean ratings for each semester was slightly higher at 0.29, with a statistically significant t-test ($t(685) = 13.22$, $p < .001$). The ICC increased to .11, and the log likelihood ratio test was statistically significant (χ^2 (1, Deviance = 156.59) $< .001$), indicating improved model fit. Allowing slopes to vary did not improve fit. Team and college were considered as possible grouping levels, and new null models were constructed. Grouping students within teams and/or by college did not yield different growth estimates, so the simple two-level model was retained, with measures grouped only by student with time as a fixed predictor.

The assumptions of normality of and independence of residuals were tested. A histogram and Q-Q plot showed normality. Scatter plots and boxplots showed no relationships between residuals and other variables.

ANOVA was done to confirm whether ratings differed for 4th and 3rd semester ratings. Results showed no difference between ratings for the two groups ($F(1, 412) = 0.247$, $p = .62$), confirming that exclusion of 4th semester ratings from the growth model was appropriate.

B. Discussion

The purpose of this study was to address the limitations of the previous study [14] and to determine whether undergraduate leadership increased over multiple semesters of participation. The previous study showed higher ratings for students by semester of participation through their third semester, but it was cross-sectional, and the seeming gains could have been the product of attrition. If students who received low ratings in their first semester did not return for a second semester, and if ratings did not change across semesters, the mean for second-semester students would be higher than for first-semester students even though their leadership roles had not changed. To address this shortcoming, the study involved two research questions. The first asked whether ratings in early semesters differed by the number of semesters students eventually completed, and the second asked how ratings changed over multiple semesters at the student level.

The first research question is important. If students with lower initial leadership role ratings leave the program at higher rates, they would have less opportunity for academic and professional growth. This was not the case. Analysis showed that 1st (and 2nd) semester leadership role ratings did not differ by the number of semesters students eventually completed. On average, students who continued in the program did not have higher initial ratings than students who left the program, implying no inequity by initial ratings.

Interestingly, mean 1st semester ratings were lowest among students who participated for four semesters, indicating that students who stayed the longest started out the weakest on average, with differences statistically significant in the ANOVA but not in the post-hoc analysis. While only 29 of the students in the sample participated for four semesters, the size of the subgroup was limited by the scope of the study. For students who began in the 2nd of the four semesters included in the study, data for their 4th semester would not have been included. A study that includes more semesters would increase the size of the 4th semester group and might show more conclusive results.

The second research question asked whether, at the student level, leadership roles changed over time. The prior cross-sectional study found that students in their 2nd semester received higher leadership role ratings than students in their 1st semester, and that students in their 3rd and later semesters received higher ratings than 2nd semester students [14]. In this study, multilevel modeling was used to model growth at the student level, eliminating the effect of attrition on the results. The results agreed with the prior study, showing gains in leadership role ratings from the 1st to 2nd semester, and from the 2nd to 3rd semester. On average, students' mean ratings increased by approximately 0.3 points per semester through the third semester.

The longitudinal analysis also agreed with the prior study, showing no difference between ratings for students in their 3rd and 4th semesters. This may indicate that students achieve their highest leadership levels in their 3rd semester, continuing with those levels into later semesters, or that the instrument does not detect leadership-related work that is not apparent to classmates, such as coordinating work with instructors, graduate mentors, or external stakeholders.

C. Implications for Research

The two questions used in the analysis provide a glimpse of team dynamics, but do not constitute thorough measurement of a construct, which usually involves at least 8-10 indicators. The high but non-perfect correlation between the items implies that the two roles (coordination of team's work, and serving as a technical/content area leader) are aspects of a leadership role construct, and more aspects could be explored. A challenge is balancing the original purpose of the peer evaluation (student assessment) with education research. Adding enough items to fully measure a

construct would nearly double the length of the evaluation, potentially decreasing response rates, which would be a disservice to instructors who rely on it.

Instructors have indicated that students in their 3rd and later semesters provide more critical/accurate evaluations of teammates, so another analysis could focus on responses from this subgroup, or a more extensive survey could be administered to returning students. The research could also be expanded to include ratings from instructors. However, if an instructor survey were administered every semester, the risk of survey fatigue would be high. Data would need to be collected over a finite period with incentives for participation.

A potential direction for program improvement would be a partnership with a campus leadership development program, to see if student participation in leadership education increases leadership growth. This may require slight modifications to offerings, because VIP differs in key ways from other contexts (i.e. teams are faculty-led), but VIP could provide valuable pre- and post-measures or treatment and control groups for the leadership education program.

Another important research question is whether similar patterns are seen in other VIP Programs. If similar peer evaluations are administered at multiple sites, leadership growth could be studied across different types of programs and institutions.

D. Implications for Practice

The findings of the study have implications for faculty, departments, and institutions. At the faculty level, the results of the study confirm that when faculty embed large student teams in their research, returning students help coordinate the teams' work and serve as technical/content area leaders. Students provide the greatest level of leadership in their third and subsequent semesters. For *faculty to maximally benefit from student leadership*, they need to *recruit students as sophomores and juniors*. This gives students enough time in their academic careers to participate for three or more semesters. Additionally, if faculty want to support their research with students who earn credit over multiple semesters, they need to *actively engage their department undergraduate curriculum committees*, to ensure the credits earned can fulfill degree requirements.

If departments value leadership development and/or seek to provide students with leadership skills sought by employers, embedding large student teams in faculty research provides a scalable model that benefits both faculty and students. While leadership development was the focus of this study, the VIP model was developed to support faculty research and to enable students to develop disciplinary skills, professional skills, and to contribute to meaningful projects. For programs such as this to succeed, *departments need to enable faculty and students to participate*. Departments that enroll students in VIP or large student teams *should provide faculty with teaching-release time*. At Georgia Tech, research-active Electrical and Computer Engineering (ECE) faculty teach three courses per year. The ECE department releases VIP instructors from *1 course per year,* producing more VIP instructors than any other academic department on campus. If 1 course of release time per instructor per year is not tenable, departments can instead provide release time during the *first two years* of team-establishment. This is when the leadership burden falls more heavily on instructors. Departments can regulate how many new teams are established each year, enabling them to distribute start-up release time over many years. As another model, NYU faculty receive *overload pay* for leading VIP teams. The model was established when teaching loads strained departments during COVID, but has since become the norm.

To achieve the leadership development seen in this study, departments will need to *find space in their curricula* for three to four semesters (~6 credits) of VIP or long-term PBL projects with large teams, and to *incentivize multiple semesters of participation*. In Georgia Tech's CS and CM programs, 3 semesters of VIP can be used to fulfill the CS/CM Junior Design sequence (one of multiple options), and 71% of CS students in the sample participated for three or more semesters. Another credit model that incentivizes multiple semesters of participation is a threshold model, which was used by ECE. In ECE's policy, if students earn 5 or fewer credits, they all count as free electives. After earning a 6th credit, 3 count as ECE-electives and 3 count as free electives. Students are also able to roll their VIP projects into Senior Design, either fully embedded within their VIP teams or by bringing their VIP project to the traditional Senior Design course. A third model was developed by the University of Pretoria in South Africa. There, students can fulfill a campus work-based learning requirement with multiple semesters of VIP [44]. The model solved a problem faced by non-liberal education institutions, which tend to have highly prescriptive degree programs with no electives.

At the institutional level, large-scale long-term PBL embedded in faculty research can provide a context for meaningful leadership development, an area institutions list as a priority but rarely assess [2]. Only a limited number of students can serve in student government or lead student teams/organizations, but every student could participate in a fully scaled VIP Program. With approximately 80 teams at Georgia Tech, 29% of students who graduated in 2022-23 had participated in VIP; Additional faculty continue to request new teams, and enrollments continue to increase. A number of papers have detailed different aspects of VIP Program establishment and expansion [32], [43], and the VIP Consortium provides an annual meeting and networks.

Resources needed for team operations differ by institution type and department. At research-intensive institutions, teams are typically embedded in ongoing faculty projects, which *leverages existing resources*. Faculty also include VIP in proposals as broader impacts (educating large/diverse groups of students, etc.). At institutions without active research programs and/or seeking to establish teams in non-research intensive departments, teams may need start-up funding. At Boise State University, *colleges contribute funding to the VIP Program* based on enrollment from their

college, and VIP instructors can submit funding proposals to the VIP Program [32]. In the College of Engineering at Virginia Commonwealth University, a $1M *endowment* from the Altria Corporation provides support for VIP teams [32], and the program is being expanded from the college to the campus level. A VIP institution that is currently restructuring its budget model plans to have a portion of tuition dollars follow students to VIP, to *have VIP redirect funds back to departments to support teams,* and to have departments use funds in ways that meet department and faculty needs (course-release time for instructors, materials and supplies, etc.). Under a responsibility center management model, this approach would prevent perceived competition for tuition dollars, because funds will make their way back to departments.

IV. Conclusion

Project-based learning has been employed in higher education, engineering education, and computer science education for decades [5], [6]. This study fills a gap in research on student leadership development and PBL. This study confirmed that in multi-semester PBL involving large teams embedded in faculty research, student leadership increased in the second and third semester of participation. In turn, student leadership decreases the burden on faculty, enabling them to lead large teams, making the model scalable.

If institutions seek to cultivate student leadership development in applied contexts, they cannot rely solely on extracurriculars. Institutions can provide meaningful contexts for leadership development by embedding large student teams in faculty research, allowing students to participate and earn 1-2 credits per semester over multiple semesters, and allowing those credits to fulfill degree requirements.

Acknowledgment

References

[1] William Moner, Phillip Motley, and Rebecca Pope-Ruark, *Redesigning Liberal Education : Innovative Design for a Twenty-First-Century Undergraduate Education*. Baltimore: Johns Hopkins University Press, 2020.

[2] S. R. Komives and M. Sowcik, "How Academic Disciplines Approach Leadership Development," *New Dir. Stud. Leadersh.*, vol. 2020, no. 165, pp. 11–21, Spring 2020.

[3] National Association of Colleges and Employers, "The Key Attributes Employers Seek on Students' Resumes." Accessed: Mar. 15, 2022.

[4] V. Chunoo and L. Osteen, "Purpose, Mission, and Context: The Call for Educating Future Leaders," *New Dir. High. Educ.*, vol. 2016, no. 174, pp. 9–20, 2016, doi: 10.1002/he.20185.

[5] National Academies of Sciences, Engineering, and Medicine, *Imagining the Future of Undergraduate STEM Education: Proceedings of a Virtual Symposium*. Washington, DC: The National Academies Press, 2022. doi: 10.17226/26314.

[6] K. Cain and S. Cocco, "Leadership Development through Project Based Learning," presented at the Canadian Engineering Education Association Conference, Montréal, Canada, Jun. 2013.

[7] R. K. Yeo, "Problem-based learning: a viable approach in leadership development?," *J. Manag. Dev.*, vol. 26, no. 9, pp. 874–894, 2007, doi: 10.1108/02621710710819357.

[8] K. S. Scott, "An Integrative Framework for Problem-Based Learning and Action Learning: Promoting Evidence-Based Design and Evaluation in Leadership Development," *Hum. Resour. Dev. Rev.*, vol. 16, no. 1, pp. 3–34, Mar. 2017.

[9] Yokota Kazutaka, Maruoka Masatomo, Takaki Junji, and Watanabe Shinichi, "PBL Based Engineering Education to Cultivate Leadership Spirit in Postgraduate Students," *J. Eng. Educ. Res.*, vol. 15, no. 5, pp. 3–7, Sep. 2012.

[10] P. Guo, N. Saab, L. S. Post, and W. Admiraal, "A review of project-based learning in higher education: Student outcomes and measures," *Int. J. Educ. Res.*, vol. 102, p. 101586, Jan. 2020.

[11] J. Rodríguez, A. Laverón-Simavilla, J. M. del Cura, J. M. Ezquerro, V. Lapuerta, and M. Cordero-Gracia, "Project Based Learning experiences in the space engineering education at Technical University of Madrid," *Adv. Space Res.*, vol. 56, no. 7, pp. 1319–1330, Oct. 2015, doi: 10.1016/j.asr.2015.07.003.

[12] J. Berbegal-Mirabent, D. G.-D. Nech, and I. S. Alegre, "Where to Locate? A Project-Based Learning Activity for a Graduate-Level Course on Operations Management," *Int. J. Eng. Educ.*, vol. 33, no. 5, pp. 1586–1597, 2017.

[13] J. S. Vogler, P. Thompson, D. W. Davis, B. E. Mayfield, P. M. Finley, and D. Yasseri, "The hard work of soft skills: augmenting the project-based learning experience with interdisciplinary teamwork," *Instr. Sci.*, vol. 46, no. 3, pp. 457–488, Jun. 2018.

[14] J. Sonnenberg-Klein and E. J. Coyle, "Experience vs. Academic Rank: Leadership Growth over Multiple Semesters in Vertically Integrated Project (VIP) Courses," in *2022 IEEE Frontiers in Education Conference (FIE)*, Oct. 2022, pp. 1–5.

[15] C. Seemiller, *The Student Leadership Competencies Guidebook: Designing Intentional Leadership Learning and Development*. Somerset, UNITED STATES: John Wiley & Sons, Incorporated, 2013. Accessed: Mar. 16, 2022.

[16] A. C. Barnes, "Higher Education and Student Affairs: Applying Expertise in Student Leadership Development and the Collegiate Context," *New Dir. Stud. Leadersh.*, vol. 2020, no. 165, pp. 99–112, Mar. 2020, doi: 10.1002/yd.20372.

[17] D. M. Rosch and J. D. Collins, "The Significance of Student Organizations to Leadership Development," *New Dir. Stud. Leadersh.*, vol. 2017, no. 155, pp. 9–19, 2017.

[18] D. L. Reyes, J. Dinh, C. N. Lacerenza, S. L. Marlow, D. L. Joseph, and E. Salas, "The state of higher education leadership development program evaluation: A meta-analysis, critical review, and recommendations," *Leadersh. Q.*, vol. 30, no. 5, p. 101311, Oct. 2019, doi: 10.1016/j.leaqua.2019.101311.

[19] R. Paul and L. G. C. Falls, "Engineering Leadership Education: A Review of Best Practices," presented at the 2015 ASEE Annual Conference & Exposition, Jun. 2015, p. 26.634.1-26.634.11. Accessed: Oct. 04, 2023.

[20] "VIP Consortium | Vertically Integrated Projects." Accessed: Mar. 24, 2021. [Online]. Available: http://vip-consortium.org/

[21] S. H. Russell, M. P. Hancock, and J. McCullough, "Benefits of Undergraduate Research Experiences," *Science*, vol. 316, no. 5824, pp. 548–549, 2007.

[22] Association of American Colleges and Universities, "College learning for the new global century," Association of American Colleges and Universities, Washington, DC, 2007. Accessed: Nov. 09, 2017.

[23] J. S. Krajcik and P. C. Blumenfeld, "Project-based Learning," in *The Cambridge Handbook of Learning Sciences*, R. K. Sawyer, Ed., Cambridge University Press, 2005, pp. 317–34.

[24] R. C. Reece, "Letter of Support for VIP, University System of Georgia Award Nomination," May 13, 2014.

[25] J. L. Huff, C. B. Zoltowski, and W. C. Oakes, "Preparing Engineers for the Workplace through Service Learning:

Perceptions of EPICS Alumni," *J. Eng. Educ.*, vol. 105, no. 1, pp. 43–69, Jan. 2016, doi: 10.1002/jee.20107.

[26] J. Ludlum, "Georgia Tech Exit Survey Research Brief: VIP Program Comparisons," Georgia Institute of Technology, Atlanta, GA, 2015.

[27] J. Sonnenberg-Klein, R. T. Abler, E. J. Coyle, and H. H. Ai, "Multidisciplinary Vertically Integrated Teams: Social Network Analysis of Peer Evaluations for Vertically Integrated Projects (VIP) Program Teams," in *2017 ASEE Annual Conference & Exposition*, Columbus, Ohio, Jun. 2017. doi: 10.18260/1-2--28697.

[28] J. Sonnenberg-Klein, R. T. Abler, and E. J. Coyle, "Social Network Analysis: Peer Support and Peer Management in Multidisciplinary, Vertically Integrated Teams," in *2018 ASEE Annual Conference & Exposition*, Salt Lake City, Utah, Jun. 2018.

[29] J. Sonnenberg-Klein *et al.*, "Dimensions of Equity: Undergraduate Research through Vertically Integrated Projects at Five Institutions," in *2023 IEEE Frontiers in Education Conference (FIE)*, College Station, TX, Oct. 2023, pp. 1–5.

[30] J. Sonnenberg-Klein, R. T. Abler, and E. J. Coyle, "Diversity and Student Persistence in the Vertically Integrated Project (VIP) Course Sequence," in *2018 CoNECD - The Collaborative Network for Engineering and Computing Diversity Conference*, Crystal City, Virginia, Apr. 2018.

[31] J. Sonnenberg-Klein, R. T. Abler, and E. J. Coyle, "Correlation between Academic Credit-use Policies and Student Persistence in Vertically Integrated Project (VIP) Courses," in *2018 ASEE Annual Conference & Exposition*, Salt Lake City, Utah, Jun. 2018. doi: 10.18260/1-2--30227.

[32] M. E. Amoo *et al.*, "Breaking Down the Silos: Innovations for Multidisciplinary Programs," in *2020 ASEE Virtual Annual Conference*, Jun. 2020. doi: 10.18260/1-2--34224.

[33] B. W. Tuckman, "Developmental sequence in small groups," *Psychol. Bull.*, vol. 63, no. 6, pp. 384–399, Jun. 1965.

[34] B. W. Tuckman and M. A. C. Jensen, "Stages of Small-Group Development Revisited," *Group Organ. Stud. Pre-1986*, vol. 2, no. 4, pp. 419–427, Dec. 1977.

[35] D. A. Bonebright, "40 years of storming: a historical review of Tuckman's model of small group development," *Hum. Resour. Dev. Int.*, vol. 13, no. 1, pp. 111–120, Feb. 2010.

[36] L. R. Offermann and R. K. Spiros, "The Science and Practice of Team Development: Improving the Link," *Acad. Manage. J.*, vol. 44, no. 2, pp. 376–392, 2001, doi: 10.2307/3069462.

[37] K. Cassidy, "Tuckman Revisited: Proposing a New Model of Group Development for Practitioners," *J. Exp. Educ.*, vol. 29, no. 3, pp. 413–417, Sep. 2007, doi: 10.1177/105382590702900318.

[38] T. Cresswell-Yeager, "Forming, storming, norming, and performing: Using a semester-long problem-based learning project to apply small-group communication principles," *Commun. Teach.*, vol. 35, no. 2, pp. 155–165, Apr. 2021.

[39] G. A. Neuman, J. Wright, G. A. Neuman, and J. Wright, "Team effectiveness: beyond skills and cognitive ability," *J. Appl. Psychol.*, vol. 84, no. 3, pp. 376–389, Jun. 1999.

[40] A. P. Field, *Discovering statistics using IBM SPSS statistics*, 4th edition. Sage, 2013.

[41] R. H. Heck, S. L. Thomas, and L. N. Tabata, *Multilevel and Longitudinal Modeling with IBM SPSS*, 3rd ed. in Qualitative Methodology Series. New York, NY: Routledge, 2022.

[42] C. K. Boscardin, S. S. Sebok-Syer, and M. V. Pusic, "Statistical points and pitfalls: growth modeling," *Perspect. Med. Educ.*, vol. 11, no. 2, pp. 104–107, Mar. 2022, doi: 10.1007/s40037-022-00703-1.

[43] B. Aazhang *et al.*, "Vertically Integrated Projects (VIP) Programs: Multidisciplinary Projects with Homes in Any Discipline," in *2017 ASEE Annual Conference and Exposition*, Columbus, Ohio, Jun. 2017. doi: 10.18260/1-2--29103.

[44] J. Bringardner *et al.*, "Level-up: Expand Undergraduate Research Capacity (and Serve Faculty) through Vertically Integrated Projects," presented at the Frontiers in Education Conference, Uppsala, Sweden, Oct. 2022.

Julia Sonnenberg-Klein earned a B.S. in Engineering Physics and M.Ed. in Education Organization and Leadership from the University of Illinois Urbana Champaign, Urbana, Illinois, USA in 1999 and 2013 respectively. She is a Ph.D. candidate in Education Policy Studies with an emphasis in Research, Measurement and Statistics at Georgia State University, Atlanta, Georgia, USA.

She is currently Executive Director of the Vertically Integrated Projects Program at the Georgia Institute of Technology, Atlanta, Georgia, USA. She has published papers on student interaction within teams, equity, and the impact of policies on student participation in VIP.

Ms. Klein is a member of the American Association for Engineering Education and is program chair for the annual meeting of the VIP Consortium. She received the Innovation and Process Improvement Award from the Georgia Tech College of Engineering (2022); the Outstanding Research, Measurement and Statistics Ph.D. Student Award for the Department of Educational Policy Studies at Georgia State University (2022); was a co-recipient of the 2020 Campus Sustainability Research Award from the Association for the Advancement of Sustainability in Higher Education; and was a co-recipient of the 2019 Innovation Award from ABET.

Edward J. Coyle (Life Fellow, IEEE) received the B.S.E.E. degree from the University of Delaware, Newark, in 1978, and the Ph.D. degree in EECS from Princeton University, Princeton, NJ, in 1982.

He is currently the J.B. Peatman Distinguished Professor of Electrical and Computer Engineering at Georgia Tech and a Georgia Research Alliance Eminent Scholar. He also serves as the Director of the VIP Program at Georgia Tech and as Chair of the Board of the nonprofit VIP Consortium, LLC. At Georgia Tech in Spring 2023, there were 86 VIP Teams with an enrollment of 1770 students and all colleges were participating in the program. 45 institutions around the world currently have VIP Programs.

Dr. Coyle was a co-recipient of the 2020 Campus Sustainability Research Award from the Association for the Advancement of Sustainability in Higher Education, the 2019 Innovation Award from ABET, and the National Academy of Engineering's 2005 Bernard M. Gordon Award for Innovation in Engineering and Technology Education. In 1998, he was named an Outstanding Engineering Alumnus of the University of Delaware. He was a co-recipient of both the ASEE's 1997 Chester F. Carlson Award for Innovation in Engineering and the IEEE Signal Processing Society's 1986 Best Paper Award for Authors under age 30. Dr. Coyle was a designated a Fellow of the IEEE in 1998 for his contributions to the theory of nonlinear signal processing. His research interests include systemic reform of higher education, signal and information processing, and cognitive networks.