Meaningful Failure: Transforming Academic Structures and Incentives through Personalization in Engineering Education

Runu Proma Das, University of Georgia
Yeongji Lee, Dartmouth College
Tathyana Moratti, The University of Georgia

PhD student

Spencer Johnathan Sabatino, Cornell University

Spencer Sabatino is a PhD candidate at the Robert Frederick Smith School of Chemical and Biomolecular Engineering at Cornell University. Spencer's research interests include studying engineering students' beliefs and their effects on broader student outcomes. He earned his Bachelor of Science degree in Chemical Engineering and his Master of Science in Chemical and Biomedical Engineering from Miami University in 2022.

David Zabner, Tufts University

David Zabner is a PhD student at Tufts university where he focuses on computer science and engineering education

Mathew Baby, University of Georgia
Katherine Becking, Cornell University
Dr. Herman Ronald Clements III, University of Nebraska - Lincoln
Michael Brown, University of Michigan
Prof. Allison Godwin, Cornell University

Allison Godwin, Ph.D. is the Dr. G. Stephen Irwin '67, '68 Professor in Engineering Education Research in the Robert Frederick Smith School of Chemical and Biomolecular Engineering at Cornell University. She is also the Associate Director of the Cornell NanoScale Science and Technology Facility and a McCormick Teaching Excellence Institute Research Fellow. Her research focuses on how identity, among other affective factors, influences diverse groups of students to choose engineering and persist in engineering. She also studies how different experiences within the practice and culture of engineering foster or hinder belonging, motivation, and identity development. Dr. Godwin graduated from Clemson University with a B.S. in Chemical Engineering and Ph.D. in Engineering and Science Education. Her research earned her a National Science Foundation CAREER Award focused on characterizing latent diversity, which includes diverse attitudes, mindsets, and approaches to learning to understand engineering students' identity development. She has won several awards for her research including the 2021 Chemical Engineering Education William H. Corcoran Award, 2022 American Educational Research Association Education in the Professions (Division I) 2021-2022 Outstanding Research Publication Award, and the 2023 AIChE Excellence in Engineering Education Research Award.

Trevion S Henderson, Tufts University

Trevion Henderson is Assistant Professor of Mechanical Engineering at Tufts University. He earned his Ph.D. in Higher Education at the University of Michigan.

Dr. Beshoy Morkos, University of Georgia

Beshoy Morkos is a professor in the College of Engineering at the University of Georgia where he directs MODEL (Manufacturing Optimization, Design, and Engineering Education Lab) Group. His research group currently explores the areas of system design, manufacturing, and their respective education. His system design research focuses on developing computational representation and reasoning support for managing complex system design through the use of Model Based approaches. The goal of Dr. Morkos' manufacturing research is to fundamentally reframe our understanding and utilization of product and

process representations and computational reasoning capabilities to support the development of models which help engineers and project planners intelligently make informed decisions. On the engineering education front, Dr. Morkos' research explores means to improve persistence and diversity in engineering education by leveraging students' design experiences.

Dr. Vitaliy Popov, University of Michigan

Vitaliy Popov is an Assistant Professor of Learning Health Sciences at the University of Michigan Medical School. His research focuses on understanding, designing, and evaluating learning technologies and environments that foster collaborative problem solving, spatial reasoning, engineering design thinking and agency. He is currently serving as a co-principal investigator on three projects funded by the National Science Foundation ranging from studying visuospatial skills development through origami to applying multimodal learning analytics in teamwork and understanding the mechanisms of an A-ha! moment. Dr. Popov completed his Ph.D. on computer-supported collaborative learning at Wageningen University & Research Center, in the Netherlands. His background allows him to utilize evidence in education science, simulation-based training and learning analytics to understand how people become expert health professionals, how they can better work in teams and how we can support these processes to foster health care delivery and health outcomes.

David J. M. Kraemer, Dartmouth College

EFRI Ideas Lab: PEL: Meaningful Failure: Transforming Academic Structures and Incentives through Personalization in Engineering Education

1. Introduction

This project focuses on identifying ways to detect individual failure signals in real-time to develop personalized teaching and learning tools for engineering students, thereby improving engagement with “meaningful failure” (MF). MF is a novel approach that encourages students to embrace uncertainty, learn from setbacks, and ultimately achieve success through navigating failure moments connected to who they are [1], [2]. We operationalize failure as both process-oriented, an impasse in learning task performance that occurs when learners lack the necessary knowledge to move forward, and outcome-oriented, a gap between desired goals and reality that can occur because of the process itself, maladaptive behaviors in response to impasse, or negative affective responses. In this research, we adopt an approach to understand MF through three strands: (1) in-vitro laboratory experiments to study cognitive, emotional, and physiological responses to failure, (2) in-situ observational studies across various engineering learning environments, and (3) administrative and socio-cultural studies to understand the conceptualization and interpretation of failure by students, instructors, and administrators.

Engineering, as a discipline, is fundamentally oriented toward designing, developing, and constructing functional socio-technical solutions to problems. Achieving these outcomes requires iterative cycles of design, construction, testing, modeling or simulation, and calculation to ensure that a solution satisfies criteria and constraints. These practices are grounded in an ethos, process, and philosophy that inherently relies on failure [3]. Nevertheless, failure is conspicuously underrepresented in both engineering curricula and in the educational tools designed to support student learning.

Our approach departs from conventional instructional models by reconsidering pedagogical practices and motivational structures that often adopt uniform strategies to enhance engineering education at a collective level. This framework seeks to create conditions in which failure can function as a meaningful and individual signal of learning. We conceptualize a *failure profile* as an individualized data structure derived from multiple relevant signals (affective, physiological, and behavioral) that captures a learner’s reactions, both adaptive and maladaptive, to failure experiences. Such profiles may support efforts to distinguish between productive failure, where a student is learning from failure, and instances in which additional scaffolding or intervention may be warranted. The Meaningful Failure framework could provide timely, targeted support aimed at mitigating the negative emotional consequences often associated with failure in engineering education.

The experience of failure is profoundly individual, shaped by a learner’s prior experiences and their affective, behavioral, cognitive, and contextual states at a given moment. As a result, even within the same learning environment, outcomes can vary substantially across individual learners and learner groups [4]. Notably, the top three reasons students cite for leaving STEM fields are: 1) a decline in initial interest and motivation, 2) ineffective STEM instruction, and 3) challenges in seeking and receiving timely, appropriate assistance [5].

In this project, we aim to answer the following research questions across the three research strands: **RQ1**: What are the useful real-time failure profile signals that would yield actionable personalization in engineering education contexts? **RQ2**: How do students' differential responses to failure present opportunities for personalization of support in engineering education contexts? **RQ3:** What changes to pedagogy, assessment, policy, and instructor professional development might be necessary to facilitate a culture of learning from and through failure in engineering education?

2. Theoretical Background

Productive Failure (PF) is a well-documented instructional approach that emphasizes the value of supporting learners' abilities to solve complex, ill-structured problems [1], [3]. PF highlights the potential of failure experiences as an essential tool for learning, embedding cognitive processes of long-term memory storage, attention, and metacognition [1], [3]. The PF literature has motivated scholars to conceptualize and evaluate how STEM students cultivate dispositions, including mindset, goal orientations, attributions, and affective factors, such as fear of failure [6]. Despite the recognized potential of failure as a productive teaching and/or learning strategy, it is frequently stigmatized and avoided when designing learning activities in classrooms [7] driven by concerns regarding the potential negative consequences associated with students' experiences of failure.

The PF framework [1], [8] demonstrates that engaging learners in ill-structured problem solving, an essential component of engineering practice prior to receiving direct instruction yields learning gains with effect sizes two to three times greater than those achieved through traditional instruction [2]. Additional empirical work provides strong evidence for the role of "desirable difficulties" in optimizing learning, promoting both transfer to novel problems and long-term retention [9]. Such difficulties include intentionally introducing variability [10] or unpredictability into the learning phase, interleaving practice across key concepts rather than postponing challenges until instruction has concluded, withholding immediate feedback during early learning stages, and using assessments as opportunities for continued learning rather than solely as measures of performance.

Although these approaches may initially appear counterintuitive, particularly given that instructors are typically encouraged to provide scaffolding and timely feedback to facilitate learning [11], a substantial body of research indicates that temporarily withholding structure can, in fact, enhance learning. By engaging in open-ended inquiry, learners are afforded the opportunity to generate their own concepts, representations, and explanations, thereby deepening understanding [12]-[17]. These findings align with well-established practices in engineering education, including problem-based, project-based, and inquiry-based learning [18], [19], [20].

However, the benefits associated with productive failure must be balanced with contextual factors such as cognitive load, stress and anxiety, demotivation, and individual learner differences [6], [7], [21] to ensure optimal outcomes. To effectively support learners, it is crucial to recognize the boundaries of what they can accomplish independently and to provide targeted assistance that enables iterative engagement with challenging tasks, thereby fostering deep learning and eventual success.

Nonetheless, existing studies do not yet account for the full range of factors that shape failure experiences, nor do they offer personalized, real-time supports that provide learners with precisely the assistance needed to foster adaptive responses and positive long-term outcomes. Instead, most investigations rely on retrospective analyses of failure, generating broad conclusions that may not adequately represent the diverse experiences of learners across different contexts. In this project, we build on PF framing to consider how unique psychosocial experiences, prior knowledge, and goals shape how students perceive and engage with failure. We have termed this personalized shift in the PF framework as Meaningful Failure (MF). In this conceptualization, we consider the progression from the pre-failure phase, where an engineering student's characteristics and academic situations come into play, through the failure experience phase, where cognitive mechanisms of failure operate, and ultimately to the post-failure phase, where affective and cognitive processes shape coping behaviors and subsequent outcomes.

3. Methods

Through a single IRB process, five engineering institutions are 1) exploring how the cognitive and affective states that students experience are reflected in biosignals that can be analyzed in real-time to understand how individual students respond differently to failure across different learning and social contexts (in-vitro strand); 2) documenting, capturing, and analyzing real-time interactions of students and instructors in the classroom, including moments of individual-level and group-level failures and successes during the experience of regular activities in several engineering courses (in-situ strand); and 3) determining the best ways to translate the insights gained into directions for future development of systems that educators and administrators can use to address the key issues in promoting meaning failure and long-term success for engineering students (administrative and socio-cultural strand).

Because each strand approaches MF differently, we employ a range of methodological approaches to address the three research questions. The in-vitro strand examined undergraduate students in a controlled laboratory environment using multimodal biosensing and behavioral measures. Participants completed an operation span task at three difficulty levels (counterbalanced order, ~40 minutes total) while we collect multimodal physiological signals synchronized via Lab Streaming Layer: four-channel EEG (Muse S headband) [22], ECG and EOG (OpenBCI Cyton and Daisy board) [23], PPG and EDA (Emotibit) [24], and facial expressions via webcam. Participants also completed a NASA-TLX questionnaire to report their subjective cognitive load. All features were baseline-corrected using a 3-minutes resting baseline to account for inter-individual variability. We extracted 43 action units and 7 emotions from facial data (PyFeat 0.6.2), and 15 physiological features including HR and HRV from ECG, theta and alpha power from each EEG channel (MNE 1.9.0), HR from PPG, and EDA metrics (Neurokit 0.2.12). These multimodal data enabled analysis of neural, physiological, and affective signatures related to cognitive load during problem-solving.

The in-situ strand involved classroom observations, collecting video, audio, EDA, and field notes supplemented with student surveys and interviews, yielding both quantitative and qualitative insights into students' experiences with failure in authentic instructional contexts. The administrative and socio-cultural strand leverages interviews with national administrators and educators to provide perspectives on current beliefs, policies, and institutional practices surrounding failure (see Table 1).

4. Results

Over the past 20 months the project has developed: 1) a standardized protocol for multi-modal physiological data collection, which enabled the collection of 18 participants' physiological data; 2) a survey instrument focusing on key psychosocial constructs, which enabled data collection from 167 participants and a workflow to perform latent profile analyses (LPA) from the survey data; 3) An interview protocol designed to understand students experiences of and beliefs around failure, which enable qualitative data collection through 15 interviews; 4) A protocol for collecting in-situ biometric and audiovisual data collection, which enabled the collection of biometric and audiovisual data on seven students during class periods; 5) eight interviews of associate deans or engineering department chairs.

Table 1. Methodologies used by each strand

Strand	In-vitro	In-situ	Administrative and Socio-Cultural
Methods	Biosignals data collection	Video/audio recording interviews, observations, surveys, and biosignals data analysis	Video/audio recording interviews, and biosignals data analysis
Instruments	Biosignals devices (Emotibit, Muse S, OpenBCI, and Webcam)	Biosignal device (Emotibit), Camera (GoPro), and Microphone (Boya)	Biosignal device (Emotibit), and Microphone (Boya)
Sample size	50	Survey: 167, Interviews: 15, and Observations: 10 classroom events	30
Data Analysis	Python codes to clean, filter, and segment data. Machine learning to identify and predict individual failure patterns	Latent profile analysis of survey responses. Critical incident techniques. In-vivo analysis of interviews. Random-forests modeling of integrated data.	Thematic analysis of failure definitions and considerations. Needs assessment for personalization tool design.

Additionally, the significant results of the first-year project were: 1) a Support Vector Machine classifier to test accuracy and mental load on a trial-by-trial basis during a demanding cognitive task, relying on different signal patterns for different students; 2) four classes of engineering student failure response related to engagement and identity were defined; 3) results indicating that at small institutions failure is viewed very negatively and mostly in the mode of academic failure (e.g., failing a course, dropping out, etc.) while these same concerns were not observed at larger research-intensive institutions.

4.1. In-vitro strand results overview

A pilot of lightweight, off-the-shelf physiological sensors indicated that they can accurately track performance during cognitive challenges in fine-grained time windows (15-second epochs). Additionally, a machine learning pipeline using Support Vector Machine (SVM) classifiers was developed to identify individual physiological signatures of cognitive struggle and failure. The analysis showed above-chance classification accuracy (mean accuracy: 59.9%). It also identified three key physiological predictors of performance: ECG heart rate variability (average F-score: 2.96), alpha power at TP10 (2.34), and theta power at TP10 (2.39). Finally, a feature importance analysis revealed highly individualized failure signatures, with participants showing distinct patterns of physiological responses to cognitive struggle. The results of these tests highlight which signals may be most useful in real-time analysis of student failure responses.

4.2. In-situ strand results overview

Survey data were collected in Spring 2025 and Fall 2025 across seven courses at three institutions. Strong correlations were observed between student responses within "positive" constructs, such as engineering role identity, mastery goal orientation (MGO), and error-learning orientation, and "negative" constructs, such as fixed mindset, stress response, and fear of failure (FoF). Following latent profile analysis, four distinct profiles were generated using the survey metrics. Profile 1 (n = 19) contains students with high FoF values and lower MGO values. Profile 2 (n = 49) contains students with very low FoF values and high MGO values. Profile 3 (n = 70) contains students with high MGO values but moderate FoF values, and Profile 4 (n = 29) contains students with roughly equal MGO and FoF values. These profiles provided preliminary insights into how students' psychosocial parameters may predict their failure responses.

In interview data (n = 15), students discussed how they define and experience failure. Students primarily defined failure in three different dimensions: 1) failing to live up to external benchmarks, such as exam averages or matching peer grades, 2) failing to live up to internal benchmarks rooted in prior experience or identities held, such as internal perceptions of effort or engagement, and 3) performance resulting in an outcome or event that disrupts a student's plans, such as internships or graduate school. These definitions often overlapped and influenced each other, and can serve as ways to develop student-centered approaches to integrating positive failure experiences in engineering classrooms.

4.3 Administrative strand results overview

Analysis of initial interviews with nine academic leaders suggested a focus on how our proposed solution would fit into existing pedagogical and curricular initiatives. Respondents expressed concern about the time required and the cognitive load involved in navigating a new instructional tool. Initial findings suggest that tools need to be designed to ease instructional burden and to facilitate pedagogical decision-making to encourage adoption and diffusion.

5. Future Work

In-vitro strand data collection and analysis are ongoing with the goal of accurately predicting failure responses. The insights from the in-vitro strand will be leveraged within the in-situ strand to focus efforts on the signals that may be most important to attend to in a classroom setting. Multiple streams of data from the in-situ strand are being analyzed collectively to develop a nuanced understanding of how individuals across contexts experience failure. Finally, the administrative and socio-cultural strand is continuing interviews with faculty to gather insights for those who would be directly engaged in using tools developed from this work.

Using the full range of data sources collected in the project, we will construct multimodal indicators that form the basis of personalized learner profiles. Qualitative data will be transformed into quantitative features to support modeling. Advanced machine learning models will be applied to determine the relative importance of contextual and individual factors in predicting coping responses and both short- and long-term outcomes, as well as to estimate individualized treatment effects (ITEs), enabling the prediction of how students—including those who did not naturally encounter meaningful failure—might respond under such conditions. This combined approach will support the future development of adaptive pedagogical tools tailored to students' characteristics, contexts, and needs, ultimately informing scalable interventions that leverage MF as a useful approach to personalization and learning.

Acknowledgments

This research is funded by NSF Grants Nos. 2432011, 2432012, 2432013, 2432014, and 2432015.

References

[1] M. Kapur, “Productive failure,” *Cogn. Instr.*, vol. 26, no. 3, pp. 379–424, Jul. 2008, doi: 10.1080/07370000802212669.

[2] T. Sinha and M. Kapur, “When Problem Solving Followed by Instruction Works: Evidence for Productive Failure,” *Rev. Educ. Res.*, vol. 91, no. 5, pp. 761–798, Oct. 2021, doi: 10.3102/00346543211019105.

[3] A. Jackson, A. Godwin, S. Bartholomew, and N. Mentzer, “Learning from failure: A systematized review,” *Int. J. Technol. Des. Educ.*, vol. 32, no. 3, pp. 1853–1873, Jul. 2022, doi: 10.1007/s10798-021-09661-x.

[4] S. L. Eddy and S. E. Brownell, “Beneath the numbers: A review of gender disparities in undergraduate education across science, technology, engineering, and math disciplines,” 2016, *American Physical Society*. doi: 10.1103/PhysRevPhysEducRes.12.020106.

[5] E. Seymour *et al.*, *Talking about Leaving Revisited*. Cham: Springer International Publishing, 2019. doi: 10.1007/978-3-030-25304-2.

[6] M. A. Henry, S. Shorter, L. Charkoudian, J. M. Heemstra, and L. A. Corwin, “FAIL is not a four-letter word: A theoretical framework for exploring undergraduate students’ approaches to academic challenge and responses to failure in STEM learning environments,” *CBE Life Sci. Educ.*, vol. 18, no. 1, Mar. 2019, doi: 10.1187/cbe.18-06-0108.

[7] A. Simpson and A. Maltese, “‘Failure Is a Major Component of Learning Anything’: The Role of Failure in the Development of STEM Professionals,” *J. Sci. Educ. Technol.*, vol. 26, no. 2, pp. 223–237, Apr. 2017, doi: 10.1007/s10956-016-9674-9.

[8] M. Kapur, “Examining Productive Failure, Productive Success, Unproductive Failure, and Unproductive Success in Learning,” *Educ. Psychol.*, vol. 51, no. 2, pp. 289–299, Apr. 2016, doi: 10.1080/00461520.2016.1155457.

[9] R. A. Schmidt and R. A. Bjork, “New Conceptualizations of Practice: Common Principles in Three Paradigms Suggest New Concepts for Training,” *Psychol. Sci.*, vol. 3, no. 4, pp. 207–218, Jul. 1992, doi: 10.1111/j.1467-9280.1992.tb00029.x.

[10] E. Butowska-Buczyńska, P. Kliś, K. Zawadzka, and M. Hanczakowski, “The role of variable retrieval in effective learning,” *Proc. Natl. Acad. Sci. U. S. A.*, vol. 121, no. 44, Oct. 2024, doi: 10.1073/pnas.2413511121.

[11] R. E. Mayer, “Should There Be a Three-Strikes Rule against Pure Discovery Learning? The Case for Guided Methods of Instruction,” Jan. 2004. doi: 10.1037/0003-066X.59.1.14.

[12] A. A. Disessa, D. Hammer, B. Sherin, and T. Kolpakowski, “Inventing Graphing: Meta-Representational Expertise in Children,” 1991.

[13] M. Kapur and K. Bielaczyc, “Designing for Productive Failure,” *Journal of the Learning Sciences*, vol. 21, no. 1, pp. 45–83, Jan. 2012, doi: 10.1080/10508406.2011.591717.

[14] R. Lesh and H. M. Doerr, “In what ways does a models and modeling perspective move beyond constructivism?,” in *Beyond constructivism*, Routledge, 2003, pp. 519–556.

[15] K. Loibl and N. Rummel, “Knowing what you don’t know makes failure productive,” *Learn. Instr.*, vol. 34, pp. 74–85, Aug. 2014, doi: 10.1016/j.learninstruc.2014.08.004.

[16] N. J. Slamecka and P. Graf, “The Generation Effect: Delineation of a Phenomenon,” 1978.

[17] K. VanLehn, S. Siler, C. Murray, T. Yamauchi, and W. B. Baggett, “Why Do Only Some Events Cause Learning During Human Tutoring?,” *Cogn. Instr.*, vol. 21, no. 3, pp. 209–249, 2003, doi: 10.1207/S1532690XCI2103_01.

[18] K. A. Smith, S. D. Sheppard, D. W. Johnson, and R. T. Johnson, “Pedagogies of engagement: Classroom-based practices,” in *Journal of Engineering Education*, Wiley-Blackwell Publishing Ltd, 2005, pp. 87–101. doi: 10.1002/j.2168-9830.2005.tb00831.x.

[19] C. E. Hmelo-Silver, R. G. Duncan, and C. A. Chinn, “Scaffolding and achievement in problem-based and inquiry learning: A response to Kirschner, Sweller, and Clark (2006),” 2007, *Routledge*. doi: 10.1080/00461520701263368.

[20] M. Prince, “Does active learning work? A review of the research,” 2004, *Wiley-Blackwell Publishing Ltd*. doi: 10.1002/j.2168-9830.2004.tb00809.x.

[21] N. Heitzmann *et al.*, “Learners’ adjustment strategies following impasses in simulations - Effects of prior knowledge,” *Learn. Instr.*, vol. 83, Feb. 2023, doi: 10.1016/j.learninstruc.2022.101632.

[22] Krigolson, O. E., Williams, C. C., Norton, A., Hassall, C. D., & Colino, F. L. (2017). Choosing MUSE: Validation of a Low-Cost, Portable EEG System for ERP Research. Frontiers in Neuroscience, 11, 109. https://doi.org/10.3389/fnins.2017.00109

[23] Frey, J. (2016). Comparison of an open-hardware electroencephalography amplifier with medical grade device in brain-computer interface applications. arXiv preprint arXiv:1606.02438.

[24] Montgomery, S. M., Nair, N., Chen, P., & Dikker, S. (2023). Introducing EmotiBit, an open-source multi-modal sensor for measuring research-grade physiological signals. Science Talks, 6, 100181. https://doi.org/10.1016/j.sctalk.2023.100181.