WINTER
2026



JOURNAL OF MANAGEMENT & ENGINEERING INTEGRATION

AIEMS
VOL 19
NO 2

ISSN:
1939-7984(Print)
3065-1433 (Online)

JOURNAL OF MANAGEMENT AND ENGINEERING INTEGRATION

Editor-in-Chief

Edwin Sawan, Ph.D., P.E.
Professor Emeritus
Wichita State University
Edwin.sawan@wichita.edu

AIEMS President

Gamal Weheba, Ph.D.
Professor and ASQ Fellow
Wichita State University
gamal.weheba@wichita.edu

Scope: The Journal of Management and Engineering Integration (JMEI) is a double-blind refereed journal dedicated to exploring the nexus of management and engineering issues of the day. JMEI publishes two issues per year, one in the Summer and another in Winter. The Journal's scope is to provide a forum where engineering and management professionals can share and exchange their ideas for the collaboration and integration of Management and Engineering research and publications. The journal will aim on targeting publications and research that emphasizes the integrative nature of business, management, computers and engineering within a global context.

EDITORIAL REVIEW BOARD

Sue Abdinnour Ph.D.
Wichita State University, USA
sue.abdinnour@wichita.edu

Gordon Arbogast, Ph.D.
Jacksonville University, USA
garboga@ju.edu

Deborah Carstens, Ph.D.
Florida Institute of Technology, USA
carstens@fit.edu

Martí Casadesús, Ph.D.
University of Girona, Spain
marti.casadesus@udg.edu

Hossein Cheraghi, Ph.D.
West New England University, USA
cheraghi@wne.edu

Mohammed Darwish, Ph.D.
Kuwait University, Kuwait
m.darwish@ku.edu.kw

Pedro Domingues, Ph.D.
University of Minho, Portugal
Pedrodomin1@sapo.pt

Aneta Kucinska-Landwojtowicz, Ph.D.
Opole University of Technology, Poland
Prorwsp@po.edu.pl

Pradeep Kumar, Ph.D.
Indian Institute of Technology, India
pradeep.kumar@me.iitr.ac.in

Luiz Carlos de Cesaro Cavaler, Ph.D.
Federal University of Rio Grande do Sul, Brazil
luiz.cavaler@satc.edu.br

Mohammad Kanan, Ph.D.
University of Business & Technology, KSA
m.kanan@ubt.edu.sa

Tamer Mohamed, Ph.D.
The British University in Egypt, Egypt
tamer.mohamed@bue.edu.eg

Bilal Obeidat, Ph.D.
Al Ain University, UAE
bilal.obeidat@aau.ac.ae

Nabin Sapkota, Ph.D.
Northwestern State University, USA
sapkotan@nsula.edu

Joseph E. Sawan, Ph.D.
University of Ottawa, Canada
jsawan@uottawa.ca

Scott D. Swain, Ph.D.
Clemson University, USA
sdswain@clemson.edu

Alexandra Schönning, Ph.D.
University of North Florida, USA
aschonni@unf.edu

John Wang, Ph.D.
Montclair State University, USA
wangj@montclair.edu

Gamal Weheba, Ph.D.
Wichita State University, USA
gamal.weheba@wichita.edu

Wei Zhan, D.Sc., PE
Texas A&M University, USA
wei.zhan@tamu.edu

REVIEWERS

The Journal Editorial Team would like to thank the reviewers for their time and effort. The comments that we received were very constructive and detailed. They have been very helpful in our effort to continue to produce a top-quality journal. Your participation and timely response are very important for providing a distinguished outlet for original articles.

Edwin Sawan, Ph.D., P.E.
Editor-in-Chief

Farhana Abedin
Navya Nikhita Agasam
Abdelhakim Al Turk
Fatih Altun
Saleh Ateiwi
Deborah S. Carstens
Muhammet Ceylan
B. Andrew Cudmore
Elizabeth Cudney
Andrzej Gapinski
Deepak Gupta
Abdelnasser Hussein
Arpita Jadav
Bassam Jaradat
Yooneun Lee
Adam Lynch
Lynn Matthews
Wilfredo Moscoso
Anusha Krishna Murthy
Arjun Nair
Paul D Nugent
Clovis Ribas
John Wang
Gamal Weheba

TABLE OF CONTENTS

Richard Sack, Ramazan Asmatulu and Adam Carlton Lynch
ACOUSTIC HOLOGRAM ENABLED CONTROL OF ELECTROSPUN NANOFIBER TRAJECTORY AND DEPOSITION 6

Nicholas Forthofer, Shellbie Reed, Mihir Wadhwa, Vivek Sharma and Brooke E. Wheeler
SPEAKING UP: AERONAUTIC STUDENT VOICE PROPENSITY AND POWER DISTANCE BY EXPERIENCE 16

Ashton Irons, Gianna Mendez, Sunkyum Kim, Vivek Sharma and Brooke E. Wheeler
AIR TRAFFIC CONTROL STAFFING DEFICIENCIES AND THEIR RELATIONSHIP TO INCIDENTS AND ACCIDENTS 28

Deborah Sater Carstens, Arjun M. Nair, Syed Ali Abbas and Anna Berenika Wojdecka
FROM CURB TO GATE: HOW USABLE IS YOUR AIRPORT APP? 39

Faisal S. Cheema and Ahmad Elshennawy
A SYSTEMATIC LITERATURE REVIEW OF INDUSTRY 5.0 CONCEPTS IN BOTTLENECK COLLEGE STEM COURSES 50

Rohan Dharmendrakumar Patel, Joyce Chiam Ziyi and Adam Carlton Lynch
TRACEABILITY WITHOUT COMPLEXITY: A SYSTEMS ENGINEERING APPROACH TO MOTOR-PUMP CORRELATION 60

Yooneun Lee, Ganesh Vaidyaa Ramanathan, KwangWook Gang and Seungbeom Kim
THE IMPACT OF INVESTMENT AMOUNT AND TIMING ON STARTUP SUCCESS: EVIDENCE FROM U.S. SEMICONDUCTOR STARTUPS 70

Scott D. Swain
ONE DEBATE, THREE STANDARDS: ON THE USE OF SYNTHETIC SURVEY RESPONDENTS 83

DOI: 10.62704/10057/59052

ACOUSTIC HOLOGRAM ENABLED CONTROL OF ELECTROSPUN NANOFIBER TRAJECTORY AND DEPOSITION

Richard Sack [1]

Ramazan Asmatulu [1]

Adam Carlton Lynch [1]

[1] Wichita State University

Richard.Sack@wichita.edu; Ramazan.Asmatulu@wichita.edu; Adam.Lynch@wichita.edu

Abstract

Electrospinning is widely used for producing micro- and nanoscale fibers, yet as a manufacturing process it lacks active control over fiber trajectory, offers limited deposition selectivity, and becomes unstable at elevated flow rates. These constraints restrict scalability and complicate the production of aligned or conductive fibers where uncontrolled interaction with grounded collectors degrade process reliability. This work presents a novel engineering system that integrates ultrasonic phased arrays generating acoustic holograms into an electrospinning environment to address these limitations. The proposed platform introduces a non-contact, reconfigurable, field-based control mechanism which operates concurrently with high-voltage electrospinning. The system comprises a phased array of ultrasonic transducers with a modular architecture designed for compatibility with electrostatic fiber formation. Experimental results demonstrate measurable in-flight influence on nanofiber behavior and selective deposition patterns that correlate with simulated acoustic pressure fields. These results establish acoustic hologram guidance as a manufacturing control capability rather than a materials modification technique. This platform provides a foundation for scalable nanofiber production, controlled alignment, multi-fiber yarn formation, and integration with downstream processing steps for various industries (e.g., biomedical, defense, aircraft and energy). Ongoing work focuses on expanded data collection, higher-frequency arrays, expanded geometries, and scaling to support increased throughput and precision.

Keywords: *Acoustic Holograms, Electrospinning, Ultrasonic Phased Arrays, Nanofiber Manufacturing, Selective Deposition, Advanced Manufacturing, Field-Based Control*

1. Introduction

Electrospinning has become a standard process for producing micro- and nanoscale fibers, supporting applications spanning filtration (Johns et al., 2020; Zhou et al., 2022), biomedical scaffolds (Almodovar, 2018; Zhou & Wang, 2021), sensors (Asmatulu et al., 2019), and functional composites (Huang et al., 2003). Despite its maturity in fiber formation, the process remains constrained by a manufacturing systems gap: the trajectory of the charged jet and resulting fiber path between the polymer source and collector is not controllable. In conventional configurations, the flight region is dominated by electrostatic forces, jet instabilities, and environmental perturbations that collectively produce diffuse deposition and limited repeatability. This restricts patterned deposition, robust alignment, and process architectures that require routing fibers and controlled deposition.

Attempts to improve controllability often embed additional electrodes at the emitter (Kim et al.,

Submitted: April 27, 2026
Revised: October 22, 2026

2006; Yang et al., 2010), specialized collectors (Zhu et al., 2015), or fields in the flight region (Bellan & Craighead, 2006; Gu et al., 2007; Zhao et al., 2016). While such methods can improve alignment or increase throughput, they remain tightly coupled to the electrostatic domain, typically introducing new tradeoffs among stability, safety margins, and flexibility. From an advanced manufacturing systems perspective, the central challenge is the absence of an independent layer that can impose trajectory constraints while preserving the electrostatic conditions required for fiber formation.

Ultrasonic phased arrays provide a candidate solution by enabling programmable acoustic pressure fields, “acoustic holograms”. Acoustic fields in air can exert radiation forces (Gor'kov, 1962) and can be reconfigured rapidly via software-defined phase patterns (Marzo et al., 2015; Ochiai et al., 2014). The core premise of this work is that an acoustic field can function as a second, orthogonal control layer within an electrospinning manufacturing system: electrostatics remain responsible for jet formation and stretching, while acoustics impose trajectory constraints and deposition targeting. This decoupling opens a path to modular architectures, such as focused deposition, fiber routing through apertured ground plates, and integration with downstream operations for bundling, coating, or sintering.

This paper contributes to an integrated system demonstration and a manufacturing-systems interpretation of its performance. The novelty of this study lies in three key contributions. Specifically, it (i) reports a scalable phased-array hardware architecture and electromagnetic interference measures required for stable operation in a high-voltage environment, (ii) implements a vortex-based hologram control formulation suitable for embedded execution at scale, and (iii) presents deposition outcomes.

2. Methods

2.1. System Architecture

The system is conceived as a cyber-physical manufacturing platform composed of decoupled subsystems; each assigned a distinct functional role. This architectural separation is central to the objective of isolating fiber formation from trajectory guidance while enabling modular scalability. A supervisory control layer provides operator-level interaction with the manufacturing system. Programs consist of a defined configuration (polymer formulation preparation, flow rate, source position, transducer voltage, phased array program, electrospinning voltage, and distance). The supervisory layer monitors and logs system health indicators to support diagnostics.

2.1.1. Electrospinning Subsystem

The electrospinning subsystem is responsible exclusively for fiber formation and stretching. It operates in an open loop during all experiments, with polymer composition, feed rate, applied high voltage, and source-collector separation selected based on voluminous previous work and held constant within each trial. No auxiliary electrodes, dynamic collectors, or active electrostatic steering mechanisms are employed. This leaves fiber trajectory unregulated by electrostatic means, establishing a baseline process against which the influence of acoustic actuation can be isolated. The subsystem outputs a charged polymer jet whose downstream path is sensitive to external fields but not actively regulated within the electrospinning domain.

2.1.2. Acoustic Actuation Subsystem

The acoustic actuation subsystem provides non-contact trajectory guidance through the

generation of programmable acoustic pressure fields within the electrospinning flight region. A cylindrical ultrasonic phased array surrounds the jet path and superimposes an acoustic field around polymer jets. The subsystem biases fiber transport by shaping pressure gradients that influence fiber trajectory. Field strength is adjusted via transducer drive voltage while acoustic hologram structure is defined by predefined programs. The acoustic subsystem is thus treated as an independent manufacturing control layer whose sole role is control of fiber trajectory and deposition. Distributed ARM Cortex-M7 controllers generate thousands of phase-resolved drive signals for the phased array and are periodically synchronized and self-calibrating. Phase coherence across the array is maintained through a shared timing reference distributed to all controller boards.

2.2. Phased Array Hardware and Electromagnetic Compatibility

Air-coupled ultrasonic transducers (Manorshi MSO-P1040H07T, nominal center frequency 40 kHz) are arranged in stacked rings to form a cylindrical phased array surrounding the electrospinning region. This transducer was selected as it has been widely used in previous work, is readily available, has a low unit cost, and enables dense multi-element arrays. System development progressed through iterative hardware revisions in which transducer density and channel count were progressively increased, culminating in a multi-layer assembly comprising 3,072 transducers with distributed driver electronics.

As device scale increased, new problems in thermal management, synchronization, and electromagnetic interference emerged. To mitigate these effects, later revisions incorporated increased physical separation between high-voltage components and control electronics, grounded conductive backings to provide shielding, multi-layer PCB layouts with large ground planes and signal wires routed within inner layers, explicit filtering of synchronization and control lines, and cooling from compressed air directed at logic and ultrasonic driver chips which produce the greatest heat.

2.3. Hologram Control Formulation

A vortex-based structure was implemented to generate acoustic pressure fields suitable for constraining fibers near the polymer source. To calculate the phase of each transducer we operate in a cylindrical coordinate system with the origin located at the center of the lowest level of transducer rings and its z axis extending through the center of each higher-level ring. The phase assigned to each transducer ($\phi_{n,L}$) is a function of its angular position (θ_n), layer index (L), and the total number of transducers in each ring (N) parameterized by a topological charge (ℓ) that governs the aperture size as shown in Equation 1. A layer shift value (S) allows for offsetting the hologram by layer.

The topological charge provides a control parameter for tuning confinement diameter and alignment tolerance. Topological charge is typically a whole number with sign indicating the direction of twist. Structures produced from this equation are highly configurable, enabling control of the acoustic field topology and confinement characteristics through manipulation of parameters.

$$\phi_{n,L} = \ell\theta_n + \frac{2\pi}{N} L \cdot S \qquad \textit{Equation 1}$$

$$P(r) = P_0 A \frac{D_f(\theta)}{d} e^{i(\phi_{n,L} + kd)} \qquad \textit{Equation 2}$$

$$k = \frac{2\pi}{\lambda_{n,L}} \qquad \textit{Equation 3}$$

The resulting pressure field developed by the transducer array can be estimated by assuming each transducer acts as a point source at angular frequency ω with wavenumber k. Given each

transducer element (n, l) is located at a position $\mathrm{r}_{n,L}$, has amplitude $a_{n,L}$, and is driven with phase $\phi_{n,L}$, a common far-field model of the complex pressure P at a point r from each transducer is given by Equation 2. The total pressure field is obtained by superposition over all transducer elements. This method is explained well in work by Dr. Asier Marzo and his development of the Ultraino software (Marzo et al., 2017).

- k is the wavenumber defined by Equation 3 and $\lambda_{n,L}$ is the wavelength emitted by each transducer
- d is propagation distance in free space front the transducer $\mathrm{r}_{n,L}$ to the point r
- P_0 is a constant that defines the transducer amplitude power
- A is the peak-to-peak amplitude of the excitation signal.
- D_f is a far-field directivity function that depends on the angle θ between the transducer normal and the point being calculated, r.

$$d = \left|\mathrm{r} - \mathrm{r}_{n,L}\right|$$

Equation 4

$$D_f = sinc\left(ka\ sin\ (\theta)\right)$$

Equation 5

Using these equations, we design and simulate acoustic fields with topological charge (ℓ) of 1, 2, 3, and 4 in Figures 1 – 4 below. These figures show increasing topological charge increases aperture size.



Figure 1. $\ell = 1$



Figure 2. $\ell = 2$



Figure 3. $\ell = 3$

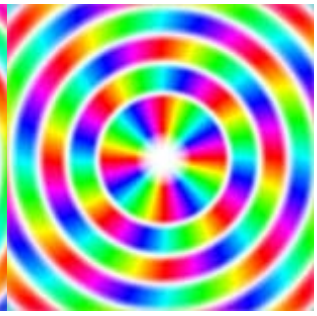


Figure 4. $\ell = 4$

Acoustic pressure fields generate acoustic radiation forces that were estimated using the Ultraino acoustic field simulation software, which implements the Gor'kov potential formulation for spherical particles in the Rayleigh regime. The phased-array geometry and transducer amplitude scaling were defined consistent with the experimental configuration. Force magnitudes were computed from the simulated pressure field. A reference particle radius of 100 nm was selected to approximate the characteristic scale of the electrospun fibers. The resulting force values therefore represent order-of-magnitude radiation-force scales associated with the structured acoustic field under the assumed pressure amplitude and field distribution.

Acoustic radiation forces were simulated for topological charges of 1, 2, 3, and 4 and are presented in Figures 1-4. Because radiation force in the Rayleigh regime scales with particle volume ($\propto a^3$), force magnitudes are reported in volume-normalized form ($\frac{F_{Acoustic}}{Volume}, \frac{N}{m^3}$), allowing the results to be rescaled to any particle radius while avoiding dependence on an arbitrarily selected reference size. This normalization facilitates interpretation for nanoscale fibers, where absolute forces are inherently small, though the spatial force gradients remain physically

meaningful. These simulations indicate that lower topological charge configurations, corresponding to tighter acoustic confinement, produce higher peak size-normalized radiation-force magnitudes within the confinement region as shown in Figure 5.

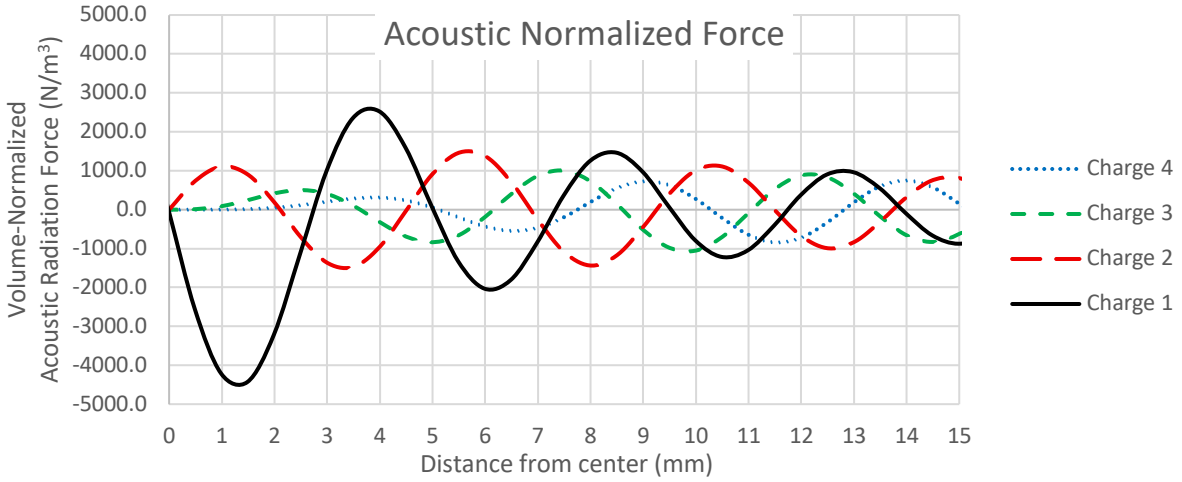


Figure 5. Force Plots

The force distributions are symmetric about the centerline, reflecting the rotational symmetry of the vortex-based acoustic field. Accordingly, the one-dimensional positive x-axis slice shown represents any radial direction in the confinement plane. The sign of the force indicates direction along the radial coordinate; inward-directed forces across the confinement region define a restoring radial force that promotes lateral confinement about the axis of symmetry.

To connect the required phase resolution directly to the previously established $\ell = 1$ confinement aperture (radius ≈ 1 mm ≈ $\frac{\lambda}{8}$), the timing architecture can be interpreted in spatial terms (see Equation 7). The control system executes $40{,}000\text{Hz} \times 96 = 3.84MHz$, yielding a phase resolution of $\frac{360}{96} = 3.75°$. Expressed as an equivalent spatial phase displacement with $\lambda \approx 8.6$mm at 40 kHz, this corresponds to:

$$\Delta x = \Delta\phi \frac{\lambda}{2\pi} \approx 0.065 \frac{8.6\text{ mm}}{2\pi} \approx 0.09\text{ mm}$$

Equation 6

Thus, the spatial phase resolution associated with the update frequency is approximately 90 µm, an order of magnitude smaller than the 1 mm confinement aperture radius observed for $\ell = 1$. The implemented update rate therefore provides phase granularity that is sufficiently fine relative to the physical confinement scale, preserving hologram symmetry and aperture fidelity at 40 kHz.

2.4. Experimental Design and Operating Conditions

A baseline-versus-treatment structure was used to isolate the effect of acoustic actuation. In the baseline condition, electrospinning was performed with the phased array inactive, establishing uncontrolled deposition behavior. In the treatment conditions, the phased array was activated using a selected hologram program and specified transducer drive voltage while maintaining nominally constant electrospinning parameters.

2.5. Evaluation Measures

Evaluation is focused on manufacturing-relevant deposition behavior and its spatial relationship to the acoustic actuation field. Deposition localization is quantified from collector images using radial image analysis centered on a deposition centroid. Centroids are detected from local intensity-weighted deposition scores within the deposited region. RGB images are converted to a whiteness-based deposition score computed from mean pixel brightness. Radial averaging is performed over concentric annuli with a bin width of 0.1 mm to obtain a one-dimensional deposition profile as a function of radius. Pixel intensity is treated as a proxy for relative deposition density under consistent illumination conditions. The annular mean profile is smoothed using a Savitzky-Golay filter. A normalized deposition metric is used where 100% corresponds to the maximum value and 0% corresponds to the minimum value within the analysis domain. To enable quantitative comparison between experimental deposition behavior and acoustic actuation fields, two characteristic metrics are defined: (i) confinement radius r_{90}, and (ii) transition width $\Delta r = r_{10} - r_{90}$. These metrics provide a compact representation of deposition localization and enable direct comparison with simulated acoustic radiation force distributions.

3. Results

Acoustic radiation force distributions were simulated for vortex holograms with topological charge $\ell = 1 - 4$, with forces reported in volume-normalized form (N/m^3) to preserve spatial gradients independent of particle size. Increasing topological charge produces a monotonic increase in confinement aperture and a corresponding reduction in peak force magnitude, consistent with redistribution of acoustic energy across a larger radial domain.

Experimental evaluation was conducted using the $\ell = 4$ configuration due to its expanded aperture and multi-band force structure. Deposition behavior was quantified using radial image analysis centered on the measured deposition centroid, yielding normalized deposition intensity profiles as a function of the radius. From these profiles, characteristic confinement metrics were extracted. The confinement radius was measured as $r_{90} = 12.8\ mm$, indicating the radial extent of the high-density deposition region. The measured confinement radius lies 1.25 mm from the nearest peak inward-directed acoustic radiation force associated with the second confinement band, which corresponds to approximately 10% of the 12.1 mm spacing between adjacent inward-directed radial force bands in the $\ell = 4$ field.

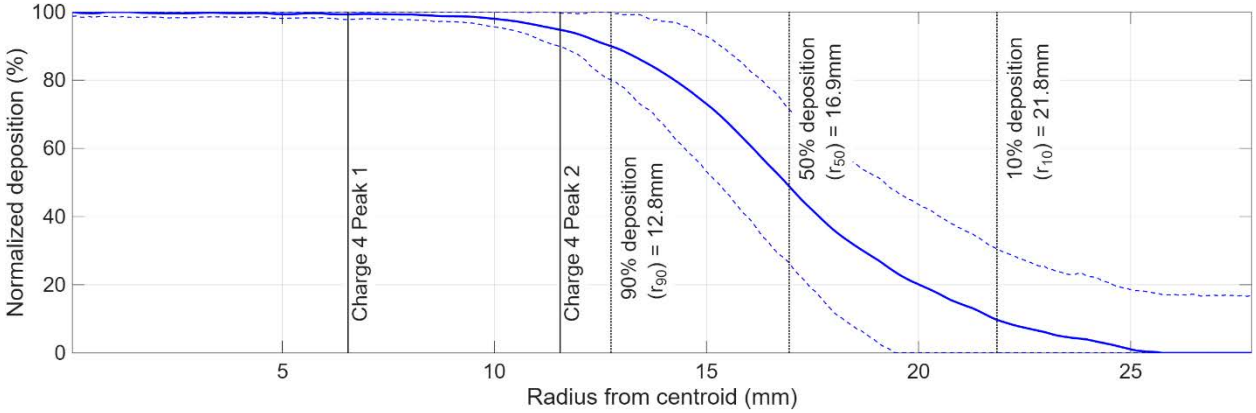


Figure 6. Deposition % VS. Distance from Deposition Centroid

The transition width between deposition and background is quantified as $\Delta r = r_{10} - r_{90} = 9.1\ mm$, demonstrating a gradual but structured decay in deposition density (Figure 6). The deposition profile exhibits a central region of approximately uniform intensity followed by a smooth, sigmoidal-like decay, consistent with the spatial structure of the simulated acoustic radiation force field. The onset of deposition decay occurs within a radial region corresponding to the second inward-directed force band predicted by the acoustic model.

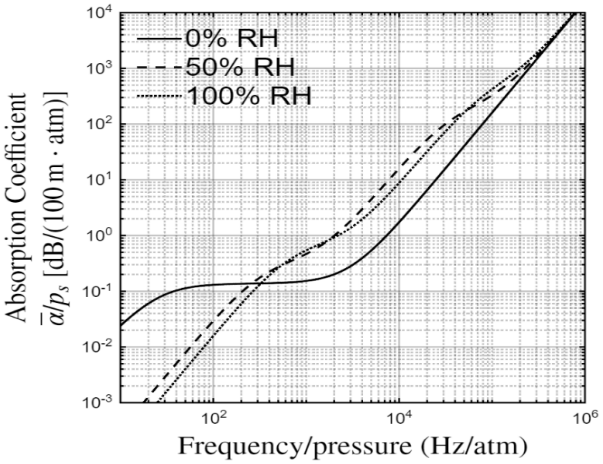


Figure 7. Atmospheric Absorption in Air

Comparison between experimental deposition and simulated force distributions shows spatial alignment between the measured confinement boundary and regions of elevated inward-directed acoustic radiation force. This correspondence indicates that acoustic pressure gradients impose a quantifiable spatial constraint on fiber trajectory as reflected in the measured confinement radius and transition width, even in the absence of precise jet positioning within the vortex center.

4. Discussion

The principal contribution of acoustic hologram guided electrospinning is the introduction of a software-defined, non-contact control layer that is not inherently coupled to electrostatic field shaping. This separation supports modularity: deposition targeting can be adjusted without reconfiguring the electrostatic environment, and the guidance layer can be repurposed for routing fibers through constrained spatial regions. Such routing is a prerequisite for apertured ground plate designs intended to enable continuous downstream processing.

Scalability depends on both physical array geometry and acoustic wavelength. The wavelength defines the scale of achievable confinement dimensions obtainable without transitioning to higher-frequency operation. The present 40 kHz implementation produces sub-centimeter-scale spatial features. Higher-frequency arrays reduce acoustic wavelength, permiting tighter confinement and finer deposition control. Acoustic attenuation in air increases with the frequency as shown in Figure 7, leading to higher losses and reduced pressure amplitude (Bass et al., 1995; ISO, 1993; Nakamura, 2012). Attenuation at Megahertz frequencies is significant, constraining effective working distance and increasing required transducer power to maintain comparable radiation forces.

Higher-frequency operation imposes stricter mechanical and electronic tolerances. As wavelength decreases, phase errors arising from timing, geometric misalignment, and fabrication tolerances represent a larger fraction of the wavelength, increasing sensitivity to synchronization integrity and structural precision. Acoustic wavelength varies with temperature through changes in sound speed, environmental conditions can introduce additional variability in confinement scale and phase alignment, emphasizing the importance of robust synchronization and structural control at elevated frequencies.

5. Future Work

System-level upgrades are required to address current limitations in source positioning and experimental repeatability, which constrain interpretation of acoustic confinement behavior. These

upgrades will enable precise, repeatable positioning of the polymer output and automated test execution. This includes development of a fixture capable of controlling axial height, radial offset, angular position, and tip alignment while maintaining electrical safety and mechanical rigidity, as well as integration of automated workflows for repeatable positioning, scripted parameter variation, timed run sequences, and synchronized data logging. Collectively, these and other improvements reduce operator-induced variability and establish a controlled experimental framework.

With these capabilities established, we will operate on a topological charge on 1 which offers a more focused aperture and higher acoustic forces. A structured experimental program will be conducted to systematically vary key control parameters (e.g., transducer drive voltage, acoustic field pattern, and polymer flow rate) and quantify their influence on deposition geometry. Data will enable creation of a process map linking control inputs to deposition behavior, identify stable operating regions, characterize sensitivity gradients, and determine dominant control variables. This framework enables evaluation of acoustic fields as a control mechanism for fiber placement.

6. Conclusions

This work demonstrates that ultrasonic phased-array acoustic holograms can be integrated into an electrospinning platform to introduce a software-defined, non-contact control layer for regulation of nanofiber deposition. By decoupling fiber formation from guidance, the system positions electrospinning from a largely stochastic deposition process to a controllable manufacturing operation. Baseline-versus-treatment evaluation confirms that acoustic activation localizes deposition within targeted regions and reduces unintended accumulation on surrounding structures, thereby establishing functional control authority beyond conventional electrostatic methods. The demonstrated platform represents a structural re-architecture of the process, enabling modular deposition windows, multi-jet guidance strategies, and integration with downstream processing such as routing, bundling, or consolidation. Collectively, these results establish acoustic hologram guidance as a potential field-based manufacturing control capability and define a pathway toward nanofiber production by controlled alignment, selective deposition, and continuous processing architectures.

7. Disclaimer Statements

Funding: None

Conflict of Interest: None

8. References

Almodovar, J. (2018). Electrospun Biomaterials and Related Technologies. Springer. https://doi.org/10.1007/978-3-319-70049-6

Asmatulu, R., Veisi, Z., Uddin, M. N., & Mahapatro, A. (2019). Highly sensitive and reliable electrospun polyaniline nanofiber based biosensor as a robust platform for COX-2 enzyme detections. Fibers and Polymers, 20, 966-974. https://doi.org/10.1007/s12221-019-1096-x

Bass, H. E., Sutherland, L. C., Zuckerwar, A. J., Blackstock, D. T., & Hester, D. (1995). Atmospheric absorption of sound: Further developments. https://doi.org/10.1121/1.412989

Bellan, L. M., & Craighead, H. G. (2006). Control of an electrospinning jet using electric focusing and jet-steering fields. Journal of Vacuum Science & Technology B: Microelectronics and Nanometer Structures Processing, Measurement, and Phenomena, 24(6), 3179-3183. https://doi.org/10.1116/1.2363403

Gor'kov, L. P. (1962). On the forces acting on a small particle in an acoustical field in an ideal fluid. Sov. Phys.-Doklady. https://doi.org/10.1142/9789814366960_0008

Gu, B. K., Shin, M. K., Sohn, K. W., Kim, S. I., Kim, S. J., Kim, S.-K., Lee, H., & Park, J. S. (2007). Direct

fabrication of twisted nanofibers by electrospinning. Applied Physics Letters, 90(26). https://doi.org/10.1063/1.2753109

Huang, Z.-M., Zhang, Y.-Z., Kotaki, M., & Ramakrishna, S. (2003). A review on polymer nanofibers by electrospinning and their applications in nanocomposites. Composites science and technology, 63(15), 2223-2253. https://doi.org/10.1016/S0266-3538(03)00178-7

ISO. (1993). Acoustics — Attenuation of sound during propagation outdoors. In Part 1: Calculation of the absorption of sound by the atmosphere (Vol. ISO 9613-1:1993, pp. 26). Geneva: ISO.

Johns, A., Qian, J., Carolan, M. E., Shaikh, N., Peroutka, A., Seeger, A., Cerrato, J. M., Forbes, T. Z., & Cwiertny, D. M. (2020). Functionalized electrospun polymer nanofibers for treatment of water contaminated with uranium. Environmental science: water research & technology, 6(3), 622-634. https://doi.org/10.1039/C9EW00834A

Kim, G., Cho, Y.-S., & Kim, W. D. (2006). Stability analysis for multi-jets electrospinning process modified with a cylindrical electrode. European Polymer Journal, 42(9), 2031-2038. https://doi.org/10.1016/j.eurpolymj.2006.01.026

Marzo, A., Corkett, T., & Drinkwater, B. W. (2017). Ultraino: An open phased-array system for narrowband airborne ultrasound transmission. IEEE transactions on ultrasonics, ferroelectrics, and frequency control, 65(1), 102-111. https://doi.org/10.1109/TUFFC.2017.2769399

Marzo, A., Seah, S. A., Drinkwater, B. W., Sahoo, D. R., Long, B., & Subramanian, S. (2015). Holographic acoustic elements for manipulation of levitated objects. Nature Communications, 6(1), 8661. https://doi.org/10.1038/ncomms9661

Nakamura, K. (2012). Ultrasonic transducers: Materials and design for sensors, actuators and medical applications. Elsevier.

Ochiai, Y., Hoshi, T., & Rekimoto, J. (2014). Pixie dust: Graphics generated by levitated and animated objects in computational acoustic-potential field. ACM Transactions on Graphics (TOG), 33(4), 1-13. https://doi.org/10.1145/2601097.2601118

Yang, Y., Jia, Z., Li, Q., Hou, L., Liu, J., Wang, L., Guan, Z., & Zahn, M. (2010). A shield ring enhanced equilateral hexagon distributed multi-needle electrospinning spinneret. IEEE Transactions on Dielectrics and Electrical Insulation, 17(5), 1592-1601. https://doi.org/10.1109/TDEI.2010.5595562

Zhao, J., Liu, H., & Xu, L. (2016). Preparation and formation mechanism of highly aligned electrospun nanofibers using a modified parallel electrode method. Materials & Design, 90, 1-6. https://doi.org/10.1016/j.matdes.2015.10.116

Zhou, Y., Liu, Y., Zhang, M., Feng, Z., Yu, D.-G., & Wang, K. (2022). Electrospun nanofiber membranes for air filtration: A review. Nanomaterials, 12(7), 1077. https://doi.org/10.3390/nano12071077

Zhou, Y., & Wang, M. (2021). Emulsion electrospun multicomponent scaffolds for tissue engineering. In Electrospun Polymers and Composites (pp. 45-75). Elsevier. https://doi.org/10.1016/B978-0-12-819611-3.00002-9

The Journal of Management and Engineering Integration (JMEI) © 2026 by AIEMS Inc., is licensed under CC BY 4.0. To view a copy of this license, visit https://creativecommons.org/licenses/by/4.0/

The Journal of Management and Engineering Integration Vol. 19, No. 2 | Winter 2026 DOI: 10.62704/10057/59053

SPEAKING UP: AERONAUTIC STUDENT VOICE PROPENSITY AND POWER DISTANCE BY EXPERIENCE

Nicholas Forthofer [1]
Shellbie Reed [1]
Mihir Wadhwa [1]
Vivek Sharma [1]
Brooke E. Wheeler [1]
[1] Florida Institute of Technology
nforthofer2022@my.fit.edu; bwheeler@fit.edu

Abstract

This study examined the difference in aeronautic student voice propensity (confidence in raising concerns) and perceived power distance (level of instructor authority) as it relates to flight training and level of experience. Voice propensity and power distance were used as the dependent measures. Across 317 students, ages 18-25, from a Part 141 collegiate aviation program, voice propensity and power distance were evaluated in four self-reported experience categories: students without a student pilot certificate, Student Pilots, Private Pilots, and Private Pilots with an instrument rating or above. Both power distance and voice propensity showed statistically significant differences between groups and increased with experience. The greatest difference between all groups was with the instrument rated pilots. There was a small effect size for all differences, but the results of this study suggest that there is a greater willingness to express dissent to instructors as experience is gained in flight training, and there is a higher perceived power distance, which could reflect a growing respect for aviation expertise.

Keywords: *Voice Propensity, Power Distance, Dissent, Student Pilot, Flight Instructor, Student-Instructor Relationship, Culture*

1. Introduction

Communication is an essential priority in aviation. Pilots learn how to communicate with their peers and superiors beginning on day one of their flight training. The quality of training is directly linked to how it is delivered by the instructor and received by the student, and therefore ultimately the quality of pilot can be impacted because of their foundational knowledge (Coehlo, 1978). Discovering how pilots perceive their instructor and how they choose to communicate with them could be a way to improve training and flight safety overall.

The purpose of this study is to determine if there is a difference in flight students' voice propensity (Kassing, 2000) and perception of power distance (Yoo & Donthu, 2002) when speaking to Certified Flight Instructors (CFIs) by aviation experience. Experience levels were self-reported and included four levels: students who have started flight training but do not have a student pilot certificate, students training with a student pilot certificate, Private Pilots, and Private Pilots with an instrument rating and above. Thus, the research questions examined were the following: a) Will

Submitted: April 29, 2026
Revised: October 19, 2026

there be a difference in student voice propensity (Kassing, 2000) by experience level? and b) Will there be a difference in perceived power distance (Yoo and Donthu, 2002) between experience levels? We expected that a) voice propensity would increase with increasing levels of flight experience, and b) perception of power distance would decrease with increasing levels of flight experience.

The study provided data on collegiate Part 141 flight students' voice propensity (Kassing, 2000) and perception of power distance (Yoo & Donthu, 2002) between themselves and their CFIs. This data aids in understanding the willingness of student pilots and private pilots to speak up when communicating with their flight instructors at Part 141 flight schools. To date, this has not been examined in the flight student population, despite the direct importance to flight safety of feeling comfortable speaking up. Furthermore, this aids in determining if pilots are more willing to communicate with their instructor as experience is gained. This could be used to improve training by making instructors aware that the atmosphere they create in the cockpit can limit the development of students' communication skills by decreasing their willingness to speak up. Moreover, if this issue exists in training, the results can be applied to the issue of first officers being hesitant to speak up to their captains or other roles in aviation with a high-power distance relationship.

2. Literature Review

Communication between crew members is an essential facet of flight deck operations. One necessary part of communication is the ability of junior crew members to speak up to their superiors in the event of an error. The relationship between the superior and subordinate crew members in respect to the willingness of junior crewmembers to speak up and how power is distributed represents an important influence on communications during flight operations. Specifically, voice propensity (Kassing, 2000) and perception of power distance (Yoo & Donthu, 2002) were analyzed with additional context of previous studies across different populations and the student and instructor relationship.

Voice propensity is described as "articulated dissent" - the willingness of a member of an organization to express their disagreement with superiors that can influence the organization (Kassing, 2000). This superior-subordinate relationship serves as the main method in which crew members express dissent. In aviation, this corresponds to the junior and senior crew members in the cockpit. To express dissent, a crewmember must have an idea to put forward (Frese et al., 1999). Secondly, they must believe that their suggestion will be taken seriously (Parker, 1993). The National Transportation Safety Board ([NTSB], 1991) found that of 37 accidents, 31 of them involved the failure of a crewmember to countermand another crewmember's mistake. The NTSB did not conclude that the errors took place because of the other crewmember's failures to challenge their captains but nevertheless emphasized the number of unchallenged errors. Of the 37 accidents, 80 percent occurred while the captain was flying, and the first officer was monitoring. The NTSB attributed the lack of corrective action to the first officer's perception that their correction may be taken as a challenge by the captain, thus inhibiting the first officer's willingness to express dissent. In response, the NTSB recommended that crew resource management (CRM) be required for airline training and is currently a required item in the Private Pilot Airman Certification Standards (Federal Aviation Administration [FAA], 2023; NTSB, 1991). This trend towards the implementation of CRM is substantiated by Kassing's (2000) finding that a higher quality superior-subordinate relationship correlated with an increased willingness to express dissent. The implementation of CRM training in curriculums ranging from private pilot training to air transportation was designed to improve the relationship between captain and first officers. In doing so, it sought to remove two of the largest

barriers to voice propensity: the first officer's lack of willingness to speak up and the captain's attitude to accept what the first officer had to say.

The FAA (1991) implemented CRM training because of the lack of voice propensity among flight crews; however, in the report, no definite measure of voice propensity took place. Specific measures between experience and voice propensity were assessed across varying populations, indicating that a relationship does exist. Parker (1993) concluded that among nurses at university hospitals, participants, who had spent more years in nursing, displayed more willingness to express dissent. However, only females made up the sample population as the five male participants' responses were discarded. Additionally, the sample population was from a medical environment operating under different circumstances compared to an aviation environment. While the sample was skewed towards female participants in a medical profession, it demonstrates that a positive relationship can exist between experience and voice propensity in hierarchical, high-risk professions. Landau (2009) found a correlation between participants' voice propensity and their belief in their own competence. The sample was not from a population consisting of pilots, but rather a population of collegiate business students and their friends. However, the scales were validated by the resulting correlation. In an aviation environment, Kirschner et al. (2014) identified that senior level students reported lower levels of stress, which correlated with their tendency to seek instrumental support compared to their freshmen level colleagues. Lower stress levels among upperclassmen were attributed to their ability to reach out to others in abnormal situations, likely due to more experience with CRM situations. While Kirschner et al. focused on stress management and experience, the results were correlational and could not identify a direct causal relationship. Additional factors included the participants' individual abilities to handle stress and how some students with higher stress levels could have opted out of the flight program. Furthermore, it cannot be assumed that students improved their ability to cope by speaking up as their experience increased. Nevertheless, Kirschner's research identified that experience related to the student's increased willingness to seek outside help with the inclusion that CRM training took place. In doing so, they demonstrated that proper training and experience could relate to an individual's willingness to speak up to others.

The second factor relating to voice propensity — the perception of the utility of their dissent, including the influence of power distance — was assessed in the following investigations and found to be prevalent across different populations. Parker (1993) concluded that among female nurses, participants would be more willing to express dissent if participants believed that their input could influence decision-making. In Parker's study, the sample population included only nurses, not pilots. However, among business students and their friends, Landau (2000) found that there was a negative relationship between voice propensity and high power distance settings. While both studies were conducted in non-aviation populations, similar relationships between perceived utility, power distance, and voice propensity were found. Therefore, these similar relationships demonstrate that across different populations, individuals' willingness to speak up could be related to their perceptions of utility and power distance.

Power distance is defined as the extent to which less powerful people "expect and accept that power is distributed unequally" (Yoo & Donthu, 2002, p. 95). A large power distance indicates that the individual is more likely to obey their superiors and follow more formal norms, rather than those of their peers and informal norms (Yoo & Donthu, 2002). Coehlo (1978) talked about instructors and how they recognize this power distance relationship, and how students react to this. These studies relate to aviation in the sense that in the flight deck, different crew members have different levels of power and seniority in each situation. The CFI-student power distance relationship is important in the study; it is crucial that the student and instructor have a proper relationship, with a low

perception of power distance.

Overall, power distance is recognized in the studies described here and should be taken into consideration when discussing students speaking up to instructors, especially with the large power distance gap in between students and CFIs.

The student pilot-instructor relationship is widely studied and is taught in ground school when pilots are training to become a CFI. The Aviation Instructor's Handbook is a key source of information from the FAA (2020) that outlines the way an instructor should teach and interact with their flight students. It was updated in 2020, but aviation is continuing to change every single year, and the book is surface-level on certain subjects, especially those that detail the student-instructor relationship. In addition, there have been some recent studies that discuss the impact of how a flight instructor interacts with their student. Baughman et al. (2025) found a statistically significant positive relationship between student-instructor rapport and the impact that their training had. Their study used two scales that have not been frequently cited; however, because they used two scales, this data can be considered more reliable (Baughman et al., 2025). Another similar study also found that the relationship between expected grades was lower, and there was a higher expected rate of classes missed when student-instructor rapport was low (Demir et al., 2019). And even though their findings are quite dated in comparison to these studies, Coehlo (1978) even explains that students and instructors should feel comfortable around each other to maintain open communication and have a higher level of efficiency in training.

Combining the ideas of voice propensity and power distance gives insight into the student-instructor relationship in aviation. Effective communication is essential to aviation training, and a good relationship between student and instructor will improve the experience for both parties (Coehlo, 1978). Safety of flight can also be negatively impacted if students have a negative perception of their relationship with their instructor. This could mean a high perceived power distance combined with a low voice propensity due to fear; this can be a dangerous combination. It is the instructor's job to decrease fear in the student, not cause so much distress that they find it nearly impossible to speak up. An approachable and responsive relationship that is formed between each party is ideal for flight training to produce the safest and most confident pilots.

There are also cultural considerations to consider because people of various cultural backgrounds will have different perceptions of power distance (Assis & Pacheco, 2020). Hofstede's (2001) Culture's Consequences has been studied by various researchers, and multiple replications of his original study have shown that there is a difference in power distance between cultures, especially American and Asian countries (Søndergaard, 1994). Cultural differences are prevalent in the aviation industry as there are many international operations and interactions between pilots. The way that pilots express their needs, wishes, and preferences in the flight deck may be vastly different and based on wording, and tone could be misconstrued, which will also change the effectiveness of communicating solutions and implementing them (Assis & Pacheco, 2020). If students learn about power distance early in their training, it could help prevent future events.

The way that pilots communicate and make requests is also related to the response they receive. Clark (1979) discussed the difference between responses when various questions are asked in his experiments on indirect speech acts. All the questions in his trials regarding the underlying meaning of speech would eventually lead to the same information being provided, but with varying levels of information attached (Clark, 1979). This should be considered as well because different levels of voice propensity and perceived power distance can alter the way a pilot will make a request, and therefore, the response they receive from their superior.

How a junior member of any organization perceives their ability to express dissent can be influenced by what they say and if their dissent will be perceived as being useful (Frese et al., 1999; Parker, 1993). High quality relationships between superiors and subordinates across different fields including aviation are essential to ensuring that effective communication takes place (Kassing, 2000). Overall, a high voice propensity and low perceived power distance is the most optimal combination to facilitate communication, safety, efficiency, confidence, and quality airmanship through a student's flight training. Several studies have been conducted to investigate voice propensity and perceptions of power distance in superior-subordinate relationships taking place in business and medical fields. Comparatively, voice propensity and power distance has rarely been assessed in flight training where the superior-subordinate relationship between the instructor and student is essential to the students' learning. Culture is also a key consideration and can have an effect outside of experience level. This information combined led to curiosity about voice propensity and power distance as they relate to flight training.

3. Methods

This research included flight students from a collegiate Part 141 flight program in central Florida. They were sampled from core aeronautics classes using stratified cluster sampling. Each participant was at least 18 years old. The target population was flight students between 18 and 25 years old currently enrolled in the college. The accessible population was approximately 550 students registered in the core courses during Fall 2025. Both male and female students with varying levels of experience were sampled. There were more male than female students, but the sample should still be representative of the aviation community because, as of 2020, only 14 percent of Student Pilots and seven percent of Private Pilots were female (Federal Aviation Administration Women in Aviation Advisory Board, 2020).

The research protocol was reviewed by the IRB (IRB-CY2025-65) and an exemption was granted on 10 October 2025. Participation was expected to pose no greater risk than everyday activities. Participation was voluntary, with the option to opt out at any point, and all data were confidential. Informed consent was obtained prior to the start of the questionnaire.

Voice propensity was measured, using Kassing's (2000) Articulated Dissent Likert-type scale. Responses ranged from 1 (strongly agree) to 5 (strongly disagree). The scale included six statements used by Landau (2009) to measure voice propensity. The questions were revised from the original 20-item organizational dissent scale (Kassing, 2000), with the shorter version excluding six items that measured displaced dissent and were removed to yield a scale that measured the variable more effectively (Kassing, 1998; Kassing & Avtgis, 1999). Landau's scale emphasized articulated dissent, and therefore did not include other items measuring latent dissent. A Cronbach's alpha value of .70 was reported (Landau, 2009), which indicates that the scale has fair internal validity.

Perceptions of power distance were measured using Yoo and Donthu's (2002) items to measure power distance. Responses ranged from 1 (strongly agree) to 5 (strongly disagree). The scale included three items from Yoo and Donthu's (2002) original 26 item Likert-type (1932) scale to measure Hofstede's (1980, 1991) dimensions at the individual level. The scale showed high internal reliability with a Cronbach's alpha value of .86 (Yoo and Donthu, 2002). The questionnaire, with both scales, items to distinguish experience level, and demographic questions, was disseminated to participants via Qualtrics.

An ex-post facto design was used for this study. Professors for aeronautics 1-4 and instructional techniques were emailed to request permission to speak to their students about participating. Participants were recruited in person, during class visits, and a link to the questionnaire was posted

in the class learning management system. Responses were downloaded from Qualtrics, and the deidentified data file was saved on a password-protected computer. Excel was used to correct the reverse-coded values for each respondent prior to analysis. The mean was then calculated from each participant's responses to all six items of the voice propensity index (Kassing, 2000) and all three items for power distance (Yoo & Donthu, 2002) to yield each participants' power distance and voice propensity scores, respectively.

Descriptive statistics were calculated for voice propensity and power distance, both overall, and by each of the four groups: student training without student pilot certificate, Student Pilot, Private Pilot, and private pilot certificate with instrument rating. A Levene's test for equal variance and Shapiro-Wilks (1965) test for normality were run to check the assumptions for an ANOVA. The ANOVA was calculated using RStudio (Posit Team, 2025) to measure voice propensity and power distance. Tukey's Honesty Significant Differences (HSD) test was used to determine the difference between groups.

4. Results

Ten aeronautics classes were attended across Aeronautics 1 through Aeronautics 4 and Instructional Techniques. A total of 388 people were sent the survey and asked to participate. The questionnaire received a total of 349 responses. There was a total response rate of 90%, and four responses were excluded from the results for mortality. Twenty-seven responses were excluded because responses to experience level items were not consistent, and thus, these responses could not confidently be categorized into an experience level. The remaining responses were Z-scored, and three outliers were identified. One Private Pilot's response was excluded from voice propensity, and two Student Pilots' responses were excluded from power distance. The sample size for analysis was 317 responses for voice propensity and 316 for power distance. Cronbach's alpha for voice propensity ($\alpha = .75$) and power distance ($\alpha = .65$) indicated fair internal reliability.

Descriptive statistics were calculated for both voice propensity (Table 1) and power distance (Table 2) overall and across the flight student experience levels. The mean for voice propensity across respondents indicated a neutral willingness to speak up ($M = 3.10$, $SD = 0.42$). The mean for power distance across respondents indicated a low power distance ($M = 1.53$, $SD = 0.56$), and the range was on the lower end of the five point scale: 1 to 3. The mean voice propensity for each certificate varied slightly but generally increased with experience (see Figure 1, Table 1). The mean power distance for flight students increased from students without their student pilot certificate to those with an instrument rating or higher, with the exception of a slight decrease in the student pilot category (see Figure 2, Table 2).

Table 1. Descriptive Statistics for Voice Propensity by Flight Student Experience Level

Experience level	*N*	Mean	Minimum	Maximum	Standard Deviation
Student	34	2.98	2.33	3.83	0.34
Student Pilot	108	3.01	2.00	4.00	0.40
Private Pilot	117	3.07	2.17	4.00	0.41
Instrument Rated	58	3.25	2.17	4.33	0.46
Voice Propensity	**317**	**3.10**	**2.00**	**4.33**	**0.42**

Table 2. Descriptive Statistics for Power Distance by Flight Student Experience Level

Experience level	*N*	Mean	Minimum	Maximum	Standard Deviation
Student	34	1.47	1.00	3.00	0.52
Student Pilot	106	1.44	1.00	3.00	0.52
Private Pilot	118	1.56	1.00	3.00	0.59
Instrument Rated	58	1.70	1.00	3.00	0.57
Power Distance	**316**	**1.53**	**1.00**	**3.00**	**0.56**

To meet the assumption of equal variance for the ANOVA test, Levene's test for homogeneity of variance was conducted based on certificate level. For voice propensity, the variances were equal between groups ($p = .17$). Power distance also had equal variances by group ($p = .86$). Accordingly, the null hypothesis that the variances are equal across groups was accepted for both variables. The Shapiro-Wilk test for normality showed that both voice propensity ($p = .002$) and power distance ($p < .001$) did not meet the normality assumption. While the distribution was not normal, under the pretext of the central limit theorem and given that the population for each group ($n > 30$) is large enough to infer that a normal distribution exists relative to the population size (Gallo et. al, 2023).

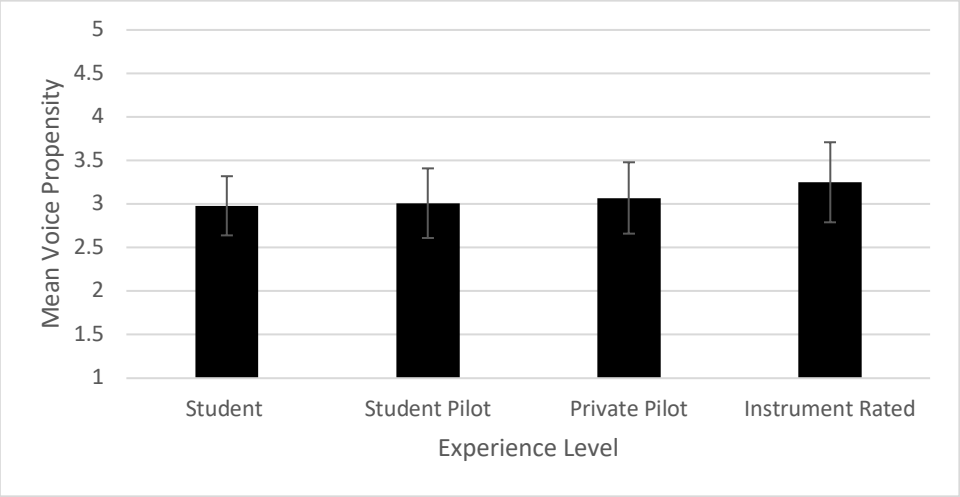


Figure 1. Voice Propensity by Flight Student Experience Level

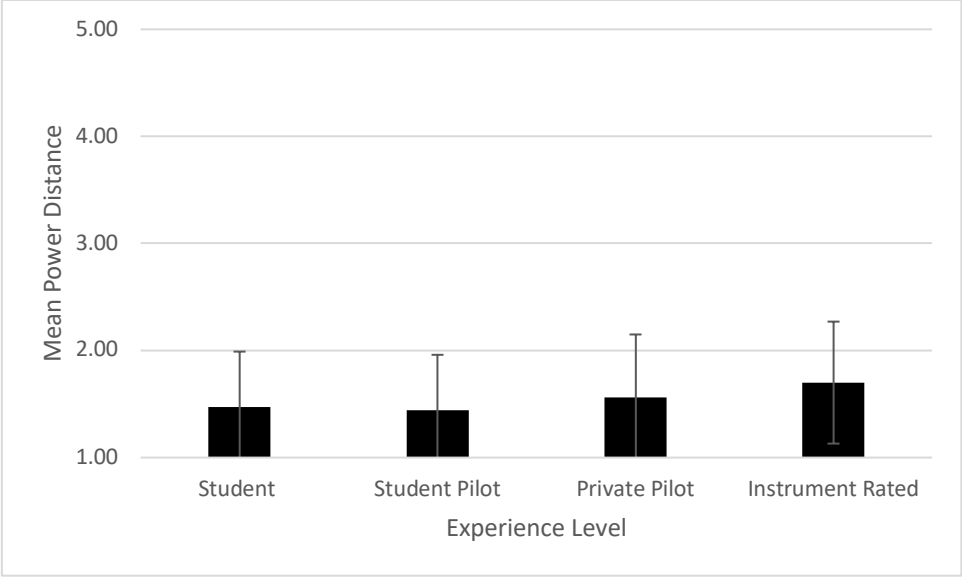


Figure 2. Power Distance by Flight Student Experience Level

The ANOVA indicated a statistically significant difference in voice propensity by experience (F(3, 313) = 3.79, p = .01). Tukey's HSD showed that significant differences existed between instrument rated private pilots and non-student pilots (p = .01, 95% CI = [-.5,-.04]), and instrument rated pilots and private pilots (p = .02, 95% CI = [-.35, -.01]). The ANOVA for power distance also found a significant difference by experience (F(3, 312) = 3.15, p = .03). Tukey's HSD showed that there was a significant difference between instrument rated private pilots compared to student pilots (p = .02, 95% CI = [-.05,-.03]). Both voice propensity (η^2 =.04) and power distance (η^2 = .03) had a small effect size.

5. Discussion

Statistically significant differences in voice propensity were found between non-student pilots and instrument pilots, as well as private pilots and instrument-rated pilots. Additionally, a slight increasing trend was noted for students' average voice propensity with experience (Figure 1); however, only the differences between instrument rated pilots and both private and non-student pilots were statistically significant. Overall, no group strongly agreed that they were willing to speak up, but the trend was for voice propensity to gradually increase with flight experience (Table 1, Figure 1). However, the magnitude of the difference in voice propensity between the instrument rated pilots and both private and student pilots was small, indicating that the increase in speaking up is very gradual. Thus, the hypothesis that more experienced pilots would have a higher voice propensity was supported.

These findings aligned with prior research showing that as students gain more experience, they will have more confidence to express their dissent (Frese et. al, 1999). The instrument rated private pilot group's relatively high voice propensity score and accompanying difference compared to non-student pilots and private pilots (Table 1, Figure 1) coincided with earlier findings that senior flight students demonstrated a higher tendency to speak up to get help (Kirschner et. al, 2014), although that study did not specifically examine dissent. Moreover, the general increase in voice propensity with experience level could support the effectiveness of the implementation of CRM training, as

students indicated that they were more likely to express their dissent to their instructors as they progressed through their training. This increase supported prior findings that more experienced students whose training included CRM training had a higher tendency to seek outside help (Kirschner. et al, 2014). Given the importance of avoiding errors that may impact safety, training students to speak up when they notice problems is crucial.

A significant difference in power distance was found between instrument-rated pilots and student pilots. While the difference was significant, the magnitude of the difference was relatively weak, but still present. However, contrary to the hypothesis, perceptions of power distance increased with experience (Figure 2, Table 2). All power distances scores were in the lower half of the scale, ranging from 1-3, likely due to the individualistic culture in the US where the flight school is located. Additionally, the sample was from a Part 141 training environment, which tends to create a highly structured atmosphere, promoting a more inflexible relationship between the student and instructor. If instructors are encouraged to maintain a high level of professionalism without any encouragement to remain amiable towards their students, the training environment may become rigid and harsh (Cohelo, 1978). Students may also gain a sense of higher power distance with increasing flight experience because they begin to recognize the amount of work that goes into flight training after they become a Private Pilot and the role of pilot in command. Student Pilots may be more prone to having a larger ego when they start flight training, with a lower understanding of exactly how much work their instructor did to become a CFI.

There could also be cultural factors that are isolated to the flight school or even the specific sample population the data were collected from. The older students might have started training in a slightly different culture than their peers with less experience. The instructors at the flight line usually change every couple of years, so the culture and knowledge of how to train students might be different for those pilots that have a higher level of experience. While the increase in power distance with experience did not support the hypothesis, it is not inherently detrimental. An increase in power distance with an increase in flight experience is conducive for maintaining respect for the instructor and supporting the transfer of learning.

While the increase in voice propensity and power distance across experience levels and identified significant differences, it cannot be concluded that the results are causal, because an experimental design was not used and other potential causal factors were not eliminated. Variations in the individual flight instructor's personality could not be accounted for as instructors from different backgrounds can build different relationships with their students. Thus, the manner in which the subordinate believes that their dissent is accepted by their superior can affect their willingness to speak up (Parker, 1993). While the findings could be generalizable to other collegiate flight programs, it would not be appropriate to attempt to generalize these findings to the general pilot population as the responses were sampled from within a collegiate Part 141 aviation environment in central Florida.

The findings are summarized in Table 3. Although the effect sizes for both voice propensity and power distance were small, there was a statistically significant difference in both variables with experience. The higher voice propensity of the instrument rated pilots underscores that in this population of pilots; they are learning to speak up. This is recommended as an emphasis in training given the importance of communication, particularly in safety-critical situations. There was also a significant difference in power distance with instrument rated pilots perceiving a greater power distance.

Table 3: Summary of Key Findings

Item	Finding
Voice Propensity	Significant difference with experience; instrument rated pilots had higher voice propensity
Power Distance	Significant difference with experience; instrument rated pilots had higher power distance than student pilots
Effect Sizes	Small for both voice propensity and power distance

6. Future Research

Further research could be conducted, especially to take into consideration the cultural factors, which were not a part of the focus of this study. Only the difference between student pilots, Private Pilots and Private Pilots that were instrument rated above were studied. There were not enough respondents who were Commercial Pilots or CFI applicants to create separate categories, so having a bigger population could grant more insight to more experienced pilots. Future research that investigates correlations between student pilot experience and voice propensity and power distance, would provide insight into the actual relationship of these variables. Conducting the study in a different operating environment may provide a means to either contradict or verify the reliability of this study depending one of similar results can be found in other flight schools.

7. Disclaimer Statements

Funding: None

Conflict of Interest: None

8. References

Assis, L. A., & Pacheco A. (2020). Communication in a multicultural flight deck environment: The influence of culture dimensions in communicative functions. *Revista CBTecLE, 4*(1), 441-462. https://revista-cbtecle.cps.sp.gov.br/index.php/CBTecLE/article/view/258

Baughman, R., Siao, D., & Birdsong, J. (2025). Instructor-Student Rapport and the Relationship to Undergraduate Aviation Students' Perception of Impact. *Proceedings of the 23rd International Symposium on Aviation Psychology, Oregon State University, Corvallis, Oregon, United States*. 216-221. https://corescholar.libraries.wright.edu/isap_2025/37

Clark, H. H. (1979). Responding to indirect speech acts. *Cognitive Psychology, 11*(4), 430–477. https://doi.org/10.1016/0010-0285(79)90020-3

Coehlo, B. A. (1978). Aeromedical aspects of flying training with special reference to instructor pupil relationship. *Aviation Medicine, 22*, 48-51.

Cohen, J. (2013). *Statistical Power Analysis for the Behavioral Sciences.* Second edition. Routledge. https://doi.org/10.4324/9780203771587

Demir, M., Burton, S., & Dunbar, N. (2019). Professor-student rapport and perceived autonomy support as predictors of course and student outcomes. *Teaching of Psychology, 46*(1), 22–33. https://doi.org/10.1177/0098628318816132

Federal Aviation Administration. (2023). *Private Pilot Airman Certification Standards*. Federal Aviation Administration. https://www.faa.gov/training_testing/testing/acs/private_airplane_acs_6.pdf

Federal Aviation Administration. (2020). *Aviation Instructor's Handbook* (FAA-H-8083-9B). U.S. Department of Transportation. https://www.faa.gov/regulations_policies/handbooks_manuals/aviation/aviation_instructors_handb

ook
Federal Aviation Administration Women in Aviation Advisory Board. (2022). *Breaking Barriers for women in aviation: Flight plan for the future.* United States Department of Transportation. https://www.faa.gov/regulations_policies/rulemaking/committees/documents/media/WIA_AB_Recommendations_Report_March_2022.pdf
Frese, M., E. Teng & C. Wijnen (1999). Helping to improve suggestion systems: Predictors of making suggestions in companies. *Journal of Organizational Behavior, 20*(7), 1139-1150. https://doi.org/10.1002/(SICI)1099-1379(199912)20:7%3C1139::AID-JOB946%3E3.0.CO;2-I
Gallo, M., Silver, I., Wheeler, B. (2023). *Fundamentals of Statistics for Aviation Research*. Routledge. https://doi.org/10.4324/9781003308300
Hofstede, G. (2001). *Culture's consequences: Comparing values, behaviors, institutions, and organizations across nations* (Second edition). Sage Publications.
Kassing, J. (1998). Development and Validation of the Organizational Dissent Scale. *Management Communication Quarterly, 12*(2), 183-229. https://doi.org/10.1177/0893318998122002
Kassing, J. (2000). Investigating the relationship between superior-subordinate relationship quality and employee dissent. *Communication Research Reports, 17*(1), 58-70. https://doi.org/10.1080/08824090009388751
Kassing, J., Avtgis, T. (1999). Examining The Relationship Between Organizational Dissent and Aggressive Communication. *Management Communication Quarterly, 13*(1), 100-115. https://doi.org/10.1177/0893318999131004
Kirschner, J., Young, J.,& Fanjoy, R. (2014). Stress and Coping as a Function of Experience Level in Collegiate Flight Students. *Journal of Aviation Technology and Engineering, 3*(2), 3 14-19. https://doi.org/10.7771/2159-6670.1092
Landau, J. (2009). To Speak or Not to Speak: Indicators of Voice Propensity. *Journal of Organizational Culture, Communications and Conflict, 13*(1), 37-38. https://www.abacademies.org/articles/jocccvol13no12009.pdf
National Transportation Safety Board (1994*). Safety study. A review of flight crew involved major accidents of U.S. air carriers, 1978 through 1990.* Washington, DC: Department of Commerce, National Technical Information Service. PB 974917001
Parker, L. (1993). When to fix it and when to leave: Relationships among perceived control, self-efficacy, dissent and exit. *Journal of Applied Psychology, 78*(6), 949-959. https://doi.org/10.1037/0021-9010.78.6.949
Posit team (2025). RStudio: Integrated Development Environment for R. Posit Software, PBC, Boston, MA. URL http://www.posit.co/.
Shapiro, S. S., & Wilk, M. B. (1965). An Analysis of Variance Test for Normality (Complete Samples). *Biometrika, 52*(3/4), 591–611. https://doi.org/10.2307/2333709
Søndergaard, M. (1994). Research note: Hofstede's consequences: a study of reviews, citations and replications. *Organization Studies, 15*(3), 447–456. https://doi.org/10.1177/017084069401500307
Yoo, B. and N. Donthu (2002). The effects of marketing education and individual cultural values on marketing ethics of students. *Journal of Marketing Education, 24*(2), 92-101. https://doi.org/10.1177/0273475302242002

The Journal of Management and Engineering Integration (JMEI) © 2026 by AIEMS Inc., is licensed under CC BY 4.0. To view a copy of this license, visit https://creativecommons.org/licenses/by/4.0/

The Journal of Management and Engineering Integration Vol. 19, No. 2 | Winter 2026 DOI: 10.62704/10057/59054

AIR TRAFFIC CONTROL STAFFING DEFICIENCIES AND THEIR RELATIONSHIP TO INCIDENTS AND ACCIDENTS

Ashton Irons [1]
Gianna Mendez [1]
Sunkyum Kim [1]
Vivek Sharma [1]
Brooke E. Wheeler[1]
[1] Florida Institute of Technology
airons2022@my.fit.edu; bwheeler@fit.edu

Abstract

Air Traffic Control (ATC) in the United States has faced staffing deficiencies for several decades. These shortages could put undue strain on the number of existing controllers, thus creating a threat to safety. This archival correlational study investigated the relationship between baseline ATC staffing deficiencies and the frequency of accident and incident reports at the 62 observed U.S. large and medium hub airports. A statistically significant but weak positive relationship was found ($r = .286$, $p = .024$). All NTSB incident and accident reports were included; therefore, the findings establish that staffing shortages have a relationship with reports. This suggests a complex relationship between the two variables, with other factors likely influencing it.

Keywords: *ATC (Air Traffic Control), Staffing Deficiencies, Controller Shortages, Workload, Aviation Accidents and Incidents*

1. Introduction

Staffing shortages within Air Traffic Control (ATC) have been an increasingly pertinent topic in both the aviation industry and public discourse, especially as highly publicized accidents have broken headlines recently within the United States. Concerns about ATC staffing shortages are longstanding, highlighting the importance of examining their potential implications for aviation safety. Investigating whether staffing deficiencies are associated with the frequency of accident and incident reports can contribute to understanding this relationship.

The purpose of this study was to examine the relationship between baseline air traffic control (ATC) staffing deficiencies and the number of aviation accident and incident reports associated with the sampled airports. Staffing data were obtained from the Air Traffic Controller Workforce Plan from 2019 to 2028, which reports actual onboard staffing as of September 29, 2018, and facility staffing ranges (FAA, 2019). Aviation incident and accident report counts for January 1, 2019, and October 16, 2025, were obtained from the National Transportation Safety Board (NTSB, n.d.) Reports were counted across the full observation period for each airport. Staffing deficiencies were calculated as the percentage difference between the high end of each facility's published staffing range and its actual total onboard staffing in the FAA's (2019) ATC Workforce Plan.

The research question for this study was as follows: What is the relationship between baseline air traffic control staffing deficiencies and the number of aviation incident and accident reports at

Submitted: April 30, 2026
Revised: October 19, 2026

the sampled airports between 2019 and 2025? A positive relationship between ATC staffing deficiencies and the number of aviation incidents and accident reports between 2019 and 2025 was expected, such that an increase in staffing deficiencies is expected to be associated with an increase in aviation incidents and accidents.

The study generated a combined dataset including controller staffing levels and published staffing ranges for facilities associated with 62 U.S. large and medium hub airports (FAA, 2019), along with the corresponding number of aviation incident and accident reports between 2019 and 2025 (NTSB, n.d.). Airports were identified using the National Plan of Integrated Airport Systems (NPIAS) 2025 to 2029 (FAA, 2024) and matched to the workforce plan (FAA, 2019). This study investigated whether baseline staffing deficiencies statistically correlate with counts of incident and accident reports. This information can contribute to understanding the relationship between controller staffing and aviation safety and inform further research on staffing and resource allocation. The findings are relevant to U.S. large and medium hub airports because of the sample's broad coverage of nearly all airports in these categories identified in the NPIAS 2025 to 2029.

2. Literature Review

The chronic shortage of air traffic controllers in the United States is a deep-rooted systemic crisis. ATC staffing deficiencies can increase operational demands within facilities, raising concerns about controller workload, fatigue, and aviation safety. These concerns involve both the availability of personnel and the operating conditions where controllers perform their duties. This section reviews the history of ATC staffing deficiencies, the operational consequences of fatigue, and their relevance to aviation safety.

2.1. Air Traffic Control Shortages

In 1981, approximately 11,347 U.S. air traffic controllers were terminated by President Reagan following a strike by the Professional Air Traffic Controllers Organization (PATCO; Shostak, 2006). This accounted for roughly two-thirds of the active controllers at the time. Additionally, the U.S. government imposed hiring freezes to reduce the federal workforce and minimize government spending. Simultaneously, further deregulation of the airline industry under Reagan led companies to lower airfare prices, thereby increasing market demand and further straining the available labor.

The Reagan-imposed hiring freeze resulted in no new incoming controllers being hired. Between 1982 and 1991, an average of 2,655 controllers were hired per year (Ruiz, 2007, p. 28). While the U.S. did move to hire a limited number of new controllers, the predominant focus was streamlining facilities and updating the technology used by controllers to expand the tower's capabilities (Fisher & Marciano, 1997). Although these technological updates and advances in automation enhanced ATC operations, workforce strain remained pronounced.

An inconsistent stop-start approach to hiring air traffic controllers created an uneven age profile within the workforce. This is particularly problematic given the career's nature as a lifelong job with no alternative source of hires due to its being a high-skilled labor position. A surge in hiring led to future bulk retirements, in which large segments of the workforce become eligible to retire simultaneously. The FAA expected an even larger shortage due to the number of active controllers currently nearing the mandatory retirement age of 56 (Casebolt et al., 2018). An example of this is the period between 2002 and 2012, during which large swaths of the ATC workforce became eligible for retirement (Ruiz, 2007, p. 28). This forced the FAA into a reactive cycle of frantic hiring, resulting in an "echo effect" where retirements increase and

decrease dramatically. This echo effect persists to this day, recurring approximately every 25 years (Fisher & Marciano, 1997).

In the latter half of the 1980s, the FAA initiated the Air Traffic-Collegiate Training Initiative (AT-CTI) to create a new stream of desperately needed air traffic controllers. The program created a pathway for two and four-year college students to become air traffic controllers, bypassing entry-level training formerly required at the academy (Coyne, 2014). The creation of the AT-CTI program is illustrative of the FAA's understanding that air traffic controllers were in short supply, thus prompting them to establish a cost-effective method to further boost hiring.

Human operators still must multitask in a high-stress, demanding environment. The human ability to process information remains limited. ATC's tasks are often complex and highly demanding, and the challenges can compound when multiple tasks arise simultaneously (Black et al., 2022), increasing the risk of human factors related errors. This sets a framework for understanding where ATC shortages originate, consistent systemic failures, and reactive corrections that were not sufficient or appropriate solutions. Repetitive band-aid solutions have now led the FAA to a larger problem with increasingly dire consequences.

2.2. ATC Operational Deficiencies and Subsequent Consequences

The structural issue of ATC staffing shortages creates operational deficiencies within facilities. One of the most critical of these is controller fatigue. These human factors issues compromise cognitive performance and pose a threat to aviation safety. Staffing deficiencies translate operationally into increased workload, overtime, and disruption of typical shift patterns, which are common catalysts for controller fatigue. Fatigue is not merely tiredness, but a biological state of decreased physical and or mental performance capacity resulting from sleep loss, excessive workload, or prolonged wakefulness (Alomar et al., 2024; International Civil Aviation Organization [ICAO], 2020). Understaffing contributes to fatigue by forcing facilities to rely on schedules that fail to adequately balance workload and rest.

Compounding the problem, the operational environment is so demanding that even monitoring fatigue presents a significant challenge. Terenzi et al. (2024) reported difficulty in obtaining consistent fatigue data from controllers, showcasing the burden of data collection on an already strained workforce. These difficulties highlight the challenges of consistently monitoring fatigue within operational ATC environments. Guo and Bard (2024) stated that the FAA lacks standardized tools for creating efficient controller schedules, which results in frequent shift adjustments that hinder performance. This operational reality is depicted in predictive safety models, such as the Loss-of-Control Accident Framework (LOCAF) developed by Ancel et al. (2015). Their model identifies ATC-related issues as a direct cause of task saturation, a cognitive state similar to the impaired characteristics observed with fatigue. In both high- and low-workload situations, which are common in understaffed facilities, controller performance is likely to decline. A high workload can overwhelm a fatigued controller, whereas a low workload often fails to provide the necessary stimulation (Guo & Guan, 2025). This is compounded by systemic inefficiencies in workforce distribution. Past research has also indicated that controllers at low-density facilities reported lower workload and boredom, demonstrating differences in operational demands across facilities (Terenzi et al., 2024). These differences highlight the importance of considering facility characteristics and work schedules when examining controller workload and fatigue.

Kontogiannis and Malakis (2009) provided an additional framework for understanding the relationship between fatigue and degradation of specific cognitive processes. They

demonstrated that effective error management depends on proactive cognitive strategies such as awareness-based detection and planning-based detection. These cognitive activities require substantial mental resources, which are significantly depleted by fatigue resulting from chronic overtime and irregular schedules (Guo & Bard, 2024). When controllers are fatigued, their ability for self-introspection and team cross-checking is impaired, creating an environment where errors in judgment and situational awareness are likely to go undetected (Kontogiannis & Malakis, 2009). This suggests that fatigue not only reduces individual cognitive effectiveness but also undermines the collective safety net of team coordination, ultimately heightening operational risk in air traffic control environments.

The fatigued state resulting from these operational deficiencies leads to a measurable degradation of core cognitive skills essential for air traffic control. Research consistently shows that fatigue impairs concentration, slows reaction times, reduces coordination, and degrades decision-making and situational awareness (Alomar et al., 2024). These impairments are observable as performance errors in an aviation controller environment, such as missed radio calls or incorrect guidance. The link between fatigue and performance degradation in ATC is not theoretical; recent technological advancements have created opportunities to detect fatigue-related indicators. Guo and Guan (2025) developed a model for real-time ATC fatigue recognition using an improved Xception network that achieved a 94.36% accuracy in identifying eye fatigue and a 91.68% recall rate (p. 14). Their findings suggest that image-based methods may support fatigue monitoring, although detection performance alone does not establish the effect of fatigue on controller performance or aviation safety.

2.3. Accidents/Incidents

The 49 Code of Federal Regulations (CFR) § 830.2 (n.d.) describes an accident as an occurrence between one or more aircraft with the intent to fly where any individual sustains serious death or injury, and/or the aircraft sustains substantial damage. An incident is defined in 49 CFR § 830.2 as a non-accident occurrence that affects or has the potential to affect safe flight. The NTSB (n.d.) utilizes these same criteria to categorize reports in its accident database.

Historical accident analysis underscores the severe consequences of fatigue. For instance, fatigue was identified as a causal factor in the 1993 Guantanamo Bay crash and the 2009 Colgan Air Flight 3407 accident (Alomar et al., 2024). Liu et al. (2023) found that human factors primarily drive the development of risk topics over time. Heng et al. (2022) described controller workload as the psychological and physical stress that controllers experience during their shift to ensure efficient and safe operation and flow of air traffic. In an ATC environment, controller workload and fatigue are often exacerbated by having an inadequate number of controllers. The causal pathway from organizational failure to accident is quantitatively demonstrated in safety analytics. The object-oriented Bayesian network (OOBN) model, as proposed by Ancel et al. (2015), illustrated how deficiencies in organizational oversight and poor resource management can propagate throughout the system. These deficiencies increase the probability of task saturation and performance deficiency, which the model directly links to a greater chance of an accident. A notable example occurred in January 2025. An American Airlines CRJ700 and Black Hawk helicopter collision in Washington, D.C., showed a lapse in ATC control coordination and procedures (Pan et al., 2025). They highlighted concerns about ATC staffing and coordination, illustrating the importance of further examining how operational conditions relate to aviation safety. These examples underscore the critical need for continued research into staffing deficiencies and their role in contributing to controller fatigue, which threatens overall aviation safety. In summary, the literature identifies fatigue and workload as important

concerns for aviation safety. Continued investigation of ATC staffing deficiencies may help clarify how workforce conditions relate to controller performance and reported safety outcomes.

3. Methods

The research design for this study was correlational. Staffing data were obtained from the FAA's (2019) Air Traffic Controller Workforce Plan 2019 to 2028. The Air Traffic Controller Workforce Plan is organized by the FAA Administrator for the Senate Committee on Commerce, Science, and Transportation and the House of Representatives Committee on Transportation and Infrastructure. The report contains information about facilities and services, staffing, lawsuits, losses, hiring plan, hiring process, training, and funding status. For this study, current total staffing for each facility, and the HIGH end of of the staffing range were used to calculate staffing deficiencies (FAA, 2019). Large and medium hub airports were identified using the National Plan of Integrated Airport Systems 2025 to 2029 (FAA, 2024) and matched by airport identifier to terminal-facility staffing records in the workforce plan (FAA, 2019). These airport categories were selected because of their commercial passenger activity and relevance to examining ATC staffing at busy airports. The sample included 62 airports, consisting of 30 large and 32 medium hubs. Airports with no accident or incident reports returned by the NTSB search were included in the dataset with a report count of zero. Honolulu International Airport was excluded because the workforce plan provided staffing for Honolulu Control Facility without a separate airport-tower staffing row like the other airports. The sample was selected using these sources and the requirement that the facility also have ATC staffing information published.

The staffing fields collected for each facility included certified professional controllers (CPC), certified professional controllers in training (CPC-IT), developmental controllers, total onboard staffing, and the lower and upper endpoints of the published staffing range. Total onboard staffing included CPC, CPC-IT, and developmental controllers. Staffing deficiency was operationalized as Staffing deficiency (%) = [(HIGH Staffing Range - TOTAL)/HIGH Staffing Range] x 100. HIGH represented the upper endpoint of the published staffing range, and TOTAL represented the actual onboard staffing as of September 29, 2018 (FAA, 2019). Positive values indicated total onboard staffing below the upper endpoint, meaning fewer controllers than the high end of the staffing range. While negative values indicated staffing above the high end of the staffing range. Expressing staffing deficiencies as percentages allowed comparison across facilities with different staffing levels. Each airport contributed one baseline staffing-deficiency value.

Accident and incident report counts were obtained from the NTSB (n.d.) Aviation Investigation Search database for events occurring between January 1, 2019, and October 16, 2025. The NTSB Aviation Investigation Search is managed by the NTSB, a credible, independent, federal agency tasked with investigating incidents and accidents concerning transportation in the U.S and its territories. The NTSB database contains data from 1962 through the present day (NTSB, n.d.) and includes reports for the investigations conducted by the agency, consisting of an analysis of the accident, probable cause, findings, and factual, medical, and pathological information if applicable. For this study, only the number of accidents and incidents (NTSB, n.d.) under 49 CFR § 830.2 for each airport were used. . Searches used each airport's three-letter identifier, the state of the airport, a 25-mile distance option, and the date range of January 1, 2019, and October 16, 2025. All other fields such as aircraft characteristics, injury level, damage, operation type, etc. were not used to restrict the search. All NTSB reports during the time period were included.

The data from both sources were integrated into a combined Excel file, matched by airport

identifier. Reports were not restricted to events involving ATC-related causes. The outcomes represented the number of airport-associated accident and incident reports across the complete observation period, rather than annual or traffic-adjusted rates.

Descriptive statistics summarized staffing deficiencies and report counts using the mean, median, range, and standard deviation. Pearson's product-moment correlation (r) was calculated to examine the strength and direction of the linear association between the two continuous variables. Statistical significance was assessed using a p-value, with an alpha value of 0.05 as the threshold for significance. The full sample of 62 airports was used for the primary analysis. Potential report-count outliers were identified using Z-scores calculated from the sample mean and sample standard deviation, with $|z| > 3$ as the criterion. An analysis excluding the outlier ORD was conducted using 61 airports. Descriptive statistics and Pearson's correlation were calculated both with and without the outlier to examine how excluding ORD affected the results. The correlation results were calculated in RStudio 2026.09.0.

4. Results

The sample included a total of 62 U.S. large and medium hub airports, with a breakdown of 30 large hub and 32 medium hub airports selected from the Air Traffic Control Workforce Plan (ATCWP) 2019-2028 (FAA, 2019). Three airports had no incident or accident reports in the NTSB database during the selected timeframe; these airports were included in the dataset with report counts of zero. A total of 194 accident and incident reports were collected from the NTSB (n.d.), and one baseline-staffing deficiency percentage was calculated for each airport using the FAA (2019) ATC workforce plan.

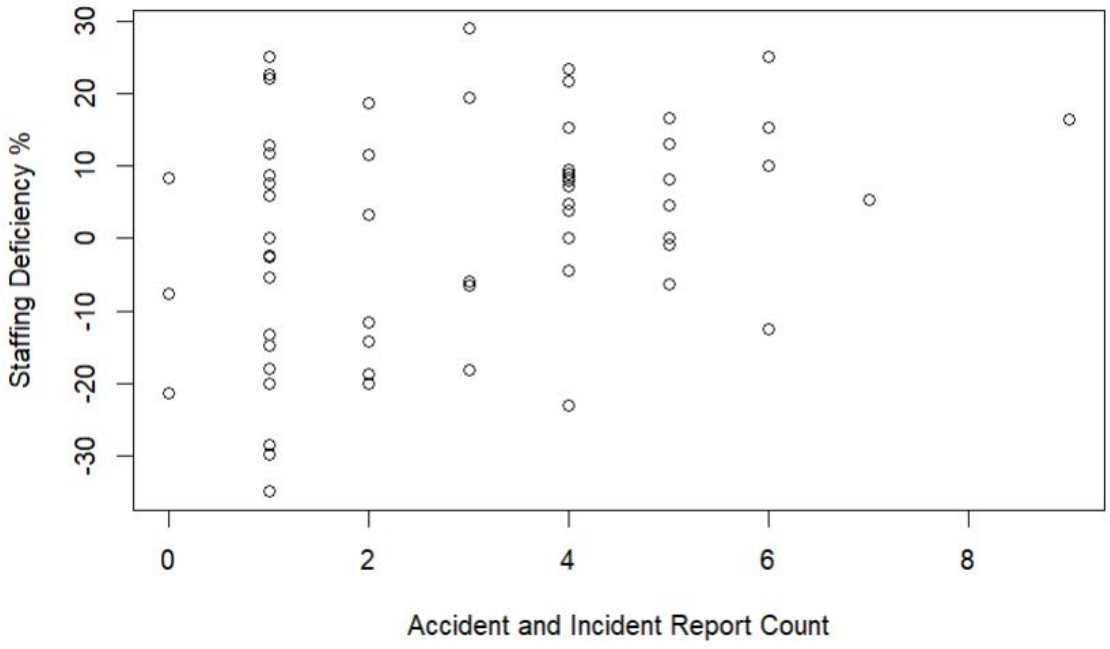


Figure 1. Correlation Between Staffing Deficiency and Accident Frequency

Table 1. Descriptive Statistics for Reports and Staffing Deficiency

	Mean	Median	Range	Standard Deviation
Reports	3.13	3.00	0.00 - 12.00	2.34
Staffing Deficiency	1.59	4.67	-35.00 - 28.95	15.17

Note. Reports represent the number of incident and accident reports associated with each airport across the observation period. The staffing deficiency was calculated as ((HIGH Staffing Range-TOTAL)/ HIGH Staffing Range) x 100. The standard deviation for staffing deficiency is expressed in percentage points.

Table 1 shows the descriptive statistics for Reports and staffing deficiencies. The mean number of reports per airport was relatively low (M = 3.13, Mdn = 3), with a standard deviation of 2.34 and a range of 0 to 12. The mean staffing deficiency was 1.59 (Mdn = 4.67%), with a range of -35.00% to 28.95% and a standard deviation of 15.17%. Negative staffing-deficiency values indicated total onboard staffing above the upper endpoint of the facility's published staffing range. Figure 1 suggests that a small number of airports have a large, negative staffing deficiency percentage, which pulls the mean downward.

The assumptions for employing Pearson's r were evaluated. Both variables are continuous ratio data, the paired data points are independent of one another as they are for unique airports, and initial analysis confirmed a linear relationship (see Figure 1). The outlier assumption was not met. The outlier was identified using Z-scores, and descriptive statistics were recalculated with the outlier excluded (Table 2), and Pearson's r was evaluated with and without the outlier. A Shapiro-Wilk test did not indicate a violation of the normality assumption for staffing deficiencies ($W = .972$, $p = .168$) but indicated a violation of normality for report counts ($W = .885$, $p < .0001$). Despite this, analysis proceeded under the Central Limit Theorem (CLT), which states that the sampling distribution of the mean will approximate a normal distribution as the sample size increases. Given the sample size ($n = 62$), the CLT states that the means derived from the data are likely to be approximately normally distributed (Gallo et al. 2023). The initial full-sample analysis revealed a statistically significant, weak, positive relationship between the number of reports and staffing deficiencies, $r(60) = 0.286$, $p = .024$. The 95% confidence interval [.039, .500] excluded zero, supporting a positive correlation, although its width indicated uncertainty about the strength of that association.

ORD was identified as an outlier; it had 12 reports and a report-count Z-score of 3.79, exceeding the screening threshold of $|z| > 3$. ORD was identified as an outlier based on its report count, not its staffing deficiency, which was 2.74%. No staffing-deficiency observations exceeded $|z| > 3$. Excluding ORD reduced the sample from 62 to 61 airports and the total number of reports from 194 to 182. Table 2 shows the descriptive statistics without ORD (N = 61). The mean report count decreased from 3.13 to 2.98, and the sample standard deviation decreased from 2.34 to 2.06. The mean staffing deficiency decreased slightly from 1.59% to 1.57% because ORD's staffing deficiency of 2.74% was above the full-sample mean. The staffing-deficiency range remained unchanged because ORD did not have either the minimum or maximum value. After excluding ORD, a slightly stronger positive correlation was observed, $r(59) = .322$, $p = .011$, $r^2 = .104$. The 95% confidence interval [.077 - .531] excluded zero, supporting a positive association, although uncertainty remained about its strength. The r^2 indicates approximately 10.4% shared variation between staffing deficiencies and report counts in the linear relationship. Both analyses indicated a

statistically significant positive association, but neither established that staffing deficiencies caused the reported accidents or incidents as the design is correlational.

Table 2. Descriptive Statistics for Reports and Staffing Deficiency without Outlier

	Mean	Median	Range	Standard Deviation
Reports	2.98	3.00	0.00 - 9.00	2.06
Staffing Deficiency	1.57	4.69	-35.00 - 28.95	15.30

5. Discussion

The findings of this study revealed a statistically significant positive relationship between ATC staffing deficiencies and aviation incident reports, as hypothesized. While the full-sample correlation ($r = .286$, $p = .024$) indicated that airports associated with larger staffing deficiencies tended to have more incident and accident reports, the effect size ($r^2 = .082$) corresponded to approximately 8.2% shared variation between the variables in the linear relationship. After excluding ORD, the positive association remained and was slightly stronger, $r = .322$, $p = .011$, $r^2 = .104$. This establishes a weak, positive, linear relationship between the variables.

The historical context of ATC staffing provides background for interpreting these findings. Shostak (2006) and Ruiz (2007) discussed workforce challenges associated with the PATCO strike and subsequent recruitment and retirement patterns. Fisher and Marciano (1997) identified "retirement bulges" that occur in 25-year cycles. This persistent strain may have led to the development of adaptation mechanisms within the ATC system, and thus a weaker relationship. The literature suggests that facilities facing persistent staffing deficiencies may use strategies, such as redistributing workloads or implementing safety redundancies to help potentially mitigate the immediate impact of staffing shortages. As examined by Terenzi et al. (2024), differences in work schedules and facility conditions may also be relevant when considering workload and fatigue. These factors could influence the relationship between staffing deficiencies and safety outcomes, although their contribution was not evaluated in this study.

The operational realities of ATC work further complicate this relationship. Kontogiannis and Malakis (2009) described cognitive strategies involved in effective error detection and identification. Fatigue and workload may be relevant to these processes, although neither was measured in the present study. The weak correlation observed leaves questions about how other operational factors may relate to staffing deficiencies and accident and incident report counts. Technological advancements, such as the improved Xception network developed by Guo and Guan (2025) for real-time fatigue detection, may offer facilities opportunities to identify and potentially mitigate fatigue-related risks before they turn into reportable incidents and accidents. Several factors must be considered when interpreting these results. The archival nature of the data may not capture the full complexity of ATC operations. Terenzi et al. (2024) report that consistent data collection on fatigue was challenging in already strained environments, suggesting that official reports may underrepresent the true safety climate of the facilities. Additionally, the low base rate of incidents and accidents in U.S. aviation contributes to a weak signal-to-noise ratio, making it challenging to detect meaningful relationships even if they exist. In the context of this study, the relationship between staffing shortages and any safety outcomes may be obscured by random variation or by the influence of numerous external factors. Variables, such as weather conditions,

technological changes, and changes in traffic volume, were not controlled in this study, and these factors likely impacted the observable relationship between staffing deficiencies and the number of accidents and incidents. Furthermore, organizational dynamics, such as management quality, safety culture, and training effectiveness, also play an important role in ATC facilities. Strong management and well-designed training programs can mitigate staffing shortages by maintaining consistent procedures, fostering clear communication, and promoting a robust safety culture. Conversely, ineffective management and/or training may exacerbate weaknesses in the facilities, allowing fatigue and miscommunication to go unchecked. In this context, the combination of low incident and accident frequency, along with the numerous variables involved, makes it difficult to clearly identify the relationship of staffing to safety outcomes.

Future research should build on these findings. Investigating whether particular levels of staffing shortages are associated with increased safety risks would benefit facility management. Additionally, incorporating measures of controller fatigue, such as the fatigue-related indicators examined by Guo and Guan (2025), could help clarify the possible role of human performance in the relationship between staffing deficiencies and accident and incident report counts. Lastly, comparing facilities with similar staffing levels but different technologies or management practices could help identify operational factors relevant to this relationship.

In conclusion, ATC staffing deficiencies and the number of incident and accident reports showed a statistically significant, but weak positive correlation. The positive association remained after ORD was excluded. These findings were consistent with the research hypothesis, establishing a relationship between staffing deficiencies and accident and incident reports. Examination of staffing, human performance, and operational conditions may contribute to further understanding this relationship. As the FAA works to improve staffing levels through initiatives, such as the AT-CTI program, understanding how these factors interact will be crucial for developing strategies that enhance both workforce stability and aviation safety.

6. Disclaimer Statements

Funding: None

Conflict of Interest: None

7. References

Ancel, E., Shih, A. T., Jones, S. M., Reveley, M. S., Luxhøj, J. T., & Evans, J. K. (2015). Predictive safety analytics: inferring aviation accident shaping factors and causation. *Journal of Risk Research*, *18*(4), 428–451. https://doi.org/10.1080/13669877.2014.896402

Alomar, I., Alomar, S., & Stecenko, I. (2024). Improvement of fatigue management methodology related to flight crew. *Aviation*, *28*(3), 148–162. https://doi.org/10.3846/aviation.2024.22146

Black, S. C., Bender, A. D., Whitney, S. J., Loft, S., & Visser, T. A. W. (2022). The Effect of Multi-Tasking Training on Performance, Situation Awareness, and Workload in Simulated Air Traffic Control. *Applied Cognitive Psychology*, *36*(4), 874–890. https://doi.org/10.1002/acp.3977

Casebolt, M. K., Depperschmidt, C. L., & Bliss, T. J. (2018). Transformation of Air Traffic Collegiate Training Initiative (AT-CTI) Hiring Process: Institutional Perspectives. *Collegiate Aviation Review International*, *36*(2), 1–17. https://doi.org/10.22488/okstate.18.100487

Coyne, W. (2014). A Review and History of the Air Traffic-Collegiate Training Initiative Program. *International Journal of Aviation, Aeronautics, and Aerospace, 1*(3). https://doi.org/10.58940/2374-6793.1025

FAA. (2019). *Air Traffic Controller Workforce Plan 2019-2028*. https://www.faa.gov/sites/faa.gov/files/2021-11/2018-ABA-200-CWP_2019_Report_508.pdf.

FAA. (2024). National plan of integrated airport systems 2025–2029: Appendix A—List of NPIAS airports.

https://www.faa.gov/sites/faa.gov/files/airports/planning_capacity/npias/current/ARP-NPIAS-2025-2029-Appendix-A.pdf

Fisher, E. G., & Marciano, V. (1997). Managing Human Resource Shortages in a Unionized Setting: Best Practices in Air Traffic Control. *Journal of Labor Research*, *18*, 287–314. https://doi.org/10.1007/s12122-997-1040-5

Gallo, A., Wheeler E., Silver, M. (2023) *Fundamentals of Statistics for Aviation Research.* Routledge.

Guo, J., & Bard, J. F. (2024). Air traffic controller scheduling. *Computers & Industrial Engineering*, *191*. https://doi.org/10.1016/j.cie.2024.110123

Guo, M., & Guan, Z. (2025). Air traffic controller work state recognition based on improved xception network. *PLoS One*, *20*(5), e0322404. https://doi.org/10.1371/journal.pone.0322404

Heng, Y., Wu, M., Wen, X., & Bedon, C. (2022). Identifying Key Risk Factors in Air Traffic Controller Workload by SEIR Model. *Mathematical Problems in Engineering*, *2022*. https://doi.org/10.1155/2022/7600754

International Civil Aviation Organization. (2020). *Manual for the Oversight of Fatigue Management Approaches.* Second Addition, Version 2 (Revised). Doc. 9966. ICAO.

Kontogiannis, T., & Malakis, S. (2009). A proactive approach to human error detection and identification in aviation and air traffic control. *Safety Science, 47*(5), 693-706. https://doi.org/10.1016/j.ssci.2008.09.007

Liu, W., Zhang, H., Shi, Z., Wang, Y., Chang, J., & Zhang, J. (2023). Risk Topics Discovery and Trend Analysis in Air Traffic Control Operations—Air Traffic Control Incident Reports from 2000 to 2022. *Sustainability*, *15*(15), 12065. https://doi.org/10.3390/su151512065

NTSB. (n.d.). *Aviation Investigation Search.* https://www.ntsb.gov/Pages/AviationQueryv2.aspx.

Ruiz, J. R. (2007). Current trends in air traffic controller recruitment and training. *Journal of Aviation/Aerospace Education & Research, 16*(2) Retrieved from http://commons.erau.edu/jaaer/vol16/iss2/4

Shostak, A. (2006). Finding Meaning in Labor's "Perfect Storm": Lessons from the 1981 PATCO Strike. *Employee Responsibilities & Rights Journal*, *18*(3), 223–229. https://doi.org/10.1007/s10672-006-9020-0

Terenzi, M., Tempestini, G., & Di Nocera, F. (2024). Air Traffic Controllers' Rostering: Sleep Quality, Vigilance, Mental Workload, and Boredom: A Report of Two Case Studies. *Aerospace*, *11*(6), 495. https://doi.org/10.3390/aerospace11060495

Weijun, P., Yinxuan, L., Yanqiang, J., Rundong, W., Yujiang, F., & Gaorui, X. (2025). Analysis of Controller-Caused Aviation Accidents Based on Association Rule Algorithm and Bayesian Network. *Applied Sciences*, *15*(17), 9690. https://doi.org/10.3390/app15179690

49 CFR § 830.2. (n.d.). *Electronic Code of Federal Regulations.* Accessed Sept 18, 2025. https://www.ecfr.gov/current/title-49/subtitle-B/chapter-VIII/part-830/subpart-A/section-830.2

The Journal of Management and Engineering Integration (JMEI) © 2026 by AIEMS Inc., is licensed under CC BY 4.0. To view a copy of this license, visit https://creativecommons.org/licenses/by/4.0/

2026 DOI: 10.62704/10057/59055

FROM CURB TO GATE: HOW USABLE IS YOUR AIRPORT APP?

Deborah Sater Carstens [1]

Arjun M. Nair [1]

Syed Ali Abbas [1]

Anna Berenika Wojdecka [1]

[1] Florida Institute of Technology

carstens@fit.edu; anair2014@my.fit.edu; awojdecka2018@my.fit.edu; sabbas2016@my.fit.edu

Abstract

The research examined the usability of an airport App prototype. The usability study consisted of think-aloud sessions, open-ended interviews, and a USE questionnaire with 14 participants. The evaluation focused on assessing the usability, clarity, and effectiveness of core functions to enhance the passenger experience. The findings identified key strengths and usability deficiencies, providing evidence-based guidance for future design iterations of the passenger App. The study implications and future research are also addressed.

Keywords: *App, Passenger, Usability*

1. Introduction

The design and implementation of new touchless interfaces for airport procedures will play a crucial role in the future of air travel (Nair et al., 2021). This includes gesture-based sensing, touchless fingerprint detection, and voice activation. Those solutions require a high-quality UX design to be accessible and intuitive, with robust built-in security measures (Iqbal, 2020). This study examined the usability of the FlyKey Passenger App prototype (see Figure 1). The evaluation focused on the clarity and effectiveness of core functions to inform recommendations for refining the App's design and passenger experience. The App is a proposed design solution to address airport and passenger concerns by enhancing the passenger experience and enabling airport management to improve passenger efficiency and safety. The design focus is on implementing technology to ease the passenger experience by making the airport experience convenient. The FlyKey platform is a digital solution that optimizes the travel experience, provides a comprehensive contactless approach, and provides the following three layers of functionality:

- **A passenger-processing FlyKey App** facilitates the whole contactless airport experience, from check-in to arrival, and provides a secure digital identification (ID) certification.
- **A tablet FlyKey-Personnel App** enables smooth contactless staff-passenger interactions with airport personnel.
- **A central FlyKey-Analytics platform** integrates passenger information and provides real-time analytics of passenger flow through the airport. Using trend and activity prediction algorithms, FlyKey radically changes the way passenger time-space allocation is optimized.

Submitted: April 30, 2026
Revised: October 20, 2026



Figure 1. FlyKey Platform in Development Encompasses Three Dimensions of Future Travel - FlyKey Passenger, Personnel, and Analytics

The concept was developed using a unique approach that fuses human-centered design research methods with emerging technological innovations. The present research focused on the first two components of the platform: usability testing of the passenger-facing App and further design research to develop and iterate the personnel-facing App. The design has benefits consisting of: (a) passengers traveling efficiently at their convenience, by flexibility, (b) airports receiving real-time analytical data about the airport's operational status, and (c) airports being able to strategically ease the peak travel bank congestion at hub airports by staggering passengers at the different waypoints.

2. Literature Review

Lund (2001) proposed the Usefulness, Satisfaction, and Ease of Use (USE) questionnaire as an instrument for measuring improvements in user interface usability. The full USE questionnaire consists of a 30-item Likert-scale survey in which participants rate products on a 7-point Likert scale, across four key usability areas – Ease of Learning, Ease of Use, Satisfaction, and Usefulness. Several studies have successfully used the USE questionnaire, including Shafitri and Sofica (2023), which assessed ride-sharing and transportation applications in Indonesia, and Gao et al. (2018), which assessed the USE questionnaire developed by Lund and evaluated its reliability and validity via the Amazon MTurk platform. Gao et al.'s findings indicate Cronbach's alpha of 0.98, indicating very high reliability of the questionnaire. In relation to the current study, the USE questionnaire complements the think-aloud activity by measuring and validating findings from the usability assessment, including those from the think-aloud study.

Saadati et al. (2022) conducted two design workshops utilizing scenarios and persona development. The researchers gained critical feedback from frontline employees to design potential technologies for deployment in airport operations. Workshops included 14 stakeholders, consisting of airport staff, user experience researchers, and airport management. The study design included qualitative research, in which the workshops were audio-recorded, transcribed, and subsequently analyzed thematically. Findings from the study suggest that design workshops, combined with domain analysis, could serve as an early-warning pilot test, identifying socio-technical problems that would typically not be identified until live trials and mass rollouts (Saadati et al., 2022). Feedback from the study suggests that pre-implementation user workshops led to more human-centered product design and helped prevent late-stage missteps.

Lee and Lee (2025) conducted a usability analysis of airport self-service check-in kiosks at the Incheon International Airport in South Korea. The study included 48 participants. The researchers

conducted video observations of passenger behavior at self-service technology kiosks. The observations were qualitatively coded, following intercept interviews with each user to record feedback. Following this, Lee and Lee conducted detailed stakeholder interviews to discuss the issues identified. The findings suggest that visually heavy interfaces created challenges for groups deemed digitally vulnerable, particularly older participants and participants with other disabilities. Qualitative follow-ups with participants via interviews were used to identify root causes of functionality issues, including a lack of audio assistance or guidance. The researchers recommended integrating a "multi-modal conversational AI assistant" into the kiosk, a recommendation further supported by the stakeholder interviews. Some of the implications of combining user observations with stakeholder feedback include future augmentation of the technology into greater accessibility and user-friendly features. As a result of Lee and Lee's findings, Incheon International Airport began exploring the implementation of an AI voice assistant into the self-service kiosk. The study integrated user experience data with real-world observations of FlyKey usage to detect issues.

Stone and Wang (2011) conducted human-computer interaction research to develop and test an airport wayfinding application named AirportLogic. The study included eight participants, namely college students who flew at least once a year. The mobile application was intended specifically for US airports to provide navigation through the airport environment. The researchers utilized usability testing subsequently iterative design method, similar to the approach the FlyKey team undertook. Stone and Wang conducted a preliminary usability study of the existing airport mobile applications using think-aloud tasks with six participants, who identified common usability issues. Following the findings from the think-aloud task, the researchers developed a prototype of their application, which was subsequently tested through task-based usability tests. The participants' actions were recorded, after which they were administered a post-test questionnaire to gain their subjective feedback regarding the application's usability. Findings from the think-aloud tasks evaluating existing airport mobile applications revealed weak information architecture and data clutter, leading to user confusion and delayed task completion. The prototype developed by the researchers was then subject to testing by all the participants, who successfully completed the tasks assigned to them and reported being satisfied with the application's Ease of Use. The think-aloud tasks provided the researchers with insights that led to the refinement of the user interface by removing points of confusion. The findings of the study highlight the importance of early-stage usability testing, user prototyping, and implementing user feedback into the product design, which subsequently led to a better user experience. The researcher suggested testing in the real airport environment, which the FlyKey team successfully integrated into our prototype testing.

Lastly, McDonald et al. (2012) investigated the use of the think-aloud method to identify gaps between the theoretical model and practical usage of the think-aloud model. The research included 207 participants who were usability professionals globally, via an internet-based survey. The study aimed to investigate the preferences of usability practitioners in the use of the think-aloud tool. Findings from McDonald et al. suggest that the practical use of think-aloud tasks remains a core testing method for usability testing, and provides for practical flexibility, which is common amongst usability practitioners. To summarize the findings, this study evaluated best practices in usability design and incorporated suggestions and recommendations from Lee and Lee (2025), Stone and Wang (2011), and Saadati et al. (2022) into the prototype development and usability testing of FlyKey's iterations.

3. Methodology and Results

A usability study was conducted to gather critical data to enhance the design of the FlyKey Personnel and FlyKey Passenger Apps. Prior to data collection, the research was approved by a

university Institutional Review Board (IRB). The objective of the research was to gather feedback on the proposed systemic solution to inform App iterations and development, further aligning the FlyKey App with user requirements. The FlyKey platform placed a key development focus on the passenger flight experience to ensure the overall system's desirability, with extensive attention to the FlyKey Passenger App prototype and platform touchpoints. The research not only provided guidance on enhancing the passenger App but also offered insights useful for further development and innovation of the FlyKey Personnel and the platform integration with the FlyKey Passenger App within the touchless interface ecosystem.

The researchers conducted usability testing of the FlyKey - passenger App with 14 participants representing a range of technological familiarity and gender balance within airports. The participants were approached by the researchers within the public waiting area at the airport while awaiting to board flights. As required by the university IRB, steps were taken to ensure participant confidentiality. The main goals of the usability study were to determine the clarity and effectiveness of the functionality. The usability testing protocol, as outlined in the IRB, consisted of the following for each of the 14 participants:

1. Think-Aloud task with the App demonstration (Adobe XD on a spare phone) and the 'map' (displayed on a tablet) as shown in Figure 2. In Table 1, the points of confusion identified during the task are noted, along with key insights. The materials included the map with marked traversal locations (see Figure 3).
2. Open-ended interview with key insights is discussed in this section.
3. USE questionnaire and key findings are identified (see Table 2).

3.1. Think-Aloud Task

The evaluation focused on the points of confusion the user may have when using the App, and if their overall experience in navigating the App was enjoyable. The time and errors were collected during the Think-Aloud task. Figure 2 depicts mock images used to simulate real-world use of the Think-Aloud task. A smartphone with the click-through AdobeXD prototype installed was utilized by participants. Sanitizing disinfection wipes were also used to disinfect the phone screen between use of participants.

The participants were tasked with booking and navigating a journey using the FlyKey App. Participants began the evaluation given the scenario that they were planning a flight to their desired destination while using the App along key journey points. The conducted 'Think-Aloud' Task consisted of participants receiving a briefing on each prototype screen and on how to progress through the tasks. This resulted in participants offering insights for design recommendations and key information about the points of confusion and main errors users encountered while navigating the simplified airport map (see Figure 3). Table 1 illustrates the format of the data collection table used, including one of the 14 annotations made during the Think-Aloud sessions.



Figure 2. The FlyKey Prototype Testing During the Think-Aloud Task

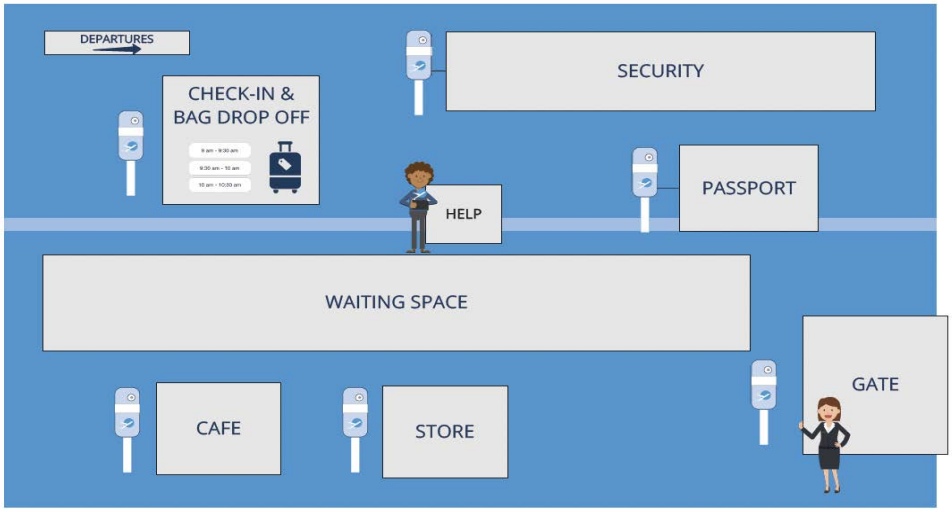


Figure 3. The Supporting Map for the Think-Aloud Task

Table 1. Time and Errors Data Sample for the Think-Aloud Task

Participant Number	Points of Confusion	Errors	Time A: Register in the App and check-in time selection	Time B: Self-check-in at the airport
Participant 1	There were four points of confusion (i.e., self-verification - photo screen, map and navigation, when to self-verify, and in finishing task A - lack of confirmation/feedback)	Time slot baggage drop-off - change, verification icon, and placing finger on the 'fingerprint' of 'verify me'	Under 2 minutes	Under 8 minutes (delay related to the confusion around the map navigation)

In summarizing the points of confusion data collected, the key redesign recommendations from the Think-Aloud Tasks are listed below:

- Incorporate 'confirmation' screens after an action is complete, which will communicate to the user that the task has been successfully completed (e.g., check-in time slot successfully selected).
- Utilize larger icons at the bottom of the screen to improve legibility.
- Utilize larger text descriptions on the main screen.
- Change the design of the 'Verify me' button to prevent confusion from the 'fingerprint' icon.
- Utilize clearer signposting and navigation during task completion (e.g., need for a 'back' button).

3.2. Open-Ended Interview

The open-ended interview consisted of the following 12 questions:

1. Could you tell me about your role at the airport (please note that this question was asked to confirm passengers were serving as the participants for the study)?
2. On the airport map, could you please mark the flow through the day, discussing the key interaction points?
3. What would you identify as the most important function the App should include? Why?
4. Which points of interaction with passengers and other stakeholders are currently most challenging? Why?
5. What are the key opportunities and challenges around touchless interfaces that the FlyKey platform could address? Why?
6. Which other stakeholders would need to be included within the FlyKey platform? Why?
7. What would be required from your perspective in order to make the FlyKey platform more inclusive and accessible to the airport environment and staff needs (i.e., The Americans with Disabilities Act (ADA), etc.)?
8. How did you find using the App to complete the task?
9. What was particularly challenging? Why?
10. What did you enjoy the most? Why?
11. What do you think could be improved?
12. Are there any other features you would like to see included in the App?

The findings from the open-ended interviews identified several themes in the data relating to the App usability and improvements. For the research team to constitute a theme, it had to be mentioned by at least three participants. The themes are listed below:

3.2.1. App Usability:

- Ease of access
- Good layout of the screens
- Clarity of navigation and the App process
- UI is user-friendly
- Saves time for passengers

3.3. Improvements:

- Desire to see the actual App versus prototype to get a better understanding of functionality (and improvements needed).
- Add an airport layout map.
- Add navigation of the walking path to your gate.
- Add a back button (undo).
- Add a menu button.
- Add a Live Chat (help assistance).
- Add a three-step verification process for logging in from a security standpoint.
- See the App interconnected with their Calendar and other Apps.
- Ability to view multiple flights that are registered for within the App.
- Link the App to Airport parking.
- Sharing permissions to be a feature to share information with others that a passenger agrees to share (i.e., such as a family member picking up a passenger once they land, and to be informed of flight delays).
- Add weather tourism destination and ground transportation for each flight itinerary.
- An intro application onboarding video on how to use the 'self-identity' functionality would be helpful.
- 'Traveling as a group' option - for parents to manage children's passports and travel information.
- Information about data ownership and which customer data is shared would be helpful, including potential General Data Protection Regulation (GDPR) related data sharing approval, such as the sharing of cookies. This could be part of the 'App privacy settings' or 'about' sections.
- 'Past flights' and 'Upcoming flights' sections.
- Airport map with gate directions.
- Possibility to amend the selected time slot.

3.4. Usefulness, Satisfaction, and Ease of Use

The 14 participants were further asked the partial questions from the Usefulness, Satisfaction, and Ease of Use (USE) questions developed by Lund (2001). Due to participants' limited time in passenger waiting areas at airports, the research team limited the questions to only those assessing usefulness and ease of use. The key findings from the USE questionnaire are summarized in Table 2.

Table 2. Summary of Findings from the USE Questionnaire

Scale Response	**Summary/Percentages** **Strongly Disagree = 1 to Strongly Agree = 7**
USEFULNESS	
It helps me be more effective	7.1% rated a 4, 21.4% rated a 5, 42.9% rated a 6, 21.4% rated a 7, and 7.1% listed a N/A
It helps me be more productive	35.7% rated a 5, 42.9% rated a 6, 14.3% rated a 7, and 7.1% listed a N/A
It is useful	7.1% rated a 5, 28.6% rated a 6, and 64.3% rated a 7
It gives me more control over the activities in my life	7.1% rated a 3, 28.6% rated a 5, 42.9% rated a 6, and14.3% rated a 7
It makes the things I want to accomplish easier to get done	7.1% rated a 5, 57.1% rated a 6, and 35.7% rated a 7
It saves me time when I use it	7.1% rated a 5, 28.6% rated a 6, 64.3% and rated a 7
It meets my needs	28.6% rated a 5, 42.9% rated a 6, and 28.6% rated a 7
It does everything I would expect it to do	7.1% rated a 4, 50% rated a 5, 28.6% rated a 6, and 14.3% rated a 7
EASE OF USE	
It is easy to use	7.1% rated a 4, 24.3% rated a 5, 14.3% rated a 6, and 64.3% rated a 7
It is simple to use	7.1% rated a 4, 21.4% rated a 5, 14.3% rated a 6, and 57.1% rated a 7
It is user friendly	7.1% rated a 4, 35.7% rated a 6, and 57.1% rated a 7
It requires the fewest steps possible to accomplish what I want to do with it	28.6% rated a 5, 28.6% rated a 6, 35.7% rated a 7, and 7.1% listed a N/A
It is flexible	7.1% rated a 4, 28.6% rated a 5, 28.6% rated a 6, 28.6% rated a 7, and 7.1% listed N/A
Using it is effortless	14.3% rated a 4, 28.6% rated a 5, 28.6% rated a 6, and 28.6% rated a 7
I can use it without written instructions	14.3% rated a 3, 28.6% rated a 5, 35.7% rated a 6, and 21.4% rated a 7

4. Discussion

The aim of the research was to gather usability feedback on the proposed systemic solution to inform the next stages of product iteration and development, so that the FlyKey App enhances the passenger experience. The research team conducted a usability study consisting of think-aloud sessions, interviews, and USE questions. The research gathered critical information to enhance the prototype design of the FlyKey Passenger App to make it more user-friendly but also provided additional insights for the Personnel App to make navigation easier. For instance, providing additional feedback to users, whether passengers or airport personnel, is beneficial for both sides of the App.

Future research to expand upon the present study is to quantify the benefits of using the FlyKey Passenger App and Flykey Personnel App. Therefore, studies using data on operational efficiency gains, such as whether the App reduces airport congestion and check-in times with fewer staff interventions, would be useful. Studying different passenger demographics and their perceptions of the App could also help researchers determine whether frequent or inexperienced travelers prefer the App. Furthermore, studying trust in the App by users and how to enhance trust, or rather have users experience calibrated trust, where the trust users have is equal to the reliability of the App, is another research area to explore. Future research also includes studies to identify stakeholder requirements for a central FlyKey-Analytics platform, including usability enhancements. This includes an expanded analysis of the needs of all the stakeholders (i.e., airport board, third-party concessionaires, and airport personnel). Therefore, interviews and focus groups will be conducted to understand the needs that would enhance airports' ability to manage operations in a safer, more efficient environment. Overall, the world is increasingly becoming more digital. Apps that can enhance safety or efficiency at busy places, such as airports, could improve the passenger experience and possibly the airport personnel's ability to further enhance it.

5. Disclaimer Statements

Funding: None

Conflict of Interest: None

6. References

Iqbal, M. Z., & Campbell, A. (2020). The emerging need for touchless interaction technologies. *Interactions*, 27(4), 51-52.

Lee, H., & Lee, K. (2025). Usability of airport self-service kiosks with multimodal AI. Archives of Design Research, 38(4), 213–228. https://doi.org/10.15187/adr.2025.11.38.4.213

Lund, A.M. (2001). *Measuring Usability with the USE Questionnaire*. STC Usability SIG Newsletter, 8:2.

McDonald, S., Edwards, H. M., & Zhao, T. (2012). Exploring think-alouds in usability testing: An international survey. IEEE Transactions on Professional Communication, 55(1), 2–19. https://doi.org/10.1109/TPC.2011.2182569

Nair, A., Wojdecka, A., Abbas, S., & Carstens, D. (2021). FlyKey [White Paper]. Florida Institute of Technology. https://vsgc.odu.edu/acrpdesigncompetition/wp-content/uploads/sites/3/2021/06/2021-ACRP-Design-Competition_1st_Environmental.pdf

Saadati, P., Abdelnour-Nocera, J., & Clemmensen, T. (2022). Co-design workshops as a step towards pilot implementation for complex workplaces: A case study of London-based airport future workplace. In G. Bhutkar, A. Joshi, & M. Rautaray (Eds.), Sense, Feel, Design. INTERACT 2021. Lecture Notes in Computer Science (Vol. 13198, pp. 421–428). Springer. https://doi.org/10.1007/978-3-030-98388-8_38

Shafitri, K., & Sofica, V. (2023). *Analisis usability aplikasi transportasi online dengan metode USE questionnaire dan usability testing*. Informatics for Educators and Professionals: Journal of Informatics, 8(2). https://doi.org/10.51211/itbi.v8i2.2534

Stone, B., & Wang, Y. (2011). AirportLogic: Usability testing, prototyping, and analysis of an airport wayfinding application. In C. Stephanidis (Ed.), *HCI International 2011 – Posters' Extended Abstracts. HCI 2011. Communications in Computer and Information Science* (Vol. 174, pp. 81–84). Springer. https://doi.org/10.1007/978-3-642-22095-1_17

The Journal of Management and Engineering Integration (JMEI) © 2026 by AIEMS Inc., is licensed under CC BY 4.0. To view a copy of this license, visit https://creativecommons.org/licenses/by/4.0/

The Journal of Management and Engineering Integration Vol. 19, No. 2 | Winter 2026 DOI: 10.62704/10057/59056

A SYSTEMATIC LITERATURE REVIEW OF INDUSTRY 5.0 CONCEPTS IN BOTTLENECK COLLEGE STEM COURSES

Faisal S. Cheema [1]

Ahmad Elshennawy [1]

[1] University of Central Florida

fa206186@ucf.edu

Abstract

With its human-centered and adaptable technologies, Industry 5.0 offers a framework to overcome ongoing barriers to student success in gateway STEM courses. Concepts such as real-time feedback, gamification, and adaptive learning can effectively simplify complex concepts to students and potentially improve student proficiency. While the implementation of such techniques seems inevitable, the lack of faculty support represents one of the barriers that must be addressed to successful integration. As such, this review aims to investigate the barriers that prevent instructors from successfully implementing them.

Keywords: *Industry 5.0, STEM Education, Human-Centered Learning, Personalized Learning, Adaptive Learning Systems, Real-Time Feedback, Gamification, Bottleneck STEM Courses*

1. Introduction

STEM (Science, Technology, Engineering, and Math) courses in the United States currently require students to undergo rigorous coursework. Students are often disinterested in science and math when they learn concepts in isolation, missing connections to related concepts and real-world applications (Kelley et al., 2016). Industry 5.0 is regarded as the latest stage of the Industrial Revolution, emphasizing the mass personalization of learning materials (Maddikunta et al., 2022). It builds on the advancements of Industry 4.0, emphasizing the integration of human capabilities and advanced technologies (Hashim et al., 2021).

Compared to Industry 4.0, Industry 5.0 is more human-centered and places greater emphasis on sustainability and resilience through the integration of technology (Supriya et al., 2024). The goal is to combine the strengths of human workers, characterized by adaptability, problem-solving abilities, and emotional intelligence. Industry 5.0 shifts educational technology toward human-centered learning, where intelligent systems support students and instructors rather than simply automating instructional processes (Adel, 2024). In STEM education, Industry 5.0 combines modern technologies and human creativity to create adaptive, personalized and low latency learning to help improve student proficiency. These principles may support personalized learning by tailoring learning experiences to individual students' needs and therefore improve students' proficiency in STEM courses. (Hashim et al., 2021).

Additionally, this review aims to investigate bottleneck courses for STEM majors. This study is a systematic literature review that examines the current Industry 5.0 frameworks adopted in STEM education. The successful implementation of Industry 5.0 concepts, such as Personalization and Gamification, can create a human-computer combination that benefits both students and instructors.

Submitted: April 30, 2026
Revised: October 20, 2026

This review aims to answer the following research questions:

1. What are the reasons for failures in STEM education?
2. What are the bottleneck courses for STEM majors?
3. What is the role of Industry 5.0 in STEM education?
4. What is the impact of Industry 5.0 techniques on STEM students' performance?
5. How does industry 5.0 in STEM education build on ideas from Industry 4.0?

2. Use of Personalization, Real-Time Feedback and Gamification in the Bottleneck Courses

In STEM education, Industry 5.0 prioritizes students' individual learning needs and supports instructors in adapting their teaching. Personalization, real-time feedback and gamification offer complementary ways to support this approach. Personalization uses adaptive learning platforms and tailored instructional strategies to address differences in students' knowledge, progress, and learning needs. Real-time feedback helps students identify errors and revise their understanding while working through complex material. Gamification may strengthen motivation and engagement by incorporating game elements into learning activities. Together, these approaches may help instructors provide responsive support in bottleneck STEM courses. Their contribution to sustainability and resilience, however, require explicit consideration and evaluation.

3. Building on Industry 4.0

The principles of personalization, real-time feedback and gamification have been a part of Industry 4.0 (Bartolomé et al., 2018; Peng et al., 2019). In Industry 5.0, the implementation of these principles has a greater human element by promoting a stronger connection between technology, instructors and students. Rather than focusing primarily on technological efficiencies, Industry 5.0 in STEM education prioritizes how these adaptive learning platforms can complement the individual needs of the students and how professors can use the data in real time to help improve proficiency among students. Within an Industry 5.0-oriented approach, instructors can use this information to provide individualized guidance, identify underlying misconceptions, and adjust teaching strategies based on students' academic and personal learning needs.

4. Methodology/Metrics

Google Scholar and the Web of Science were the two databases used for this review. Several keywords were used to search the databases for the research, including "Industry 5.0", "Industry 5.0 in STEM education", "Bottleneck courses in STEM education", "Personalized of STEM content", "Gamification of STEM content in college", "Reasons for failures in STEM courses", "Assessment techniques in industry 5.0 in STEM education" and "How to help students in STEM courses". 33 articles out of a total of 610 articles were selected based on the following criteria:

1. Articles must be published on or after January 1, 2014.
2. Articles must be peer reviewed.
3. Articles must discuss how Industry 5.0 concepts can be linked to STEM education.
4. Articles must be written in English.
5. Articles must discuss one of the following topics:
    a. How Industry 5.0 adaptive and personalized systems can impact STEM learning for college students.
    b. Bottleneck courses in STEM education and the reasons for students' poor performance in these courses.
    c. Concepts of gamification and real-time feedback, and how providing real-time feedback and gamifying the STEM content can help struggling students.

d. Ideas to help students thrive in STEM education and how Industry 5.0 concepts can help.

It is worth noting that Industry 5.0 emerged around 2020 to 2021 and the principles of personalization, real-time feedback and gamification were part of Industry 4.0 literature. It was important to understand how these principles would be implemented differently under Industry 5.0 where the ideas of human-centric approach, resilience and sustainability are encouraged.

4.1. Relevance to Research Questions

Articles discussed Industry 5.0, STEM education, and failure rates in STEM courses, as well as personalizing STEM content. The articles were screened based on their titles, abstracts, and conclusions. Each article had to address one of the following concepts: personalization of STEM education, gamification in STEM Education, real-time feedback in STEM education, or causes of failure among STEM students. Additionally, non-peer-reviewed content was ignored. Based on the above content, gaps in the current literature were identified to inform the development of a platform and future research topics.

4.2. Data Extraction

Data was extracted from each article by recording the authors' names, the study's aim, key findings related to topics such as personalized adaptive learning, gamification, and real-time feedback, and the conclusion drawn from each article.

4.3. Inclusion and Exclusion Criteria

For initial screening, the titles and abstracts were reviewed to eliminate irrelevant studies. The articles that did not have either an abstract, an introduction, or a conclusion were disqualified.

For the next part of screening, articles were narrowed down after reviewing the title, abstract, and conclusion. Any articles that did not directly fit into content linked to Industry 5.0 in STEM education or the reasons for failure among STEM students were discarded. Other reasons for exclusion include the following:

1. Published before 2014.
2. Not in English.
3. Did not include human participants.
4. Were not peer reviewed or were conference proceedings.

For complete screening, a full text for each article was retrieved. The extracted findings were then grouped into the following categories: Student Engagement through Gamification, Adaptive learning, Real-time feedback, bottleneck courses for STEM students, and reasons for failure. To assess the overall risk of bias in the included studies, the following three aspects were looked at:

1. Bias from the Randomization Process
2. Bias Due to Deviations from Intended Principles linked to Industry 5.0 in STEM education
3. Bias Due to Missing Outcomes

To test the heterogeneity, the characteristics (population, intervention, outcomes, methods) and testing patterns (subgroup analysis, meta-regression, sensitivity analysis) were determined for each article. To understand potential sources of heterogeneity in study results, the population, interventions, and outcomes were examined. This literature review aims to understand the use of Industry 5.0 concepts, such as personalization, to help students struggling in the gateway STEM courses. Any articles showing a high degree of bias or significant heterogeneity were excluded from the review.

4.4. Operational Definition of Industry 5.0

For this Literature Review Industry 5.0 is defined as human centric technological approaches to help improve student proficiencies. These include Adaptive learning systems, intelligent personalized learning tools, instructor led personalization, low latency learning through real time feedback and gamification of STEM content.

5. Results

The articles were selected according to the PRISMA flow. Table 1 shows the number of articles selected by category and year.

Table 1. Number of Articles by Year

Article Year	Number of Articles
2014	2
2015	1
2016	2
2017	0
2018	2
2019	2
2020	4
2021	4
2022	4
2023	2
2024	6
2025	2
2026	2

5.1. Literature Review

The review revealed that students are vulnerable in the initial years of their academic programs. It was found that STEM students' relationship with their math courses is one of the leading factors for both student success and the choice to stay in the STEM field (Gandhi-Lee et al., 2015).

Students are often disinterested in science and math when concepts are taught in isolation, with little connection to the real world (Kelley et al., 2016). Freeman et al. (2014) support this idea by showing that lecture-heavy teaching is associated with poorer STEM performance. Hashim et al. (2021) also state that student motivation linked to interest and learning habits is a major hurdle for STEM students. Allen et al. (2016) claim that students struggle when teaching and learning are adaptive, and they struggle even more when feedback is nonexistent or delayed. The suggestion was to ensure that feedback is quick and actionable so students can quickly recognize their errors and fix them. Cviko et al. (2014) claim that teachers lack the tools and training to integrate the latest technology into their lessons.

5.2. Math is One of the Biggest Barriers STEM Students Face

It was found that STEM students' relationship with their math courses is one of the leading factors for both student success and the choice to stay in the STEM field (Gandhi-Lee et al., 2015). Students are often disinterested in science and math when concepts are taught in isolation,

missing connections to real-world applications (Kelley et al., 2016). Gandhi-Lee et al. (2015) also state that student success is greatly linked to mathematical orientation.

Wester et al. (2021) found that STEM undergraduates' engagement declined during COVID-19-driven remote learning, and students learn better when they are in person. This means that purely online learning may not be the best way to teach Calculus and other math-heavy classes. Allen et al. (2016) reported a lack of belonging among STEM students and limited academic support during the initial years of college STEM study.

5.3. Personalization and Industry 5.0

The studies reveal that concepts of Industry 5.0, which advocate personalizing content in a human centric way, can greatly help struggling STEM students. Ahmad et al., 2023 focuses on a human-centric, personalized learning approach towards STEM education. Industry 5.0 offers more opportunities for data analysis and there is an increased need to create personalized learning pathways (Adel, 2022).

Industry 5.0 offers opportunities for both students and teachers to think critically and creatively. Therefore, the idea of a dynamic assessment system and the AI tools currently used to provide instant feedback to students and teachers were explored. Adaptive learning can provide personalized learning, adaptive assessments, and intelligent feedback. This could lead to improved learning outcomes (Saniuk et al., 2024).

Personalization of learning is considered one of the themes when utilizing Industry 5.0 in STEM education (Barkoczi et al., 2024). This can be done with advanced technologies such as artificial intelligence to create a human-centric synergy between humans and machines. Maddikunta et al. (2022) support the idea by describing Industry 5.0 as an opportunity to enable mass personalization through advanced technologies such as AI, data analytics and a strong human connection. Project-based personalized learning is highly impactful, but more studies are needed that focus primarily on bottleneck courses. Lin et al. (2024) showed that personalization through a more human centric approach can help with the motivation issues faced by STEM students by increasing learning interest and enhancing proficiency. Providing individualized instruction has requires great amount of instructor time, the reviewed literature suggests that with the use of artificial intelligence the adaptive learning platforms and real time feedback offer potential mechanisms for addressing these challenges (Supriya et al., 2024).

5.4. Low Latency Learning Through Real-Time Feedback and Gamification

There are different ways in which Industry 5.0 concepts can be used to help struggling STEM students. The use of real-time feedback in STEM courses can make learning both adaptive and engaging (Kim et al., 2024). Real-time technology enables educators to analyze data and interpret results through results-based decision-making (Supriya et al., 2024). These changes can adapt to students' needs and provide real-time feedback closely related to industry needs.

Real-time feedback mechanisms can personalize instruction and enhance engagement in STEM contexts (Meacham et al., 2020). Real-time feedback can also boost motivation, especially for struggling STEM students, and improve proficiency (Troussas et al., 2020). Meacham et al. (2020) also showed that adaptive systems personalize instruction in real time while also supporting personalization in online STEM environments. Villegas-Ch et al. (2025) reveal that adaptive tutoring systems provide personalized and real-time feedback that can significantly improve student proficiency.

Gamification enhanced student engagement and improved conceptual understanding in

students' formative years, leading to higher achievement scores (Gui et al., 2023). Gamification and personalization may help these students become more engaged in their work and apply these ideas in subsequent courses (Troussas et al., 2020). Wang et al. (2022) claimed that digital game-based STEM learning improves student achievements. There is a special emphasis in the literature on real-time feedback, Personalization, and adaptive learning through Gamification (Chen, 2020, and Gui et al., 2023). Gamification was also linked to higher achievement scores (Wang et al., 2022). The main theme of the articles was that Industry 5.0 enables stakeholders in the education sector to ensure efficiency and effectiveness in teaching and learning (Supriya et al., 2024).

5.5. Challenges in Implementing Industry 5.0 Principles and Future Direction

The issue of overcrowding class sizes, where students are not receiving personalized attention, is a major concern, as the current approach is not scalable. One of the most significant barriers to implementing Industry 5.0 in STEM education is the time required of faculty to design and monitor Industry 5.0 systems, such as Adaptive learning and real-time feedback. Implementing Industry 5.0 ideas can be time-consuming and may put additional pressure on instructors. Saniuk et al. (2024) claimed that faculty workload and insufficient preparation time are two major reasons why Industry 5.0 adoption poses a major hurdle. Troussas et al. (2020) found that instructors often resist adaptive systems due to unfamiliarity and lack of training.

Without institutional support and workload adjustments, many instructors view Industry 5.0 adoption as impractical despite its proven educational benefits. Saniuk et al. (2024) emphasized that institutions struggle with outdated systems and limited access to advanced educational technologies, underscoring the need to invest in robust infrastructure. Developing adaptive learning systems along with a reliable infrastructure and training the instructors on how to best use these systems creates several challenges. Additionally, it may be difficult to convince the instructor to abandon some of the traditional teaching methods and replace them with the Industry 5.0 principles.

6. Conclusion and Future Research

The reviewed literature identifies several challenges in STEM education including large classes and teaching and learning of complex content. Personalization, Real Time feedback and gamification may help address these challenges and support students in bottleneck STEM courses. Integrating these approaches within an Industry 5.0 framework offers a potential pathway for more human centered, sustainable, and resilient learning environment. However, further empirical evidence is needed to establish the benefits of this integration.

Future research should examine how these instructional approaches can translate Industry 5.0 principles into educational practice. Priorities include evaluating how adaptive learning platforms respond to individual student needs and how their design support sustainability and resilience. Rather than focusing mainly on the development of adaptive learning technologies, future researchers should examine how artificial intelligence and instructor expertise can complement one another to provide personalized learning experiences and help students improve proficiency in the bottleneck STEM courses.

Research should also access the extent of adoption in bottleneck STEM courses and investigate barriers to implementation, including instructor training, workload, institutional support and technological infrastructure. Finally comparative empirical studies should determine whether learning environments explicitly designed around Industry 5.0 principles improve student outcomes

relative to traditional approaches.

7. Limitations

Some studies included in the review had convenience and purposive sampling biases, but they were still included because the topic was extremely relevant to the concepts of Industry 5.0. These databases were last searched on March 26, 2026. D.G., Kaynak, O., & Sait, S.M. (2021); Cviko, A., McKenney, S., & Voogt, J. (2014); and Gui et al. (2023) reported moderate population-related heterogeneity.

8. Disclaimer Statement

Funding: None

Conflict of Interest: None

9. References

Adel, A. Future of Industry 5.0 in Society: Human-Centric Solutions, Challenges, and Prospective Research Areas. *J Cloud Comp* 11, 40 (2022). https://doi.org/10.1186/s13677-022-00314-5

Ahmad, S., Umirzakova, S., Mujtaba, G., Amin, M. S., & Whangbo, T. (2023). *Education 5.0: Requirements, Enabling Technologies, and Future Directions. arXiv* preprint arXiv:2307.15846. https://doi.org/10.48550/arXiv.2307.15846

Allen, M., Webb, A. W., & Matthews, C. E. (2016). Adaptive Teaching in STEM: Characteristics for Effectiveness. *Theory Into Practice*, 55(3), 217–224. https://doi.org/10.1080/00405841.2016.1173994

Ang, H. Q., & Domingo, B. (2025). An innovative and universal teaching model in engineering education to foster Industry 5.0 skills—EPIC principles. *Education Sciences, 15*(9), 1096. https://doi.org/10.3390/educsci15091096

Bahr, P. R. (2019). The labor market returns to a community college education for non-completing students. *The Journal of Higher Education*, *90*(2), 210-243. https://doi.org/10.3102/01623737251360029

Bartolomé, A., Castañeda, L., & Adell, J. (2018). Personalization in Educational Technology: The Absence of Underlying Pedagogies. *International Journal of Educational Technology in Higher Education, 15*, Article 14. https://doi.org/10.1186/s41239-018-0095-0

Broo, D.G., Kaynak, O., & Sait, S.M. (2021). Rethinking engineering education at the age of Industry 5.0. *J. Ind. Inf. Integr., 25*, 100311. https://doi.org/10.1016/j.jii.2021.100311

Chen PH (2020). The design of applying gamification in an immersive virtual reality laboratory for powder-bed binder jetting 3DP training. *Educ Sci* 10(7):172. https://doi.org/10.3390/educsci10070172

Cviko, A., McKenney, S., & Voogt, J. (2014). Teachers as co-designers of technology-rich learning activities for early literacy. *Technology, Pedagogy and Education*, 24(4), 443–459. https://doi.org/10.1080/1475939X.2014.953197

D. Zhao *et al.*, "An Innovative Multi-Layer Gamification Framework for Improved STEM Learning Experience," in *IEEE Access*, vol. 10, pp. 3879-3889, 2022. https://doi.org/10.1109/ACCESS.2021.3139729

Freeman, S., Eddy, S. L., McDonough, M., Smith, M. K., Okoroafor, N., Jordt, H., & Wenderoth, M. P. (2014). Active learning increases student performance in science, engineering, and mathematics. *Proceedings of the National Academy of Sciences, 111*(23), 8410–8415. https://doi.org/10.1073/pnas.1319030111

Gandhi-Lee, E., Skaza, H., Marti, E., Schrader, P. G., & Orgill, M. (2015). Faculty perceptions of the factors influencing success in STEM fields. *Journal of Research in STEM Education*, *1*(1), 30-44. https://doi.org/10.51355/jstem.2015.7

Gui, Y., Cai, Z., Yang, Y. *et al.* Effectiveness of digital educational games and game design in STEM learning: a meta-analytic review. *IJ STEM Ed* 10, 36 (2023). https://doi.org/10.1186/s40594-023-00424-9

Hashim, Suhaizal, et al. "Association of interest, attitude, and learning habit in mathematics learning towards enhancing students' achievement." *Indonesian Journal of Science and Technology* 6.1 (2021): 113-122. https://doi.org/10.17509/ijost.v6i1.31526

Hurley, M., Butler, D., & McLoughlin, E. (2024). STEM teacher professional learning through immersive STEM learning placements in industry: A systematic literature review. *Journal for STEM Education Research*, 7, 122–152. https://doi.org/10.1007/s41979-023-00089-7

Karimi, H., & Pina, A. (2021). Strategically addressing the soft skills gap among STEM undergraduates. *Journal of Research in STEM Education*, 7(1), 21–46. https://doi.org/10.51355/jstem.2021.9

Kelley, T.R., Knowles, J.G. A conceptual framework for integrated STEM education. *IJ STEM Ed* 3, 11 (2016). https://doi.org/10.1186/s40594-016-0046-z

Kim, J. (2024). Leading teachers' perspective on teacher-AI collaboration in education. *Education and Information Technologies*, 29(7), 8693–8724. https://doi.org/10.1007/s10639-023-12109-5

Klieger, A. (2026). Aligning STEM education with industry needs: Strategies to bridge the skills gap. *Industry and Higher Education*, *40*(5), 884-896. https://doi.org/10.1177/09504222261418077

Lai, C.-S. (2018). Using inquiry-based strategies for enhancing students' STEM education learning. *Journal of Education in Science, Environment and Health*, 4(1), 110–117. https://doi.org/10.21891/jeseh.389740

Lin, L., Lin, X., Zhang, X. *et al.* The Personalized Learning by Interest Effect on Interest, Cognitive Load, Retention, and Transfer: A Meta-Analysis. *Educ Psychol Rev* 36, 88 (2024). https://doi.org/10.1007/s10648-024-09933-7

Maddikunta, P. K. R., Pham, Q.-V., Prabadevi, B., Deepa, N., Dev, K., Gadekallu, T. R., Ruby, R., & Liyanage, M. (2022). Industry 5.0: A survey on enabling technologies and potential applications. *Journal of Industrial Information Integration, 26*, Article 100257. https://doi.org/10.1016/j.jii.2021.100257

Meacham S, Pech V, Nauck D (2020). AdaptiveVLE: An Integrated Framework for Personalized Online Education Using MPS JetBrains Domain-Specific Modeling Environment. In: IEEE Access, (8):184621-184632. https://doi.org/10.1109/ACCESS.2020.3029888

Peng H., Ma S., Spector J.M. Personalized adaptive learning: an emerging pedagogical approach enabled by a smart learning environment. *Smart Learning Environments*. 2019;6. https://doi.org/10.1186/s40561-019-0089-y

Shah, R., & Gillen, A. L. (*2024). A systematic literature review of university-industry partnerships in engineering education. European Journal of Engineering Education*, 49(3), 577–603. https://doi.org/10.1080/03043797.2023.2253741

Supriya, Y., Bhulakshmi, D., Bhattacharya, S., Gadekallu, T. R., Vyas, P., Kaluri, R., Sumathy, S., Koppu, S., Brown, D. J., & Mahmud, M. (2024). Industry 5.0 in smart education: Concepts, applications, challenges, opportunities, and future directions. *IEEE Access, 12*, 81938–81967. https://doi.org/10.1109/ACCESS.2024.3401473

Saniuk, S., Grabowska, S., & Thibbotuwawa, A. (2024). Challenges of industrial systems in terms of the crucial role of humans in the Industry 5.0 environment. Production Engineering Archives, 30(1), 94–104. https://doi.org/10.30657/pea.2024.30.9

Troussas, C., Krouska, A., & Sgouropoulou, C. (2020). Collaboration and fuzzy-modeled personalization for mobile game-based learning in higher education. *Computers & Education, 144*, Article 103698. https://doi.org/10.1016/j.compedu.2019.103698

Urrea, C. (2026). Industrial robotics and adaptive control systems in STEM education: Systematic review of technology transfer from industry to classroom and competency development framework. Applied Sciences, 16(4), Article 2026. https://doi.org/10.3390/app16042026

Villegas-Ch, W., Buenano-Fernandez, D., Maldonado Navarro, A., & Mera-Navarrete, A. (2025). Adaptive intelligent tutoring systems for STEM education: Analysis of the learning impact and effectiveness of personalized feedback. *Smart Learning Environments, 12*, Article 41. https://doi.org/10.1186/s40561-025-00389-y

Wang, L.-H., Chen, B., Hwang, G.-J., Guan, J.-Q., & Wang, Y.-Q. (2022). Effects of digital game-based STEM education on students' learning achievement: A meta-analysis. *International Journal of STEM Education, 9*, Article 26. https://doi.org/10.1186/s40594-022-00344-0

Wester, E. R., Walsh, L. L., Arango-Caro, S., & Callis-Duehl, K. L. (2021). Student engagement declines in STEM undergraduates during COVID-19–driven remote learning. *Journal of Microbiology & Biology Education, 22*(1). https://doi.org/10.1128/jmbe.v22i1.2385

Zhai, X., Haudek, K. C., Shi, L., Nehm, R. H., & Urban-Lurain, M. (2020). From substitution to redefinition: A framework of machine learning-based science assessment. *Journal of Research in Science Teaching*, 57(9), 1430–1459. https://doi.org/10.1002/tea.21658

The Journal of Management and Engineering Integration (JMEI) © 2026 by AIEMS Inc., is licensed under CC BY 4.0. To view a copy of this license, visit https://creativecommons.org/licenses/by/4.0/

The Journal of Management and Engineering Integration Vol. 19, No. 2 | Winter 2026 DOI: 10.62704/10057/59057

TRACEABILITY WITHOUT COMPLEXITY: A SYSTEMS ENGINEERING APPROACH TO MOTOR-PUMP CORRELATION

Rohan Dharmendrakumar Patel [1]

Joyce Chiam Ziyi [1]

Adam Carlton Lynch [1]

[1] Wichita State University

rxpatel11@shockers.wichita.edu

Abstract

In practice, manufacturers often test components beyond regulatory requirements to reduce the risk of downstream failures. In the studied facility, motors were tested prior to assembly, but the absence of traceability made it difficult to understand whether this additional testing contributed measurable value. This study focuses on addressing this gap by establishing a practical traceability mechanism between motor and pump testing. The work applies a systems engineering perspective to the design of a barcode-based traceability framework that fits within existing shop-floor constraints. An exploratory Excel-based model with embedded VBA logic was developed to generate motor serial numbers, capture pass/fail test outcomes, and link these records to pump testing through barcode scanning. Emphasis was placed on offline operation, low implementation cost, and ease of use for operators. The framework was implemented directly on the factory floor without disrupting production activities. Motor test results were consistently logged, while the system provided an initial mechanism for associating motor and downstream pump testing data and reducing reliance on manual documentation. The solution integrated with existing equipment and supported daily motor testing operations. Although this pilot study did not extend to full motor–pump performance correlation, the implemented system establishes an initial foundation for future analysis. The results suggest that modest, well-aligned digital interventions, when guided by systems engineering principles, can support traceability and process decisions in small and mid-sized manufacturing environments.

Keywords: *Systems Engineering, Traceability, Automation, Manufacturing, Process Optimization*

1. Introduction

In manufacturing, traceability across components is essential for quality control, diagnostics, and continuous improvement; however, it remains difficult to implement in resource-constrained environments due to reliance on manual data collection and disconnected testing processes. This study addresses this challenge by developing a practical, low-cost traceability framework using an Excel-based barcode system to link motor and pump test data within a manufacturing facility. By applying a systems engineering approach and using accessible tools, the proposed solution demonstrates how effective digital traceability can be achieved without the complexity of full-scale ERP or RFID systems (INCOSE, 2023).

1.1. Background

The manufacturer of retail products identified key inefficiencies in its motor testing process, specifically a lack of digital traceability. The existing process operated independently of

Submitted: April 30, 2026
Revised: October 21, 2026

downstream pump testing and lacked a structured mechanism for capturing and retaining test data at the unit level. Although motors were verified prior to assembly, the absence of serialized identification and data linkage limited the ability to trace component history across the production lifecycle. This disconnected testing approach constrained failure diagnostics and prevented systematic evaluation of relationships between motor performance and final pump outcomes, highlighting the need for an integrated yet practical traceability mechanism within existing manufacturing constraints (Womack & Jones, 1997; Bhamu & Sangwan, 2014).

1.2. Literature Review

Traceability and systems integration are essential in modern manufacturing for ensuring quality control, validating systems, and maintaining data integrity throughout the lifecycle (INCOSE, 2023). From a systems engineering viewpoint, establishing traceable relationships between components improves transparency, maintainability, and decision-making (Blanchard & Fabrycky, 2011). However, small and mid-sized manufacturers often lack cost-effective and scalable traceability solutions that match their operational constraints. The lean manufacturing approach also focuses on process efficiency, waste reduction, and continuous improvement in its production systems (Bhamu & Sangwan, 2014; Womack & Jones, 1997). Additionally, integrating information management within lean systems shows how crucial structured data flow and visibility are at various manufacturing stages (Bevilacqua et al., 2015). Despite these improvements, actual traceability implementation on the shop floor is still limited, especially in environments with limited resources and low digital infrastructure (Abad et al., 2009). This study aims to fill that gap by using systems engineering principles to create a low-cost, user-friendly solution that connects subcomponent testing to final product performance.

1.3. Problem Statement

At the manufacturer's facility, motors were tested independently of pumps with no system in place to track motor performance, assign serial numbers, or associate test results with final pump output. This lack of traceability limited quality analysis and made it difficult to evaluate whether motor testing added meaningful value to the process.

1.4. Research Question

- **RQ1**: What system design requirements allow traceability in manufacturer's motor testing process?
- **RQ2**: How can the system support correlation of motor-pump data to inform testing decisions?

2. Methodology

At the outset of the project, the manufacturer's motor testing process operated independently from pump testing and did not include a system for serializing motors or tracking individual test results. While the testing ensured motor functionality at the time of assembly, the absence of a traceability mechanism made it difficult to analyze long-term performance or associate motor behavior with pump outcomes. This created an opportunity to explore digital solutions that could enhance traceability, support quality monitoring, and enable more data-informed process decisions. To analyze and understand the current workflow, a UML Activity Diagram was developed to map the existing motor testing process (Figure 1). This diagram highlighted the sequential operator actions, decision points, and illustrated the absence of data logging of test results.

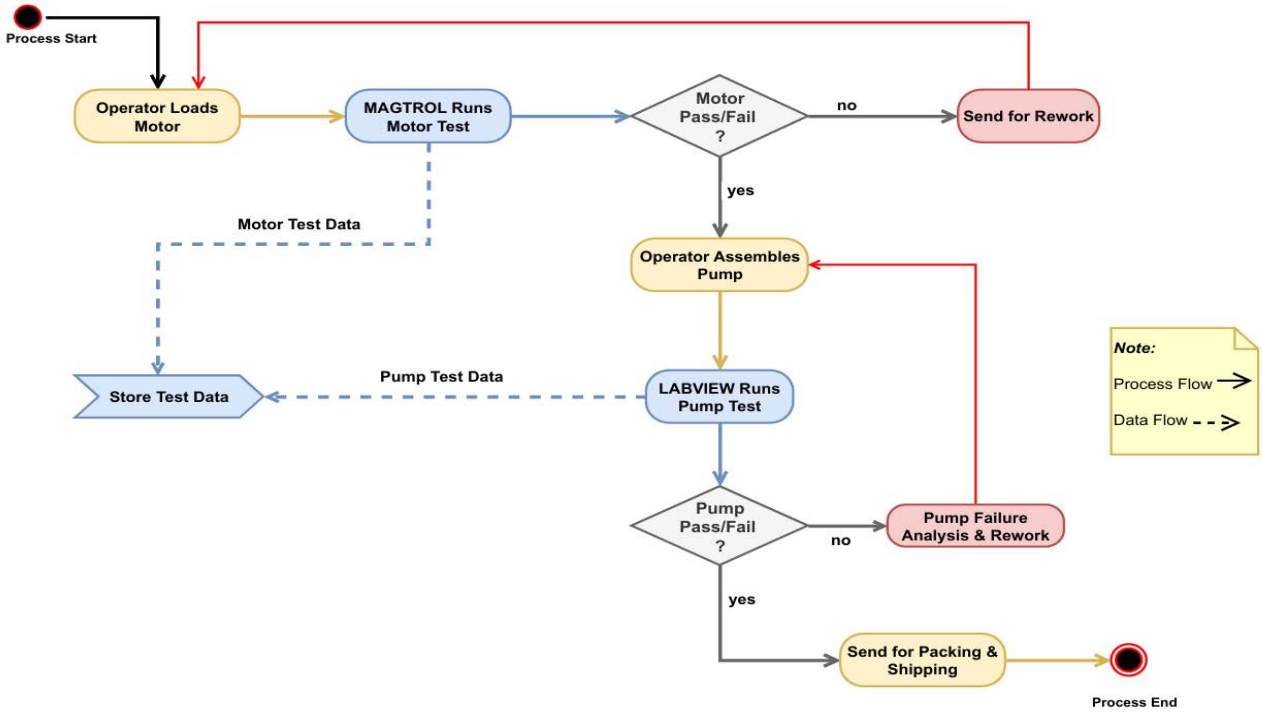


Figure 1. UML Activity Diagram

Based on stakeholder input and a systems engineering approach, a digital traceability solution was designed to address these gaps. The team developed a barcode-based tracking system, implemented using Microsoft Excel and Visual Basic for Applications (VBA). The tool, named R-Matrix, enabled operators to log motor pass/fail outcomes. For each passing motor, a timestamped serial number was automatically generated and printed on a 2x2 inch barcode label via a Zebra printer. Motors that did not pass the test were logged separately without a serial number. During pump testing, the barcode was scanned to associate motor and pump data digitally, enabling traceability.

To model the structure of the new system and its integration within manufacturer's workflow, a SysML Internal Block Diagram (Figure 2) was created. This diagram illustrates the interactions between hardware components (e.g., motor test station, barcode printer), software logic (e.g., serial generator), and operator inputs (INCOSE, 2023).

During the system design phase, Pareto analysis was first conducted to identify the most frequent and impactful issues affecting motor testing quality. This analysis guided the team in focusing on a core subset of challenges. Based on these prioritized issues, a Fishbone (Ishikawa) diagram was developed to explore root causes of traceability limitations and process inefficiencies. Key contributors were identified across six categories: environment, people, machines, methods, materials, and measurement systems. Notably, the absence of automated data logging and the lack of correlation between motor and pump testing data were central to the design requirements of the proposed solution (Womack & Jones, 1997).

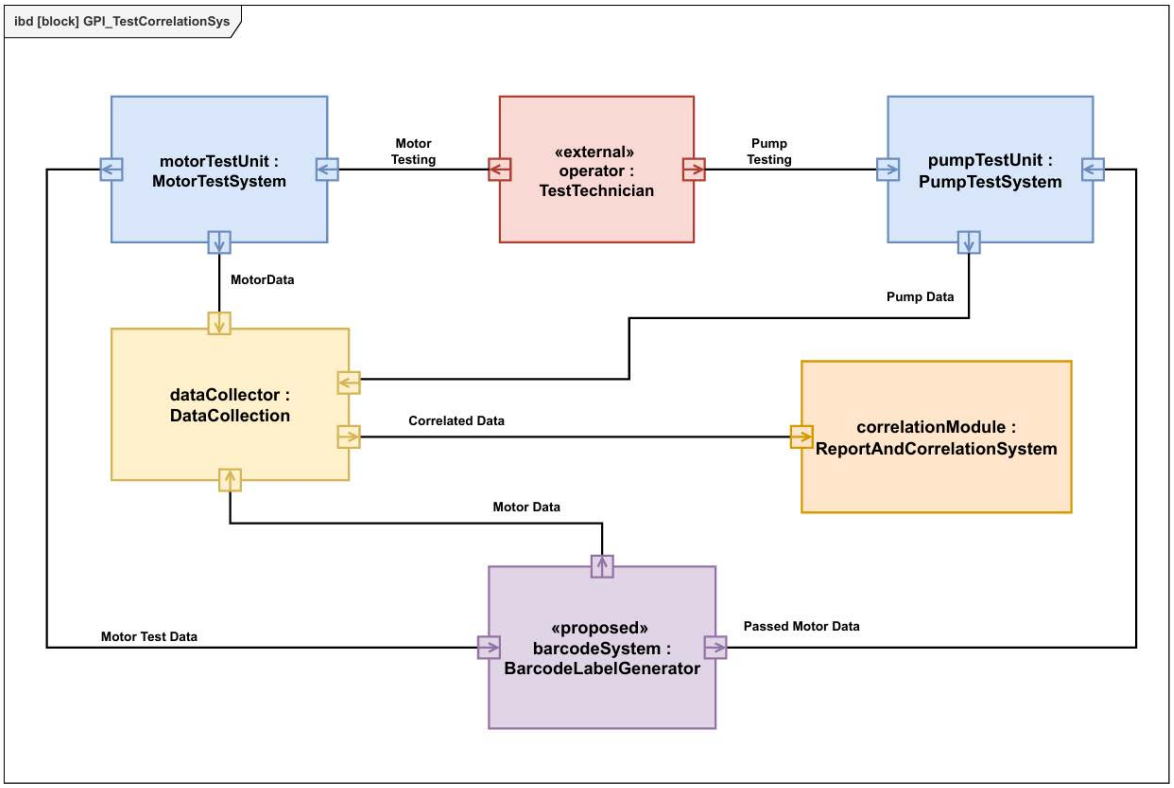


Figure 2. SysML Internal Block Diagram

Several alternative solutions were considered during the design phase to address traceability and testing data management, including RFID-based test stands, AI-based video tracking, and motor serial number pre-assignment. Each option was assessed based on feasibility, setup complexity, scalability, cost.

When considering methods to track motor testing at industry, consideration was given to advanced solutions like RFID and ERP systems. Quickly, it became clear that these were not practical options. RFID setups need specialized hardware, like scanners and tags, plus dedicated software, and IT support. ERP systems offer strong features, though are expensive, require long implementation times, and need people with specialized training (Table 1) (Liker, 2020).

Instead, the researchers selected a simpler and practical solution - Microsoft Excel with built-in VBA coding. Excel was familiar to the operators at the manufacturing facility, which meant there was no learning curve. With VBA, the researchers were able to create a barcode-based tracking system that worked offline, no cost, could be customized, and tested in real time. This approach gave the researchers needed flexibility, without overwhelming our team or the company's (Pande et al., 2007).

Table 1. KPI Matrix

KPI	RFID Tracking	AI Video	Pre-Assigned Serial	Barcode Tracking
Low Hardware Cost	❌	❌	✅	✅
Minimal Setup Complexity	❌	❌	✅	✅
Operator-Friendly	✅	❌	⚠️ (some training)	✅
Scalable for Future	✅	❌	⚠️	✅
Works Offline	✅	❌	✅	✅
Low Maintenance	⚠️ (RFID tags)	❌	✅	✅
High Data Accuracy	✅	⚠️ (camera error)	⚠️	⚠️
Traceability Throughout Tests	✅	⚠️	⚠️	✅

2.1. Materials Used

- Zebra ZD420 Barcode Printer for generating 2x2inch barcode labels with motor serial numbers.
- Dedicated PC near MAGTROL motor test stand for label generation and data logging.
- Microsoft Excel with VBA used to build the R-Matrix tool for serial number generation, pass/fail logging, and barcode printing.
- 2” x 2” Thermal Barcode Labels applied to motors after passing the test.
- Handheld barcode scanner used during pump testing for data retrieval and correlation.
- MAGTROL Motor test stand used for performance testing and integrated with the labeling process.

3. Results

The proposed barcode-based traceability system was successfully integrated into the manufacturer’s motor testing workflow using existing shop-floor tools. The system enables digital tracking of test outcomes, supports operator-friendly implementation, and lays the foundation for future motor–pump data correlation. This section addresses the two defined research questions by examining the system’s design requirements, technical implementation, and long-term potential.

3.1. Addressing RQ1: System Requirements and Design

The traceability system was built around five key requirements: offline operation, low-cost tools, ease of use, compatibility with existing test stations, and support for future scalability. These requirements guided the development and implementation of the Excel-VBA-based R-Matrix system on the shop floor.

The final system included the following key elements:

- A VBA-based serial number generator and logger
- A barcode label printing module
- An operator-facing worksheet for quick data entry
- Integration into existing MAGTROL motor testing workflow
- Use of a handheld barcode scanner during pump testing

Following the initial pilot, the R-Matrix system remained in use at the motor build line for approximately six months, supporting consistent recording of motor test outcomes. During

preliminary shop-floor testing, the logging and barcode printing cycle was observed to take less than 10 seconds in a limited number of trials.

The R-Matrix system was developed using VBA to automate serial number generation, test result logging, and barcode preparation. Figure 3 presents a selected code excerpt illustrating how passing motors are assigned unique identifiers and timestamps, while failed motors are recorded separately. The complete program includes additional functions supporting the overall workflow.

```
Sub PassMotor()
    Dim ws As Worksheet
    Set ws = ThisWorkbook.Sheets("Motor Test Log")

    Dim lastRow As Long
    lastRow = ws.Cells(ws.Rows.Count, 1).End(xlUp).Row + 1

    Dim serial As String
    serial = "[COMPANY_CODE]" & Format(Date, "yymmdd") & Format(lastRow - 1, "00")

    ws.Cells(lastRow, 1).Value = "Pass"
    ws.Cells(lastRow, 2).Value = Now
    ws.Cells(lastRow, 3).Value = serial
    ws.Cells(lastRow, 4).Value = "*" & serial & "*"
    ws.Cells(lastRow, 4).Font.Name = "Free 3 of 9 Extended"
    ws.Cells(lastRow, 4).Font.Size = 30

    ' Auto-print barcode (cell D)
    'ws.Range(ws.Cells(lastRow, 4), ws.Cells(lastRow, 4)).PrintOut Copies:=1
End Sub

Sub FailMotor()
    Dim ws As Worksheet
    Set ws = ThisWorkbook.Sheets("Motor Test Log")

    Dim lastRow As Long
    lastRow = ws.Cells(ws.Rows.Count, 1).End(xlUp).Row + 1

    ws.Cells(lastRow, 1).Value = "Fail"
    ws.Cells(lastRow, 2).Value = Now
End Sub
Sub PrintLastBarcode()
```

Figure 3. VBA Code Extract

3.2. Addressing RQ2: Enabling Future Correlation

The barcode system was designed to enable the manufacturer to associate each tested motor with its corresponding pump through scanning (Figure 4). While this phase did not involve full-scale data analysis, an initial framework for traceability has been established. Earlier analysis identified traceability gaps including lack of data logging, standardized labeling, and no connection between motor and pump performance data. The implemented system addresses these requirements by:

- Logging motor test outcomes
- Applying unique IDs to passing motors
- Capturing timestamps
- Supporting downstream identification through barcode scanning

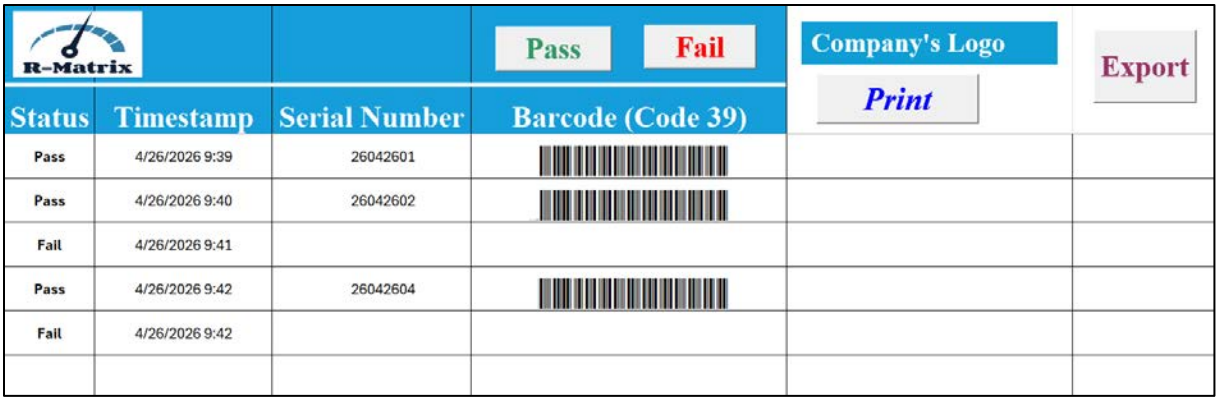


Figure 4. Developed Barcode Generation System

This system provides a foundation for future statistical analysis of motor-pump correlation. With improvements in barcode reliability and data consistency, the manufacturer could assess performance trends and identify opportunities to optimize testing processes through evidence-based decision making.

4. Discussion

4.1. Practical Impact of the Traceability System

The implementation demonstrated the feasibility of a low-cost, barcode-based traceability approach using existing manufacturing resources. The R-Matrix system supported motor test data logging within the established workflow and provided an initial foundation for downstream motor–pump data linkage. While barcode-based traceability is an established technology, the contribution of this study lies in its practical application through a systems engineering approach tailored to the manufacturer's operational requirements (Tortorella & Fogliatto, 2017).

4.2. Alignment with Systems Engineering Principles

The project followed a systems engineering approach by integrating stakeholder needs, functional requirements, and implementation constraints into the solution design. UML activity diagram and SysML internal blocks clarified the current process and the redesigned workflow. Iterative development and validation with users ensured successful integration. (Blanchard & Fabrycky, 2011).

4.3. Operational Limitations and Data Integrity Considerations

Although the proposed system demonstrated practical feasibility, several operational limitations and potential sources of data bias were identified during implementation:

- **Manual Data Entry:** Motor test outcomes were entered manually into R-Matrix, introducing potential recording errors and differences between actual testing and logging timestamps.
- **Barcode Reliability:** Occasional incomplete barcode printing and shop-floor contamination from grease and dust affected label readability and downstream scanning reliability.
- **Data Integrity:** A temporary dummy barcode was introduced to maintain production continuity during scanning difficulties. Its continued use made some records unsuitable

for reliable motor–pump correlation.

- **Scalability:** The Excel-based architecture lacks centralized database integration, potentially limiting data management and scalability as production requirements increase.

These limitations highlight the importance of reliable identification, standardized data recording, and improved validation procedures to support consistent traceability and future motor–pump correlation analysis (Pyzdek & Keller, 2014).

4.4. Future Opportunities and Scalability

Future improvements could focus on enhancing barcode reliability through shorter barcode formats, QR codes, improved printer alignment, and more durable labels suitable for shop-floor conditions. Improvements within R-Matrix could help reduce manual data entry errors and improve timestamp consistency. Integration with the existing LabVIEW pump testing program could also be explored to strengthen motor-pump data linkage and support future correlation analysis. Further integration with ERP/MES systems or centralized databases could enhance data management and long-term scalability.

5. Conclusion

This project addressed the absence of test data traceability in a motor and pump testing process. By leveraging systems engineering methods and operator-friendly tools like Excel and VBA, the team developed and implemented a barcode-based solution that logs motor test results and links them to pump performance data. The solution was successfully deployed within the company's existing workflow and remained in use at the motor build line for approximately six months. It met key functional goals of offline operation, real-time barcode generation, and integration with existing hardware. While large-scale correlation analysis was outside the project scope, the system now enables data-driven decisions and provides a path to long-term quality improvement (Fet & Haskins, 2023).

Beyond the technical setup, the project shows the importance of designing digital solutions that work with real-world challenges. These challenges include limited connectivity, operator workload, and non-standardized processes. Using visual modeling tools like UML and SysML helped facilitate communication among technical and non-technical stakeholders. This support allowed for iterative development and practical validation. The downstream scanning challenges highlighted the need for improved data reliability. Nevertheless, this work demonstrates how small, low-cost digital improvements, guided by systems thinking and engineering management principles, can support manufacturing traceability and provide a foundation for future modernization.

6. Disclaimer Statement

Funding: None

Conflict of Interest: None

7. References

Abad, E., Palacio, F., Nuin, M., De Zarate, A. G., Juarros, A., Gómez, J. M., & Marco, S. (2009). RFID smart tag for traceability and cold chain monitoring of foods: Demonstration in an intercontinental fresh fish logistic chain. *Journal of food engineering*, *93*(4), 394-399. https://doi.org/10.1016/j.jfoodeng.2009.02.004

Bhamu, J., & Singh Sangwan, K. (2014). Lean manufacturing: literature review and research issues. *International journal of operations & production management*, *34*(7), 876-940. https://doi.org/10.1108/IJOPM-08-2012-0315

Bevilacqua, M., Ciarapica, F. E., & Paciarotti, C. (2015). Implementing lean information management: the case study of an automotive company. *Production Planning & Control*, *26*(10), 753–768. https://doi.org/10.1080/09537287.2014.975167

Blanchard, B. S., & Fabrycky, W. J. (2011). Systems engineering and analysis. Pearson.

Fet, A. M., & Haskins, C. (2023). Systems engineering. In A. M. Fet (Ed.), Business transitions: A path to sustainability (Chapter 12). https://doi.org/10.1007/978-3-031-22245-0_12

INCOSE. (2023). INCOSE systems engineering handbook: A guide for system life cycle processes and activities (5th ed.). John Wiley & Sons.

Liker, J. (2020). *The Toyota way: 14 management principles from the world's greatest manufacturer*. McGraw-Hill.

Pande, P.S., Neuman, R.P., Cavanagh, R.R. (2007). The Six Sigma Way. In: Boersch, C., Elschen, R. (eds) Das Summa Summarum des Management. Gabler. https://doi.org/10.1007/978-3-8349-9320-5_24

Pyzdek, T., & Keller, P. A. (2014). The six sigma handbook. McGraw-Hill Education.

Tortorella, G., & Fogliatto, F. (2017). Implementation of lean manufacturing and situational leadership styles: An empirical study. *Leadership & organization development journal*, *38*(7), 946-968.doi: https://doi.org/10.1108/LODJ-07-2016-0165

Womack, J. P., & Jones, D. T. (1997). Lean Thinking—Banish Waste and Create Wealth in your Corporation. *Journal of the Operational Research Society*, *48*(11), 1148. https://doi.org/10.1057/palgrave.jors.2600967

The Journal of Management and Engineering Integration (JMEI) © 2026 by AIEMS Inc., is licensed under CC BY 4.0. To view a copy of this license, visit https://creativecommons.org/licenses/by/4.0/

The Journal of Management and Engineering Integration Vol. 19, No. 2 | Winter 2026 DOI: 10.62704/10057/59058

THE IMPACT OF INVESTMENT AMOUNT AND TIMING ON STARTUP SUCCESS: EVIDENCE FROM U.S. SEMICONDUCTOR STARTUPS

Yooneun Lee [1]

Ganesh Vaidyaa Ramanathan [2]

KwangWook Gang [3]

Seungbeom Kim [4]

[1] University of Dayton

[2] Nidec Motor Corporation

[3] Salisbury University

[4] Hongik University

ylee03@udayton.edu; Ganesh.Ramanathan@nidec-motor.com; kxgang@salisbury.edu; sbkim@hongik.ac.kr

Abstract

This study examines how the amount and timing of investment relate to the success of U.S. semiconductor startups, where success is defined as a successful exit through an initial public offering (IPO) or acquisition. Using company- and transaction-level data from CB Insights covering 1,152 U.S. semiconductor startups, 9,543 funding records, and 2,758 investors over more than four decades (1980–2025), we apply Kaplan–Meier estimation and Cox proportional hazards regression, with survival time measured from each startup's first observed funding. Successful exits are infrequent and slow: only 32.1% of startups exited during the observation window, the estimated cumulative probability of exit remains below 10% for the first two decades after first funding, and exits are concentrated two to three decades after first funding. In the Cox models, company valuation is strongly associated with a faster rate of successful exit, and the size of an individual financing round is associated with faster exit when valuation is not controlled for, whereas cumulative total funding is not a significant predictor in either specification. The results suggest that the structure and timing of financing rounds, rather than the total amount of capital raised, distinguish startups that reach a successful exit. The findings offer guidance to investors, entrepreneurs, and policymakers on how capital deployment relates to long-run startup outcomes in a capital-intensive, policy-sensitive industry, and contribute sector-specific survival evidence to the literature on venture finance and innovation.

Keywords: *Semiconductor Startups, Startup Success, Successful Exit, Venture Capital, Investment Timing, Funding Amount, Kaplan–Meier Estimation, Cox Proportional Hazards Model, Survival Analysis, IPO, Acquisition*

1. Introduction

The semiconductor industry forms the foundation of today's global digital economy, powering technologies from smartphones and electric vehicles to defense systems and high-performance computing. As the demand for advanced chip capabilities intensifies, semiconductor startups play a pivotal role in driving innovation at the edge of design, architecture, and materials science. However,

Submitted: June 30, 2026
Revised: October 21, 2026

despite their critical function, these startups face disproportionately high barriers to success due to capital intensity, long development timelines, and high technical risk. Although multiple policy efforts—most notably the CHIPS and Science Act—have injected significant public and private capital into the ecosystem, a deeper understanding of how that capital influences startup outcomes is still lacking (SIA, 2023; SIA, 2022).

Startups in this domain face distinct challenges compared to those in software or consumer-facing sectors. Unlike apps that can scale rapidly with modest resources, semiconductor ventures often require tens to hundreds of millions in early-stage capital just to reach manufacturable proof-of-concept (Fontana & Malerba, 2010). At the same time, global policy competition—from the U.S., EU, China, and others—has created a rapidly evolving incentive landscape for semiconductor R&D and manufacturing (SIA, 2021; SIA, 2023). These macro trends are driving unprecedented investment into startups; yet the specific timing and structuring of this funding remain critical. A growing body of research suggests that when a startup receives capital, and how much it raises at each stage, may be as important as the technology itself in predicting survival or exit (Nanda & Rhodes-Kropf, 2013; Paik, 2014).

The types of investors supporting semiconductor startups are diverse and expanding. Traditional venture capital firms continue to lead early-stage investments, focusing on scalability and exit potential, while strategic corporate venture arms invest in startups aligned with internal roadmaps. Government funds, such as those driven by national security or economic competitiveness goals, play a foundational role in de-risking deep-tech innovation (SIA, 2020; Paik, 2014). Early-stage investment decisions are often based on a mixture of perceived technical feasibility, founder credibility, market momentum, and—crucially—whether the firm has previously been funded (Fontana & Malerba, 2010; Paik, 2014). Repeat founders with successful exits, for example, tend to receive funding faster and in larger amounts. Prior funding also serves as a validation signal for follow-on investors, shaping both deal flow and due diligence outcomes (Nanda & Rhodes-Kropf, 2013).

Importantly, investment timing is deeply entangled with market cycles. Startups backed during “hot” periods of investor enthusiasm may secure faster capital but face higher exit pressure, whereas those funded during downturns may benefit from longer development timelines and more patient capital (Gross, 2015; Nanda & Rhodes-Kropf, 2013). Within the semiconductor space, this sensitivity is heightened due to the sector’s cyclicality, foundry bottlenecks, and policy shifts. Furthermore, startup exit scenarios—IPO, acquisition, merger, or liquidation—are shaped by the availability and sequencing of capital. IPOs remain the most visible form of success, yet are rare due to regulatory and scale thresholds. Acquisitions are more frequent, often led by integrated device manufacturers (IDMs) or platform companies seeking talent, intellectual property, or market entry points (Paik, 2014). Meanwhile, a significant share of startups fail quietly—either dissolving due to funding constraints or being absorbed into other ventures in acquisition and hire arrangements.

While recent initiatives such as the CHIPS and Science Act and semiconductor-specific tax credits have greatly improved funding access, the question of how investment decisions—particularly timing and amount—impact startup outcomes remains largely underexplored in empirical academic research. Prior studies tend to either generalize across industries or focus on capital availability without modeling its temporal effects on firm trajectory. This study directly addresses that gap by investigating how financial structuring influences the success of semiconductor startup firms, with success operationalized as a successful exit through IPO or acquisition and analyzed in a time-to-event framework that accounts for startups whose outcomes are not yet observed. Using Kaplan–Meier estimation to describe when exits occur and Cox proportional hazards regression to relate exit

timing to deal size, cumulative funding, and valuation, the research aims to offer actionable insights into what kinds of capital support semiconductor startups not just at launch, but all the way through to exit.

2. Literature Review

The semiconductor sector continues to represent a cornerstone of technological and economic advancement, particularly in the United States, where rapid investment growth and policy incentives are driving startup formation in chip design, AI accelerators, and fabrication solutions (SIA, 2024). Despite a growing number of semiconductor startups, venture capital (VC) funding remains pivotal in determining which firms ultimately survive and succeed. The capital-intensive nature of the semiconductor industry and long time-to-market cycles make early financial support a key differentiator for growth. Alvarez-Garrido and Dushnitsky (2016) emphasized that success in startups should encompass both financial exits and innovation outputs such as patents. Fontana and Malerba (2010) confirmed that early investments increase the likelihood of survival in semiconductor ventures, particularly by signaling startup quality to future investors. Similarly, Gross (2015) and Nanda and Rhodes-Kropf (2013) found that market timing strongly affects performance outcomes, with startups launched during hot investment cycles exhibiting higher variance in results. Paik (2014) highlighted that prior founder experience plays a protective role in venture survival, especially within the capital-intensive semiconductor domain.

Several empirical studies employ survival analysis methodologies to analyze startup success trajectories over time. Shin, Kraemer, and Dedrick (2017) applied survival modeling to semiconductor manufacturing firms, demonstrating that R&D investment and external funding significantly influence longevity. Wu et al. (2018) used similar techniques to show how survival curves diverge in memory manufacturing startups depending on early capital flows. Bonaventura et al. (2020) and Gastaud, Carniel, and Dalle (2019) also explored survival through the lens of network dynamics, indicating that employee movement and investor connectivity strongly correlate with venture lifespan. Startup ecosystems have also been studied using deep learning approaches. For example, Lyu et al. (2021) and Gao and Xiao (2025) introduced graph-based models to predict startup outcomes using venture capital relationships and multivariate time series patterns. Potanin et al. (2023) further simulated VC portfolio strategies using data-driven modeling from Crunchbase to guide investment decisions. These works demonstrate that prediction models incorporating founder attributes, investment rounds, and innovation metrics can outperform heuristic or intuition-based investment decisions.

Recent literature has also shifted toward examining founder-centric and psychological models of success. Virágh (2024) proposed a five-dimensional framework grounded in founder interviews, capturing facets such as firm performance, personal fulfillment, team cohesion, community impact, and financial security. Their findings revealed that founders frequently value internal satisfaction, social recognition, and autonomy above financial exit outcomes, echoing earlier frameworks from Wach et al. (2016) and Fisher et al. (2014) that advocate a subjective and multidimensional view of entrepreneurial success. Sevilla-Bernardo et al. (2022) conducted a structured review of 60 papers to identify the most consistent predictors of startup success, including idea quality, founder leadership, and business model robustness. Kovalov et al. (2024) presented a typology of startup failure modes grouped into financial, leadership, product-market, regulatory, and strategic design categories, underscoring how systemic errors in early stages frequently cascade into venture shutdown.

In parallel, other scholars have embraced computational methods to augment traditional theory.

Asgari et al. (2022) applied sentiment analysis to Twitter discourse and revealed that positive founder sentiment, frequent investor engagement, and high transparency are public signals linked to favorable funding rounds and user traction. Meanwhile, Bakarich et al. (2023) emphasized that the inclusion of diverse perspectives from technical and business co-founders enhances adaptive decision-making in fast-evolving technology markets. Studies such as those by Zhang (2011) and Prohorovs et al. (2019) continue to highlight that founder education, sector-specific experience, and network capital play an outsized role in early-stage performance. Del Bosco et al. (2021) contributed findings suggesting that regional context and demographic factors shape access to accelerators, which in turn affect growth outcomes. Similarly, Díaz-Santamaría and Bulchand-Gidumal (2021) concluded from econometric modeling that location, promotional commitment, and founder engagement were significantly associated with higher success probability.

Taken together, this literature suggests that startup success, particularly within the semiconductor industry, is not reducible to singular metrics like valuation or exit events. Instead, it reflects a composite of early investment quality, founder characteristics, innovation timelines, team and network dynamics, and environmental context. Traditional survival analysis and Cox hazard regressions remain powerful tools for assessing time-dependent outcomes, while newer AI-based methods provide granular, real-time projections. Importantly, subjective founder definitions of success—including autonomy, health, impact, and fulfillment—must be considered alongside economic indicators. As the field evolves, integrated approaches that blend statistical rigor, behavioral insights, and contextual nuance appear most promising for understanding and predicting startup trajectories.

3. Dataset and Data Preprocessing

This study uses proprietary company- and transaction-level data obtained from CB Insights. The dataset includes information on private companies' funding histories, investors, financing rounds, industry classifications, and exit events, including acquisitions and initial public offerings (IPOs) over more than four decades, which enables the tracking of semiconductor startups from their observed financing activity through potential exit outcomes.

The final company-level analytical sample comprises 1,152 U.S. semiconductor startups, representing 9,543 underlying funding and investment records involving 2,758 unique investors. The observed funding activity spans from December 1, 1980, to April 6, 2025, the latter serving as the common administrative censoring date. After consolidating duplicated investor-level observations, the data contained 4,421 unique company-round records. Table 1 (Panel A) summarizes the terminal outcomes observed in the sample. Of the 1,152 startups, 370 (32.1%) experienced a successful exit, consisting of 307 acquisitions (26.6%) and 63 IPOs (5.5%); acquisitions therefore outnumber IPOs by roughly five to one, consistent with the industry pattern in which integrated device manufacturers and platform companies absorb startups for talent and intellectual property. Sixteen startups (1.4%) experienced an observed failure event, whereas the remaining 766 (66.5%) had no recorded terminal outcome and were administratively right-censored at the end of the observation period. Panel B reports descriptive statistics for the observed post-funding duration, the disclosed funding amounts, and the number of observed funding rounds.

Table 1. Sample Composition and Descriptive Statistics

Panel A. Terminal outcome (company level, n = 1,152)			
Outcome		**n**	**% of sample**
Successful exit (event of interest)		370	32.1
Acquisition		307	26.6
IPO		63	5.5
Observed failure (censored at failure date)		16	1.4
No terminal outcome recorded (censored)		766	66.5
Total		**1,152**	**100.0**
Panel B. Descriptive statistics			
Statistic	**Observed post-funding duration (years; company level, n = 1,152)**	**Disclosed funding amount per record (US$ millions; investor-level records, n = 8,020)**	**Observed funding rounds (company level, n = 1,152)**
Mean	11.246	20.52	3.12
Standard deviation	8.834	71.30	2.68
Minimum	0.063	0.01	—
25th percentile	4.488	3.96	1.00
Median	8.200	10.80	2.00
75th percentile	16.895	21.00	4.00
Maximum	44.345	3,600.00	—

Note: Percentages in Panel A are of the 1,152 startups. Observed post-funding duration is measured from the first observed funding transaction to successful exit, observed failure, or the censoring date; it is not business age, because founding dates are not observed. Funding amounts are as recorded by CB Insights (US$ millions) for the 8,020 investor-level records with a disclosed amount; because syndicated rounds appear once per investor, these records can repeat the same round. Minimum and maximum of the number of rounds are not reported.

Two features of Panel B matter for the analysis that follows. First, the observed post-funding duration is wide-ranging (median 8.2 years, interquartile range 4.5–16.9 years, maximum 44.3 years), so the follow-up available for individual startups differs by more than an order of magnitude; this is precisely the situation in which right-censoring must be handled explicitly rather than by discarding incomplete observations. Second, funding amounts are extremely right-skewed: the mean deal (US$20.5 million) is nearly twice the median (US$10.8 million), the standard deviation (US$71.3 million) is more than three times the mean, and the largest deal (US$3.6 billion) is more than 300 times the median. A small number of very large rounds therefore dominates the dollar scale, which has implications for how the amount variables are entered in the regression models (Section 4.2).

The initial dataset contained investment records involving semiconductor companies. Since a single financing round may involve multiple investors, the raw data included multiple transaction-level observations for the same startup. To avoid treating repeated investments or investor records as independent observations, the data were aggregated to the company level. The final analytical sample was limited to companies headquartered in the United States with a valid observed funding date. For each startup, the time origin was defined as the date of its first observed funding transaction in the dataset. Funding amounts were recorded by CB Insights in US$ millions. To reduce distortion from repeated investor-level entries in syndicated financing rounds, financing variables were constructed from unique company-round observations. The Kaplan–Meier analysis in Section 4.1 uses the company-level sample. The Cox regressions in Section 4.2 are estimated on the

underlying funding records, because the investment-related covariates are observed at the transaction level: deal amount is the disclosed amount of the individual funding transaction, company total funding is the startup's cumulative disclosed funding, and company valuation is the company valuation reported with the funding record.

The event of interest was a successful exit, defined as an IPO, acquisition, acquisition of a business unit, acquisition financing transaction, acquisition of a peer-to-peer company, or take-private transaction. The event date was defined as the first observed date associated with one of these exit outcomes. Startups without an observed successful exit by the end of the observed study period were treated as right-censored. Survival duration was calculated as the number of days from each startup's first observed funding date to either its successful exit date, observed failure date, or the common censoring date. The 16 startups with an observed failure event were thus censored at the date of failure; because failure precludes a subsequent successful exit, this is a conservative simplification that, given the small number of failures (1.4% of the sample), has negligible influence on the estimates. This variable is referred to as observed post-funding duration rather than business age, because the data identify the first observed funding transaction rather than each startup's founding date. This structure allows the Kaplan–Meier estimator to accommodate differences in observation periods across startups while accounting for firms that had not experienced a successful exit by the end of the study period.

4. Analysis

4.1. Kaplan-Meier Estimation

The Kaplan-Meier (KM) estimator is a nonparametric method widely used in survival analysis to estimate the probability that an observation remains event-free beyond a given point in time. Unlike conventional regression methods, the Kaplan–Meier method accommodates right-censored observations, which occur when the event of interest has not been observed for some firms by the end of the study period. In the present study, the event of interest is a successful exit as defined in Section 3, that is, an initial public offering (IPO) or an acquisition-type transaction. Startups that had not experienced a successful exit by the common censoring date were treated as right-censored.

For each startup, survival time was measured from the date of the first observed funding transaction to the date of successful exit or the end of the observation period. The estimated survival function, S(t), represents the probability that a startup has not yet experienced a successful exit by time t. Accordingly, a decline in the survival curve indicates the occurrence of IPOs or acquisitions over time. This interpretation differs from traditional clinical survival analysis, where a decline typically represents failure or death. In this study, a decline in the curve reflects the occurrence of a favorable terminal outcome.

Figure 1 presents the KM estimate for the full sample of 1,152 startups. Panel (a) shows the conventional survival function, S(t), with its 95% confidence interval, plotted against years since first observed funding; the segment beyond 40 years is drawn in grey because only a small number of startups remain under observation at those durations. Panel (b) shows the same estimate expressed as the cumulative probability of successful exit, 1 − S(t), on a magnified scale, with the estimate marked at 5, 10, 20, 30, and 40 years. Because the event is a favorable outcome, panel (b) is the more natural reading of the curve: it rises as startups exit. The curve in panel (a) begins at one because all startups are initially at risk of a successful exit following their first observed funding event.

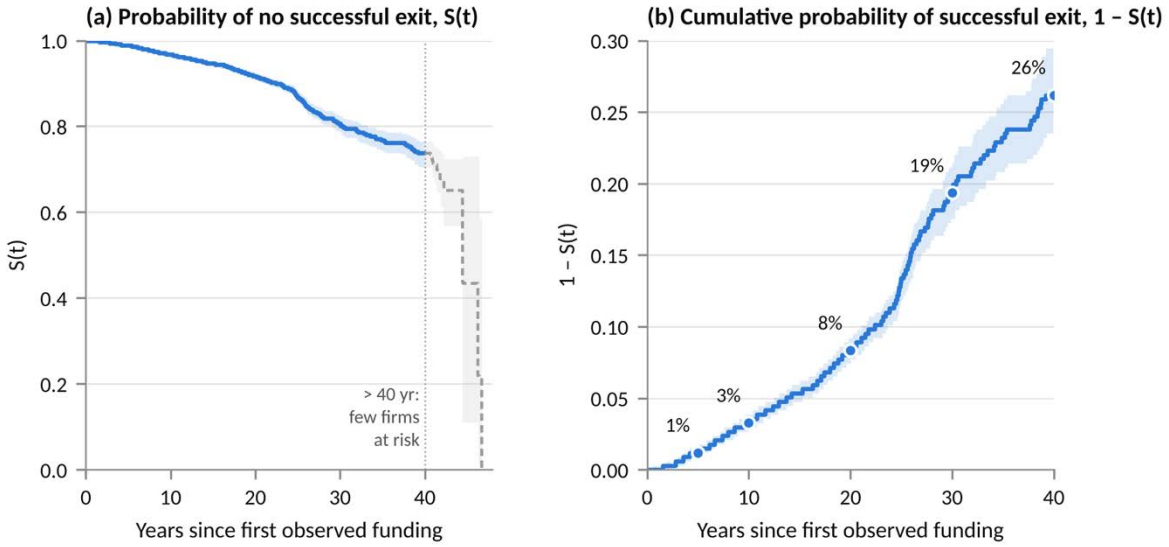


Figure 1. Kaplan–Meier Estimate of Time to Successful Exit for U.S. Semiconductor Startups (n = 1,152; 370 successful exits). (a) Probability of remaining without a successful exit, S(t), with 95% confidence interval (shaded); the dashed grey segment beyond 40 years rests on very few firms at risk. (b) The same estimate expressed as the cumulative probability of successful exit, 1 − S(t); markers give the estimate at 5, 10, 20, 30, and 40 years after first observed funding. Startups without an observed exit are right-censored.

Three features of the curve stand out. First, successful exits are slow to materialize. The estimated probability of remaining without a successful exit is approximately 0.99 at five years, 0.97 at ten years, and 0.92 at twenty years after first observed funding; equivalently, only about 1%, 3%, and 8% of startups are estimated to have exited by those horizons (Figure 1b and Table 2). Second, the rate of exit is not constant over time. The curve declines gently during the first two decades and then more steeply between roughly 20 and 30 years after first funding, when the cumulative exit probability rises from about 8% to about 19%, before flattening again; by 40 years the cumulative exit probability is roughly 26%. The stepwise pattern of the curve reflects the discrete timing of observed exit events, with larger downward movements indicating periods in which several startups exited in close succession. Third, as the number of startups remaining under observation decreases at longer follow-up durations, the confidence interval widens markedly. Beyond approximately 40 years the estimate rests on a handful of firms, and the steep drops in that region (grey segment in Figure 1a) should not be interpreted substantively.

Because the estimated probability of remaining without a successful exit stays above 0.70 through 40 years of follow-up, a median time to successful exit cannot be meaningfully estimated from these data: the curve falls below 0.50 only in the extreme tail of the distribution, where very few startups remain at risk and the estimate is unstable. This is consistent with the descriptive statistics—only 370 of the 1,152 startups (32.1%) experienced a successful exit during the observation period—and with the capital intensity and long development cycles that characterize the semiconductor industry. The concentration of exits two to three decades after first observed funding also deserves comment. Part of this pattern is plausibly genuine: semiconductor ventures require long periods to reach manufacturable products and to become attractive acquisition targets. Part of it may also reflect the structure of the data, in which startups first funded in the 1980s and 1990s have had far longer to exit than those funded more recently, so that the long-duration segment of the curve is dominated by earlier cohorts that faced different market and

policy conditions. Overall, the figure indicates that successful exits are relatively infrequent in the sample and that a substantial share of semiconductor startups remains without an observed IPO or acquisition throughout the available follow-up period.

Table 2. Kaplan–Meier Estimates at Selected Horizons after First Observed Funding

Years since first observed funding	*S(t)*	Cumulative exit probability, 1 − *S(t)* (%)
5	0.99	1.2
10	0.97	3.3
15	0.95	5.4
20	0.92	8.3
25	0.87	13.4
30	0.81	19.3
35	0.77	23.2
40	0.74	26.2

Note: *S(t)* is the Kaplan–Meier estimate of the probability of remaining without a successful exit *t* years after first observed funding (n = 1,152; 370 successful exits).

4.2. Cox Proportional-Hazard Model

The Cox proportional hazards (PH) model was used to examine the relationship between investment-related characteristics and the timing of successful exit among U.S. semiconductor startups. Whereas the Kaplan–Meier estimator provides an unadjusted description of the probability of remaining without a successful exit over time, the Cox model permits the simultaneous evaluation of multiple covariates while accounting for right-censored observations. In this study, the event of interest was successful exit, defined as an IPO, acquisition, acquisition of a business unit, acquisition financing transaction, acquisition of a peer-to-peer company, or take-private transaction.

The Cox model estimates the hazard rate, that is, the instantaneous rate at which the event occurs at time t among startups that have not yet experienced it. The logarithm of the hazard is modeled as a linear function of the covariates x, so that each covariate acts multiplicatively on a common baseline hazard. The Cox proportional hazards model is expressed as:

$$h(t) = h_0(t)\exp\left(\beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_p x_p\right)$$

where $h(t)$ is the hazard of successful exit at time t for a startup with covariate values $\left(x_1, x_2, \ldots, x_p\right)$; $h_0(t)$ is the unspecified baseline hazard function, and the coefficients $\left(\beta_1, \beta_2, \ldots, \beta_p\right)$ measure the effect of each covariate on the log hazard. The model does not require the researcher to specify the functional form of the baseline hazard. Instead, it estimates the relative effect of each explanatory variable on the hazard of successful exit through the exponentiated coefficient, $exp(\beta)$, commonly interpreted as a hazard ratio. A hazard ratio greater than one indicates a faster rate of successful exit (a shorter expected time to exit), and a hazard ratio below one a slower rate, holding the other covariates constant.

To evaluate the effects of investment-related variables, two Cox proportional hazards specifications were estimated. Model 1 includes deal amount, company valuation, and company total funding, to examine whether financial resources and market-based signals are jointly

associated with the rate of successful exit. Model 2 retains deal amount and total funding but omits valuation and therefore isolates the relationship between investment amount and exit timing. Both models are estimated on funding records rather than on the company-level sample of Section 4.1: Model 1 uses the 4,114 records with a reported company valuation, among which 221 successful-exit events are observed, and Model 2 the 7,954 records with a disclosed deal amount, among which 54 events are observed. Because the two models are estimated on different sets of records, they are not nested and their fit statistics are not directly comparable. Model 2 was fitted with an L2 (ridge) penalty of 0.20 on the coefficients to stabilize estimation given its small number of events; the penalty shrinks coefficients toward zero, so its estimates are conservative. In both models, a hazard ratio greater than one indicates that a startup is associated with a faster rate of successful exit, conditional on not having previously exited. Monetary covariates enter the models in their original units; at that scale the coefficients and hazard ratios round to 0.000 and 1.000 at three decimals, so the tables report, in addition, the z-statistics and p-values, which are invariant to the unit of measurement. Model fit is summarized by the concordance index (the probability that, of two randomly chosen records, the one with the higher predicted hazard exits first) and by the likelihood-ratio test of the fitted model against a model with no covariates.

Table 3. Cox Proportional Hazards Regression Results for Financial Characteristics and Successful Exit (Model 1)

Variable	Coefficient (β)†	Hazard ratio exp(β)†	*z*	*p*
Deal amount	0.000	1.000	0.51	0.61
Company valuation	0.000	1.000	7.81	<0.005
Company total funding	0.000	1.000	0.47	0.64
Observations (funding records)		4,114		
Successful-exit events		221		
Concordance index		0.70		
Likelihood-ratio test		$\chi^2(3) = 28.28$, $p < 0.001$		

Note: Monetary covariates enter in their original units. † Per-unit coefficients are positive but round to 0.000 (hazard ratios to 1.000) at three decimals; the *z*-statistic and *p*-value are invariant to the unit of measurement. Model 1 is estimated on the funding records with a reported company valuation.

As shown in Table 3, company valuation is strongly associated with the timing of successful exit ($z = 7.81$, $p < 0.005$): startups with higher reported valuations experienced IPOs, acquisitions, or related exits at a markedly faster rate. In contrast, neither deal amount ($z = 0.51$, $p = 0.61$) nor company total funding ($z = 0.47$, $p = 0.64$) is statistically significant once valuation is included. Two interpretations are consistent with this pattern. The first is substantive: valuation summarizes investors' forward-looking assessment of a startup's technological potential, commercial viability, and expected growth, and this market-based signal carries more information about exit prospects than the quantity of capital raised. The second is partly mechanical: valuation and funding amounts are strongly correlated, because larger rounds are typically raised at higher valuations, so valuation may absorb much of the variation that deal amount would otherwise explain. It should also be noted that valuations are often reported close to the time of a late-stage round or an exit, so part of the association between valuation and exit timing may reflect proximity to the exit event rather than a causal effect; we therefore interpret valuation as a leading indicator of exit rather than as a determinant of it.

The concordance index of Model 1 is 0.70, indicating moderate ability to rank observations by their time to successful exit, and the likelihood-ratio test rejects the null hypothesis of no covariate effects ($\chi^2(3) = 28.28$, $p < 0.001$). Model 1 is estimated on the 4,114 funding records with a reported valuation, among which 221 successful-exit events are observed; the different estimation sample relative to Model 2 should be kept in mind when the two models are compared.

Table 4. Cox Proportional Hazards Regression Results for Investment Amount and Total Funding (Model 2)

Variable	Coefficient (β)†	Hazard ratio exp(β)†	*z*	*p*
Deal amount	0.000	1.000	2.64	0.01
Total funding	0.000	1.000	0.19	0.85
Observations (funding records)		7,954		
Successful-exit events		54		
Concordance index		0.92		
Likelihood-ratio test		$\chi^2(2) = 6.98$, $p = 0.03$		
L2 (ridge) penalty on coefficients		0.20		

Note: Monetary covariates enter in their original units. † Per-unit coefficients are positive but round to 0.000 (hazard ratios to 1.000) at three decimals; the *z*-statistic and *p*-value are invariant to the unit of measurement. Model 2 is estimated on the funding records with a disclosed deal amount; partial log-likelihood = −422.69, partial AIC = 849.39.

Table 4 presents Model 2, which examines deal amount and cumulative funding without conditioning on valuation. Here deal amount is significantly associated with the timing of successful exit ($z = 2.64$, $p = 0.01$), whereas total funding is not ($z = 0.19$, $p = 0.85$). The contrast with Model 1, in which deal amount was not significant, is informative: the association between round size and exit timing appears to operate through, or to be subsumed by, valuation. The size of an individual financing round is therefore more closely related to faster successful exit than the cumulative amount of capital raised across the startup's funding history, which suggests that concentrated investment at particular stages of development—rather than the accumulation of capital per se—is what distinguishes startups that reach a successful exit sooner. Model 2 is estimated on 7,954 funding records with 54 successful-exit events; its likelihood-ratio test is significant at the 5% level ($\chi^2(2) = 6.98$, $p = 0.03$) and its concordance index is 0.92. The high concordance should be read with caution: with only 54 events and a ridge penalty on the coefficients, the index is imprecise and the estimates are conservative, so Model 2 is best viewed as supporting the direction of the deal-amount effect rather than as a precise measure of its size.

Taken together, the Cox results suggest that the relationship between investment and startup success is more complex than a simple increase in cumulative capital. Company valuation appears to be the strongest indicator of successful-exit potential, while the amount of a specific investment round is also associated with faster exit when valuation is not controlled for. Cumulative total funding is not a significant predictor in either specification. These findings support the view that the quality and timing of financial support, rather than total funding alone, shape the trajectory of semiconductor startups.

Several caveats apply to these estimates. First, the monetary covariates enter the models linearly in their original units, so each coefficient describes the change in the log hazard per unit of currency regardless of a startup's size; given the extreme right-skew of funding amounts documented in Table 1, rescaling to larger units or a logarithmic specification would yield more

interpretable magnitudes and would be less sensitive to the handful of very large rounds. Second, because the models are estimated on funding records, several records can belong to the same startup; the standard errors therefore treat related records as independent and may be understated, so the p-values should be read as indicative rather than exact. Third, the proportional-hazards assumption—that the effect of each covariate on the hazard is constant over follow-up—was not formally tested. Finally, the models are not adjusted for the year of first funding, so the estimates pool startups whose exit opportunities differed with the semiconductor cycle and the policy environment. These extensions are left for future work.

5. Conclusion

This study examined the relationship between investment characteristics and successful exit among 1,152 U.S. semiconductor startups using Kaplan–Meier estimation and Cox proportional hazards regression. Three findings emerge. First, successful exits in this industry are infrequent and slow: fewer than one-third of startups exited during more than four decades of observation, the estimated cumulative probability of exit remains below 10% for the first two decades after first observed funding, and exits are concentrated 20 to 30 years after first funding. Second, cumulative total funding is not a consistent predictor of exit timing in either Cox specification. Third, what does predict faster exit is the market-based assessment of the startup—company valuation—and, when valuation is not controlled for, the size of individual financing rounds. Taken together, these results indicate that how and when capital is deployed matters more than how much is accumulated: investors' valuation signals and the structure of individual rounds carry information about exit prospects that total funding does not.

These results should be interpreted in light of several limitations. Many startups remained active without an observed exit or failure, resulting in substantial right-censoring, and failure outcomes may be underreported or inconsistently classified in the dataset; both features mean that the estimated exit probabilities describe recorded exits rather than the full distribution of startup fates. Successful exit was defined narrowly as IPO or acquisition rather than broader measures such as profitability, sustained growth, or technological impact. The analysis focuses on investment-related variables and does not incorporate founder experience, patent activity, technology maturity, investor reputation, or broader market conditions, nor does it adjust for the year of first funding, so cohort and market-cycle effects are pooled. The Cox models are estimated on funding records rather than on startups, with monetary covariates entered linearly despite their pronounced skewness, and company valuation may partly reflect proximity to an exit rather than an independent driver of it. Natural extensions include stratifying the survival analysis by funding cohort and round stage to address timing directly, re-estimating the Cox models at the company level with rescaled or log-transformed amounts and cluster-robust standard errors, and modeling failure as a competing risk. Nevertheless, the findings suggest that investors and founders should consider the timing and structure of financing rounds, rather than relying solely on cumulative funding, when evaluating pathways toward successful exit in the semiconductor sector.

6. Disclaimer Statements

Funding: This work was supported by the Ministry of Education of the Republic of Korea and the National Research Foundation of Korea (NRF-2024S1A5C3A02043476).

Conflict of Interest: None

7. References

Alvarez-Garrido, E., & Dushnitsky, G. (2016). Are entrepreneurial ventures' innovation rates sensitive to investor complementary assets? Comparing biotech ventures backed by corporate and independent VCs. *Strategic Management Journal, 37*(5), 819–834. https://doi.org/10.1002/smj.2359

Bonaventura, M., Ciotti, V., Panzarasa, P., Liverani, S., Lacasa, L., & Latora, V. (2020). Predicting success in the worldwide start-up network. *Scientific Reports, 10*, Article 345. https://doi.org/10.1038/s41598-019-57230-3

Del Bosco, B., Mazzucchelli, A., Chierici, R., & Di Gregorio, A. (2021). Innovative startup creation: The effect of local factors and demographic characteristics of entrepreneurs. *International Entrepreneurship and Management Journal, 17*(1), 145–164. https://doi.org/10.1007/s11365-019-00618-0

Díaz-Santamaría, C., & Bulchand-Gidumal, J. (2021). Econometric estimation of the factors that influence startup success. *Sustainability, 13*(4), Article 2242. https://doi.org/10.3390/su13042242

Fontana, R., & Malerba, F. (2010). Demand as a source of entry and the survival of new semiconductor firms. *Industrial and Corporate Change, 19*(5), 1629–1654. https://doi.org/10.1093/icc/dtq043

Gastaud, C., Carniel, T., & Dalle, J.-M. (2019). *The varying importance of extrinsic factors in the success of startup fundraising: Competition at early-stage and networks at growth-stage* [Preprint]. arXiv. https://doi.org/10.48550/arXiv.1906.03210

Lyu, S., Ling, S., Guo, K., Zhang, H., Zhang, K., Hong, S., Ke, Q., & Gu, J. (2021). *Graph neural network based VC investment success prediction* [Preprint]. arXiv. https://doi.org/10.48550/arXiv.2105.11537

Nanda, R., & Rhodes-Kropf, M. (2013). Investment cycles and startup innovation. *Journal of Financial Economics, 110*(2), 403–418. https://doi.org/10.1016/j.jfineco.2013.07.001

Paik, Y. (2014). Serial entrepreneurs and venture survival: Evidence from U.S. venture-capital-financed semiconductor firms. *Strategic Entrepreneurship Journal, 8*(3), 254–268. https://doi.org/10.1002/sej.1161

Potanin, M., Chertok, A., Zorin, K., & Shtabtsovsky, C. (2023). *Startup success prediction and VC portfolio simulation using CrunchBase data* [Preprint]. arXiv. https://doi.org/10.48550/arXiv.2309.15552

Prohorovs, A., Bistrova, J., & Ten, D. (2019). Startup success factors in the capital attraction stage: Founders' perspective. *Journal of East-West Business, 25*(1), 26–51. https://doi.org/10.1080/10669868.2018.1503211

Sevilla-Bernardo, J., Sánchez-Robles, B., & Herrador-Alcaide, T. (2022). Success factors of startups in research literature within the entrepreneurial ecosystem. *Administrative Sciences, 12*(3), Article 102. https://doi.org/10.3390/admsci12030102

Virágh, E. A. (2024). Startup success from the founder's perspective. *Society and Economy, 46*(4), 404–424.

Wach, D., Stephan, U., & Gorgievski, M. J. (2016). More than money: Developing an integrative multi-factorial measure of entrepreneurial success. *International Small Business Journal, 34*(8), 1098–1121. https://doi.org/10.1177/0266242615608469

CHIPS and Science Act of 2022, Pub. L. No. 117-167, 136 Stat. 1366.

Semiconductor Industry Association. (2020). *2020 state of the U.S. semiconductor industry*.

Semiconductor Industry Association. (2022). *2022 state of the U.S. semiconductor industry*.

Semiconductor Industry Association. (2023). *2023 state of the U.S. semiconductor industry*.

Semiconductor Industry Association. (2024). *2024 state of the U.S. semiconductor industry*.

The Journal of Management and Engineering Integration (JMEI) © 2026 by AIEMS Inc., is licensed under CC BY 4.0. To view a copy of this license, visit https://creativecommons.org/licenses/by/4.0/

The Journal of Management and Engineering Integration Vol. 19, No. 2 | Winter 2026

DOI: 10.62704/10057/59059

ONE DEBATE, THREE STANDARDS: ON THE USE OF SYNTHETIC SURVEY RESPONDENTS

Scott D. Swain [1]

[1] Clemson University

sdswain@clemson.edu

Abstract

The use of large language models (LLMs) as synthetic survey respondents has prompted debate over whether they can replace human participants in research. The present research argues that this debate is less about what LLMs are capable of than about which standards (i.e., epistemic criteria) should be used to decide what kinds of claims synthetic responses can support. A distinction is made between three types of knowledge claims (measurement, prediction, and simulation), each implying a different standard for evaluating synthetic responses. To examine this framework, 654 comments were analyzed from 73 public discussions posted in the previous four years, focusing on 420 comments that explicitly invoked an epistemic criterion. Measurement-oriented comments raised ontological and provenance objections and expressed strongly negative attitudes. Prediction-oriented comments focused on model performance and validation and expressed conditional attitudes. Simulation-oriented comments endorsed preliminary and bounded applications and expressed generally positive attitudes. Better agreement with human responses can strengthen prediction claims without making synthetic outputs direct measurements of people's attitudes or experiences. The paper proposes a brief claim-and-evidence statement specifying the intended claim, its supporting evidence, and the limits of inference to guide research reporting, peer review, and synthetic respondent vendor assessment.

Keywords: *Synthetic Respondents, Large Language Models, Measurement Validity, Prediction, Simulation, Attitudes, Content Analysis*

1. Introduction

The time and cost of collecting human survey data have prompted debate over using large language models (LLMs) as synthetic respondents in place of human participants (Argyle et al., 2023; Dillion et al., 2023; Wang et al., 2026). This study examines that debate through validity theory, distinguishing claims that responses reflect respondents' attributes from claims that they predict a criterion (Kelley, 1927; Cronbach & Meehl, 1955).

Proponents and critics (e.g., Argyle et al., 2023; Gao et al., 2025; Lin, 2025) each evaluate synthetic responses against a standard of validity that often goes unstated, so apparent disagreements about the technology may instead be disagreements about which standards apply. The study distinguishes measurement, prediction, and simulation claims and examines how the corresponding epistemic criteria relate to commenters' arguments and attitudes toward synthetic respondents.

2. Three Types of Claims and the Epistemic Criteria They Evoke

Survey responses often serve as the basis or reference point for three types of knowledge claims: measurement, prediction, and simulation. These different types of claims tend to evoke in

Submitted: July 30, 2026
Revised: October 19, 2026

researchers a particular epistemic criterion or basis by which a researcher may decide whether the claim is valid (Kelley, 1927; Cronbach & Meehl, 1955).

A measurement claim asserts that an attribute exists and contributes causally to the observed response (Borsboom et al., 2004; Borsboom, 2023; van Bork et al., 2024). The epistemic criterion is therefore that a real human attribute causally affects the response, which requires recognizing the distinct provenances (origins) of human and synthetic responses (Quattrociocchi et al., 2025). A human survey response reflects a cognitive and communicative process in which a person interprets a question, retrieves or constructs relevant information, forms a judgment, and reports it (Tourangeau et al., 2000). Such responses can be inattentive, strategic, stylistic, or erroneous, but they always result from a process within a human whose attributes (perceptions, memories, preferences, attitudes, goals, intentions, history) inform it (Groves et al., 2009; Krosnick, 1991; Swain et al., 2008; Bhaktha et al., 2026).

In contrast, a synthetic survey response is an LLM output, generated by sampling from a learned conditional distribution. The sampling is determined by the persona prompt (a description of a target respondent), the survey question, prior context, and tuning or decoding settings. A persona prompt is therefore not a query of any human proxy but a query of statistical regularities encoded in language artifacts produced by many humans.

Thus, under the causal theory of validity, persona-conditioned text produced by synthetic respondents fails the epistemic criterion for measurement because no target human's attribute values enter the generative process. Even perfect correspondence with a human response distribution would not make the output a measurement of a particular person's attitude, memory, or experience. Discourse participants who comment on measurement claims would thus be expected to make ontological or provenance objections and express negative attitudes towards the use of synthetic respondents as data.

This conclusion rests on a causal conception of validity. A functionalist alternative, in the spirit of criterion-oriented validation (Cronbach & Meehl, 1955), holds that a system whose outputs are indistinguishable from human responses across the relevant range of inputs may license at least some of the inferences that human responses license. Within this framework, output equivalence is the standard that defines a prediction claim, so the functionalist case for synthetic data is relocated rather than dismissed. That is, it is a case for prediction, judged by out-of-sample correspondence in the target population and domain. However, output equivalence fails to provide a causal link from a target person's attribute to the response, and that link is what separates measurement from prediction.

A prediction claim pertains to whether an output corresponds to what a target human would say or do. It is an empirical performance claim whose epistemic criterion is the sufficiency of out-of-sample fit in the target population and domain (Argyle et al., 2023; Wang et al., 2026). Prediction does not require that the synthetic response arise from the same causal process as the human response. Rather, it only requires that the output reliably anticipate the relevant human outcome. Predictive validity is therefore conditional on the population, domain, time, and task in which performance is established, and correspondence in one setting does not by itself establish transportability to another. Discourse participants concerned with prediction claims should therefore contribute arguments about training data, representation, calibration, verification, and held-out human benchmarks, and their attitudes should vary with the perceived adequacy of the evidence for prediction.

A simulation claim pertains to what should follow from a stated response model paired with

instructions for the model. Thus, the epistemic criterion for simulation claims is transparency about model behavior. Simulation need not claim that its outputs correspond to the responses of an actual target population. Instead, it only needs to claim that its conclusions are conditional consequences of specified assumptions, model properties, and instructions. Discourse participants concerned with simulation claims should therefore be expected to make arguments about the transparency of premises, reproducibility of results, and sensitivity to model and prompt changes (Cui et al., 2025; Taday Morocho et al., 2026). Additionally, their attitudes should be generally positive and associated with endorsements of ideation, pretesting, scenario exploration, and other bounded uses.

Table 1. Claims, epistemic criteria, and comments from the corpus

Claim	Epistemic criterion	Typical argument	Attitude	Representative comments
Measurement	A real human attribute causally affects the response	Ontological/ provenance objection	Negative	"It's not user research if there isn't a user."
Prediction	Out-of-sample fit in the target population/ domain	Model/data limitations	Conditional	"Depends on how accurately it predicts human survey responses."
Simulation	Transparent premises, reproducibility, and sensitivity	Preliminary or bounded-use endorsement	Positive	"Making the space of possibilities visible."

3. Discourse Corpus and Analytical Approach

3.1. Constructing and Coding the Discourse Corpus

Public discussions of synthetic respondents between 2022 and 2026 were located and scraped, resulting in 654 comments from 73 discussions. This included 385 Reddit comments from 40 threads, 238 Hacker News comments from 25 threads, and 31 comments from eight industry-press discussions. The thread title, root post or article, and immediate parent comment were used as interpretive context. Each comment was then coded by two research assistants trained in content analytic methods. The researchers worked independently, coding one construct at a time. Coding was facilitated by an interactive coding application that presented comments in randomized order, along with surrounding context. The researchers were not informed of the hypothesized relationships among the constructs.

Each construct had its own decision rules. The epistemic criterion was assigned from the rationale a comment treated as decisive, irrespective of the verdict reached, and each argument code required the coder to identify a passage in the comment that licensed it. Reliability was computed on the codes before disagreements were resolved through discussion.

Epistemic criterion: For each comment, coders identified whether and which of the three hypothesized epistemic criteria (measurement, prediction, or simulation) provided the dominant rationale (i.e., the criterion whose removal would change or collapse the comment's conclusion). The coding revealed that 89 (13.6%) comments used the measurement criterion, 264 (40.4%) used the prediction criterion, and 67 (10.2%) used the simulation criterion. The remaining 234 (35.8%) comments used non-epistemic criteria.

Arguments: For each comment, coders identified the presence or absence of four types of arguments: (1) an objection based on ontology or provenance, (2) an objection or endorsement based on model or data limitations, (3) an endorsement of preliminary or exploratory uses such as ideation, instrument pretesting, and low-stakes concept screening, and (4) an endorsement of any bounded use, which included preliminary or exploratory uses but also extended to human-data-anchored augmentation and researcher-side assistance. These argument codes were not mutually exclusive, and their frequencies by criterion are reported in Table 2.

Attitude toward synthetic respondents: For each comment, coders rated the expressed attitude toward synthetic respondents, with ratings ranging from 1 (strongly negative) to 5 (strongly positive). A rating of 3 indicated a neutral or mixed attitude. Of the 654 comments, 97 (14.8%) were rated strongly negative, 224 (34.3%) negative, 166 (25.4%) neutral or mixed, 144 (22.0%) positive, and 23 (3.5%) strongly positive. The mean attitude rating was 2.65 (SD = 1.08).

3.2. Coding Validation

For the epistemic criterion, exact agreement across the 654 comments was 80.7% (Cohen's κ = 0.71). Since Cohen's κ understates agreement when categories are imbalanced, Gwet's AC1 was also computed and found to indicate strong reliability (0.75). Among the 420 comments coded as using an epistemic criterion, the coders assigned the same focal criterion in 84.3% of cases (Cohen's κ = 0.71, Gwet's AC1 = 0.81), counting assignment of a different criterion or of a non-focal category as disagreement; among the 392 comments that both coders classified as using an epistemic criterion, agreement on the particular criterion was 90.3% (κ = 0.80, Gwet's AC1 = 0.87).

For attitude, the coders assigned exactly the same rating to 68.7% of comments and ratings within one scale point of each other for 98.9%. For directional attitude (any rating other than 3), agreement was 83.2% (Cohen's κ = 0.56, Gwet's AC1 = 0.73).

The four argument categories were coded as separate binary indicators. Because one category was relatively uncommon, reliability was also assessed using F1, which evaluates agreement on positive cases rather than being dominated by joint absences. Strong reliability was observed for each category: ontological/provenance objections (F1 = 0.80, Cohen's κ = 0.77, AC1 = 0.94), model/data limitations (F1 = 0.82, Cohen's κ = 0.69, Gwet's AC1 = 0.70), preliminary or exploratory uses (F1 = 0.86, Cohen's κ = 0.83, Gwet's AC1 = 0.93), and any bounded use (F1 = 0.85, Cohen's κ = 0.80, Gwet's AC1 = 0.88).

3.3. Analytical Strategy

The focal analyses involved the 420 comments that used an epistemic criterion (measurement, prediction, or simulation). Models 1-4 test whether the criteria that commenters use are associated with the arguments and endorsed uses they express in their comments. Because the argument profiles were nonexclusive, four binary outcomes were estimated separately using linear probability models (Models 1-4, $j = 1, \ldots, 4$). To assess robustness to this analytical choice (Aczel et al., 2026), fixed-effects logistic models were also estimated and yielded the same substantive conclusions. Model 5 replaced the binary argument outcome with the five-point attitude rating and was estimated by ordinary least squares. Finally, Model 6 added three argument indicators to Model 5 to test whether the measurement and simulation coefficients are reduced in magnitude once the arguments expressed in the comments are considered. The models were specified as follows, and parameters and error terms are model-specific:

Models 1-4: $\text{Argument}_{it}^{(j)} = \alpha_t^{(j)} + \beta_M^{(j)} M_{it} + \beta_S^{(j)} S_{it} + \boldsymbol{\gamma}^{(j)\prime} \boldsymbol{X}_{it} + \varepsilon_{it}^{(j)}$

Model 5: $\text{Attitude}_{it} = \alpha_t + \beta_M M_{it} + \beta_S S_{it} + \boldsymbol{\gamma}' \boldsymbol{X}_{it} + \varepsilon_{it}$

Model 6: $\text{Attitude}_{it} = \alpha_t + \beta_M M_{it} + \beta_S S_{it} + \delta_O OP_{it} + \delta_D MD_{it} + \delta_B BU_{it} + \boldsymbol{\gamma}' \boldsymbol{X}_{it} + \varepsilon_{it}$

$\text{Argument}_{it}^{(j)}$ is the j^{th} argument profile for comment i in thread t: ontological/provenance objection, model/data limitation, preliminary/exploratory use, or any bounded use (in the order of Models 1-4 in Table 3). Attitude_{it} is the attitude rating. Model terms and coefficients are as follows:

- α_t is the thread fixed effect.
- M_{it} and S_{it} indicate whether the comment applied the measurement or simulation criterion, respectively (prediction is the reference category), with coefficients β_M and β_S.
- $\mathbf{X}_{it}$ is a control vector of variables known to influence comment content (Swain, 2023; Weathers et al., 2015) such as log word count, an opening-post indicator, reply depth, relative position within the thread, a fragment indicator, and a question indicator. $\boldsymbol{\gamma}$ is the corresponding coefficient vector.
- OP_{it}, MD_{it}, and BU_{it} indicate an ontological/provenance objection, a model/data limitation, and any bounded use, respectively, with coefficients δ_O, δ_D, and δ_B.
- ε_{it} is the error term.

All models included thread fixed effects and thread-clustered standard errors. The focal sample contained 420 comments from 66 threads (31 with more than one focal criterion). Because the comments in a thread are posted within a short window, thread fixed effects absorb not only platform, topic, and product but also the period in which a discussion took place, including model capabilities and framing conventions at the time, so the criterion coefficients are identified from differences among comments within the same discussion. With 66 clusters, wild cluster bootstrap p-values (Rademacher weights, null imposed, 9,999 replications) for the criterion coefficients lead to the same conclusions as the cluster-robust p-values reported, and the within-thread sign tests in Section 4.2 do not rely on cluster asymptotics. Table 2 reports raw percentages. Table 3 reports the same differences estimated within threads, with the six comment controls included.

4. Results

4.1. Epistemic Criteria and Arguments

Table 2 shows the expected argument profiles: measurement comments emphasized ontological/provenance objections, prediction comments emphasized model/data limitations, and simulation comments endorsed preliminary and bounded uses. These differences persist in within-thread comparisons, controlling for six comment characteristics (Models 1-4, Table 3). A Holm–Bonferroni adjustment across the criterion coefficients in Table 3 leaves the conclusions unchanged.

Table 2. Arguments Appearing in Comments According to Which Epistemic Criterion is Used

Epistemic criterion used	*Ontological/ provenance*	*Model/data limitation*	*Preliminary/ exploratory use*	*Any bounded use*
Measurement	74/89 = 83.1%	34/89 = 38.2%	4/89 = 4.5%	7/89 = 7.9%
Prediction	1/264 = 0.4%	172/264 = 65.2%	44/264 = 16.7%	67/264 = 25.4%
Simulation	2/67 = 3.0%	21/67 = 31.3%	43/67 = 64.2%	47/67 = 70.1%

4.2. Epistemic Criteria and Attitudes Toward Synthetic Respondents

Mean attitude differed across the three epistemic criteria. Measurement comments averaged 1.43, prediction comments 2.81, and simulation comments 3.75. Model 5 (Table 3) tests whether these differences in mean attitude are explained by thread composition or comment characteristics by comparing comments within the same thread and controlling for six comment characteristics. They are not. Measurement comments remained 1.23 points more negative than prediction comments (SE = 0.12) and simulation comments 0.90 points more positive (SE = 0.14), both $p < 0.001$.

The rating distributions make the differences more concrete. Among measurement comments, 96.6% were negative, 1.1% were neutral or mixed, and 2.2% were positive. Prediction comments were distributed across negative (48.1%), neutral or mixed (23.9%), and positive (28.0%) attitudes. Among simulation comments, 7.5% were negative, 11.9% were neutral or mixed, and 80.6% were positive. The ordering also appeared within discussions. Measurement comments were more negative than prediction comments in 22 of 23 threads containing both criteria, with one reversal. Prediction comments were more negative than simulation comments in 19 of 24 threads, with four ties and one reversal. Measurement comments were more negative than simulation comments in all 18 threads containing both criteria. Exact sign tests were significant ($p < 0.001$) for all three comparisons. Within each period of the window (2023-2024, 2025, and 2026), measurement comments remain by far the most negative, and the share of simulation comments is no higher in 2026 than in 2023-2024, which argues against reading the criterion differences as a by-product of improving models or shifting framing conventions.

Table 3. Models of Arguments (Models 1-4) and Attitude (Models 5-6) as A Function of The Epistemic Criterion Used

Epistemic criterion used or argument made	**Model 1:** *Ontological/ provenance*	**Model 2:** *Model/data limitation*	**Model 3:** *Preliminary/ exploratory use*	**Model 4:** *Any bounded use*	**Model 5:** *Attitude*	**Model 6:** *Attitude*
Measurement	74.7 (6.4)***	−16.0 (6.8)*	−14.2 (4.5)**	−21.4 (6.0)***	−1.23 (0.12)***	−0.80 (0.16)***
Simulation	2.6 (3.5)	−35.0 (7.6)***	40.6 (9.8)***	32.8 (8.5)***	0.90 (0.14)***	0.38 (0.13)**
Ontological/provenance						−0.47 (0.18)*
Model/data limitation						−0.69 (0.10)***
Any bounded use						0.89 (0.12)***

*Notes. Coefficients with thread-clustered standard errors in parentheses. Models 1-4 are linear probability models in percentage points; Models 5-6 are OLS regressions in attitude-scale points. Prediction is the reference criterion. All models include six comment controls and thread fixed effects (n = 420 comments, 66 threads, 31 multi-criterion). * p < 0.05; ** p < 0.01; *** p < 0.001.*

The augmented attitude model (Model 6) added the argument profiles. Ontological/provenance objections (−0.47, $p < 0.05$), model/data limitations (−0.69, $p < 0.001$), and endorsement of any bounded use (+0.89, $p < 0.001$) were each associated with attitude in the expected direction, and their inclusion reduced the measurement coefficient from −1.23 to −0.80 and the simulation coefficient from 0.90 to 0.38, reductions of approximately 35% and 58%, respectively.

5. Discussion

5.1. Summary of Findings

Among comments invoking an epistemic criterion, arguments and attitudes differed by criterion within and across discussions. Measurement comments emphasized ontological/provenance objections and were predominantly negative; prediction comments focused on model/data limitations and expressed varied attitudes; simulation comments endorsed bounded uses and were most positive.

5.2. Implications for Research Practice and Management

The proposed framework shifts the practical question from whether synthetic respondents are acceptable in general to what their responses are intended to establish. Improved correspondence with human responses may strengthen a prediction claim without establishing the causal connection required by measurement. Conversely, simulation may support exploratory work without demonstrating predictive accuracy for a target population. For researchers, reviewers, and managers, these distinctions imply different requirements for evaluating and reporting synthetic responses.

Under the causal conception of measurement, persona-conditioned LLM outputs should not be reported as direct observations of target individuals' attitudes, memories, or experiences. This boundary concerns what the responses represent, not simply how closely they resemble human survey data.

Prediction claims instead require evidence of out-of-sample correspondence with human responses in the target population and domain. Authors and vendors should identify the human benchmark, survey questions, predictive performance, prompts, model version, and validation date. A successful comparison supports prediction under the conditions evaluated. Extending the claim to other populations, questions, or models requires additional evidence.

Simulation claims concern what follows under stated assumptions and model conditions. Their evaluation should therefore emphasize transparent assumptions and procedures, reproducibility, and sensitivity to changes in the model, prompts, and settings. Reporting should explain which conclusions depend on these choices rather than imply that simulated responses describe a target population.

A practical application would be a brief claim-and-evidence statement identifying the intended claim, its supporting evidence, and the limits of inference. Researchers could use this statement to delimit conclusions, reviewers to assess whether the evidence addresses the claim, and managers to judge whether a vendor's validation is relevant to the proposed application. When a study combines exploratory simulation and prediction, evidence supporting each use should be distinguished. Such reporting would make the stated basis of evaluation explicit without presuming all disagreements would be resolved. A related distinction is between AI that stands in for participants versus AI that assists the researcher and whose output people review before use. For example, in a large-scale reproducibility test, AI-assisted teams performed similarly to human-only teams on most outcomes but detected fewer major coding errors. Additionally, both outperformed AI-led teams (Brodeur et al., 2026).

5.3. Limitations and Future Research

The reported cross-sectional associations do not establish whether epistemic criteria shape attitudes toward synthetic respondents or whether existing attitudes influence which criteria

commenters invoke. Repeated observations of the same commenters could examine whether changes in criteria precede or follow changes in attitudes. A complementary experiment could hold synthetic survey results constant while presenting them as supporting measurement, prediction, or simulation claims. Researchers, reviewers, and managers could then indicate whether they accept the stated claim and what supporting evidence they require. Varying the evidence provided could further test whether acceptance responds to evidence relevant to each claim type.

The opt-in corpus compiled in this research captures selected public discussions rather than the distribution of views among researchers and practitioners more broadly. Comparisons with academic discussions could examine differences in how often each epistemic criterion appears and how each is associated with arguments and attitudes toward synthetic respondents. Repeated sampling over time could examine whether criterion prevalence and the associations between criteria, arguments, and attitudes change as model capabilities and framing conventions evolve. The 234 comments classified as using non-epistemic criteria suggest a further extension: examining how those considerations interact with measurement, prediction, and simulation criteria.

Finally, future research could test the practical value of an explicit claim-and-evidence statement. A separate experiment could keep the intended claim, synthetic results, and available evidence constant while varying whether the report includes such a statement. Researchers, reviewers, and managers could be asked which conclusions the results support and what additional evidence they require. The experiment could also assess whether participants revise their conclusions when claim-relevant evidence changes. The aim would not be greater acceptance of synthetic respondents, but closer alignment between the intended claim, the evidence participants consider, and the conclusions they draw.

6. Disclaimer Statements

Funding: None

Conflict of Interest: None

7. References

Aczel, B., Szaszi, B., Clelland, H. T., Kovacs, M., ..., Swain, S. D., ..., & Nosek, B. A. (2026). Investigating the analytical robustness of the social and behavioural sciences. *Nature*, *652*, 135-142. https://doi.org/10.1038/s41586-025-09844-9

Argyle, L. P., Busby, E. C., Fulda, N., Gubler, J. R., Rytting, C., & Wingate, D. (2023). Out of one, many: Using language models to simulate human samples. *Political Analysis, 31*(3), 337-351. https://doi.org/10.1017/pan.2023.2

Bhaktha, N., Silber, H., & Lechner, C. M. (2026). A unified framework for characterizing response quality in surveys with multi-item scales. *Communications Psychology, 4*, Art. 86. https://doi.org/10.1038/s44271-026-00463-2

Borsboom, D. (2023). Psychological constructs as organizing principles. In L. A. van der Ark, W. H. M. Emons, & R. R. Meijer (Eds.), *Essays on contemporary psychometrics* (pp. 89-108). Springer. https://doi.org/10.1007/978-3-031-10370-4_5

Borsboom, D., Mellenbergh, G. J., & van Heerden, J. (2004). The concept of validity. *Psychological Review, 111*(4), 1061-1071. https://doi.org/10.1037/0033-295X.111.4.1061

Brodeur, A., Valenta, D., Marcoci, A., …, Swain, S. D., …, & Zozoungbo, C. (2026). AI-assisted teams outperform AI-led teams but not human-only teams in assessing research reproducibility in quantitative social science. *Proceedings of the National Academy of Sciences*, *123*(22), e2524747123. https://doi.org/10.1073/pnas.2524747123

Cronbach, L. J., & Meehl, P. E. (1955). Construct validity in psychological tests. *Psychological Bulletin, 52*(4), 281-302. https://doi.org/10.1037/h0040957

Cui, Z., Li, N., & Zhou, H. (2025). A large-scale replication of scenario-based experiments in psychology and management using large language models. *Nature Computational Science, 5*, 627-634. https://doi.org/10.1038/s43588-025-00840-7

Dillion, D., Tandon, N., Gu, Y., & Gray, K. (2023). Can AI language models replace human participants? *Trends in Cognitive Sciences, 27*(7), 597-600. https://doi.org/10.1016/j.tics.2023.04.008

Gao, Y., Lee, D., Burtch, G., & Fazelpour, S. (2025). Take caution in using LLMs as human surrogates. *Proceedings of the National Academy of Sciences, 122*(24), e2501660122. https://doi.org/10.1073/pnas.2501660122

Groves, R. M., Fowler, F. J., Couper, M. P., Lepkowski, J. M., Singer, E., & Tourangeau, R. (2009). *Survey methodology* (2nd ed.). Wiley.

Kelley, T. L. (1927). *Interpretation of educational measurements*. World Book Company.

Krosnick, J. A. (1991). Response strategies for coping with the cognitive demands of attitude measures in surveys. *Applied Cognitive Psychology, 5*(3), 213-236. https://doi.org/10.1002/acp.2350050305

Lin, Z. (2025). Six fallacies in substituting large language models for human participants. *Advances in Methods and Practices in Psychological Science, 8*(3), 1-19. https://doi.org/10.1177/25152459251357566

Quattrociocchi, W., Capraro, V., & Perc, M. (2025). Epistemological fault lines between human and artificial intelligence. *arXiv*. https://doi.org/10.48550/arXiv.2512.19466

Swain, S. D. (2023). Unhelpful and unaware of it: A dyadic analysis of online product reviews. *Journal of Management and Engineering Integration*, *16*(1), 48-56. https://doi.org/10.62704/10057/25983

Swain, S. D., Weathers, D., & Niedrich, R. W. (2008). Assessing three sources of misresponse to reversed Likert items. *Journal of Marketing Research, 45*(1), 116-131. https://doi.org/10.1509/jmkr.45.1.116

Taday Morocho, E. E., Cima, L., Fagni, T., Avvenuti, M., & Cresci, S. (2026). Assessing the reliability of persona-conditioned LLMs as synthetic survey respondents. In *Companion Proceedings of the ACM Web Conference 2026*. Association for Computing Machinery. https://doi.org/10.1145/3774905.3795477

Tourangeau, R., Rips, L. J., & Rasinski, K. (2000). *The psychology of survey response*. Cambridge University Press. https://doi.org/10.1017/CBO9780511819322

van Bork, R., Rhemtulla, M., Sijtsma, K., & Borsboom, D. (2024). A causal theory of error scores. *Psychological Methods, 29*(4), 807-826. https://doi.org/10.1037/met0000521

Wang, M., Zhang, D. J., & Zhang, H. (2026). Large language models for market research: A data-augmentation approach. *Marketing Science, 45*(4), 728-751. https://doi.org/10.1287/mksc.2025.0009

Weathers, D., Swain, S. D., & Grover, V. (2015). Can online product reviews be more helpful? Examining characteristics of information content by product type. *Decision Support Systems, 79*, 12–23. https://doi.org/10.1016/j.dss.2015.07.009

The Journal of Management and Engineering Integration (JMEI) © 2026 by AIEMS Inc., is licensed under CC BY 4.0. To view a copy of this license, visit https://creativecommons.org/licenses/by/4.0/

WINTER
2026



AIEMS

AIEMS is committed to furthering the integration between business management, engineering, & industry. Our objective is to promote research, collaboration, & practice in these multidisciplinary areas. AIEMS seeks to encourage local, national, & international communication & networking via conferences & publications open to those in both academia & industry. We strive to advance professional interaction & lifelong learning via human & technological resources, and to influence and promote the recruitment and retention of young faculty and industralists.

AIEMS VOL 19 NO 2