2026 ASEE ANNUAL CONFERENCE June 21–24, 2026 | Charlotte Convention Center, NC
Host Sponsor UNC Charlotte
Paper ID #52525

A Gateway to Educational Research: Development of a Toolkit to Make Statistical Analysis Accessible

Dr. Darby Rose Riley, The Ohio State University
Tyler James Stump, The Ohio State University

Tyler Stump is a first year Ph.D. student in the Department of Engineering Education at The Ohio State University. Tyler received his B.S.in Biosystems Engineering at Michigan State University in 2022 and received his M.S. from Michigan State University in 2023. His engineering education interests include first-year engineering student experiences, computing education, and how to foster and develop creativity within programming courses.

Dr. Connor Jenkins, The Ohio State University

Connor Jenkins received his PhD in Electrical and Computer Engineering at The Ohio State University. His engineering education research interests include first-year engineering, teaching assistant programs, and technical communication

Dr. Krista M Kecskemety, The Ohio State University

Krista Kecskemety is an Associate Professor in the Department of Engineering Education at The Ohio State University. Krista received her B.S. in Aerospace Engineering at The Ohio State University in 2006 and received her M.S. from Ohio State in 2007. In 2012, Krista completed her Ph.D. in Aerospace Engineering at Ohio State. Her engineering education research interests include investigating first-year engineering student experiences, faculty experiences, and the research to practice cycle within first-year engineering.

A Gateway to Educational Research: Development of a Toolkit to Make Statistical Analysis Accessible

Abstract:

Engineering education currently struggles to recruit technical engineering faculty to perform educational research in their classrooms. A few common barriers for technical engineering faculty in improving their classroom are access to introductory educational research tools, lack of time and energy to dedicate towards extracurricular activities, and lack of incentive to participate in and publish research. Many technical instructors tend to monitor the students' success using grades (a measure of more technical skills), but there is a need to measure student impact beyond the technical. One method for conceptualizing these technical and non-technical skills is through entrepreneurially-minded learning (EML), a pedagogical approach to supporting the development of engineering students' attitudes, habits, and behaviors. This paper details the development and testing of a statistical toolkit that any engineering instructor, regardless of their experience with statistics, can use to measure student development of an entrepreneurial mindset (EM). The completed resource will include three survey instruments (developed as part of an ongoing KEEN grant to measure EM using the 3Cs), an excel worksheet which automates introductory statistical analysis, and written documentation describing how to use the survey and worksheet with appropriate rigor. The documentation describes each step of the process, starting from deploying the survey in a classroom and ending with reporting results to a conference venue such as ASEE. The statistical toolkit is currently being designed and improved through think aloud sessions with graduate teaching assistants and teaching faculty at a large R1 university in the Midwest. The statistical toolkit currently walks users through descriptive statistics and a paired-sample t-test, with plans to incorporate an ANCOVA model using regression analyses for student learning gains. The development team aims to disseminate this statistical toolkit broadly via conference publications and workshops, hopefully creating a user-friendly gateway into the world of educational research.

Introduction:

As part of an ongoing effort to introduce entrepreneurially-minded learning (EML) into first-year engineering classrooms, there has been a push to find methods for quantitatively measuring student outcomes [1]. Many researchers have turned to the "3Cs" framework developed by the Kern Entrepreneurial Engineering Network (KEEN). This framework posits that an Entrepreneurial Mindset (EM) "consists of three key elements: Curiosity, Connections, and Creating Value—the 3Cs" [2],[3]. While this framework provides a solid basis for qualitative evaluation of student outcomes [4]-[6], only one of the 3Cs had an existing validated tool that could be used for quantitative evaluation: the Five-Dimensional Curiosity Scale (5DC) [7]. As such, there existed a strong motivation to develop validated tools for measuring the remaining 2 Cs, a task undertaken by a research team at the Ohio State University [8].

These instruments are now in their final stages of development, with contextual validation for undergraduate engineering attained and plans to more widely test and validate the instruments in a variety of contexts [9]. A key upcoming step in this project is identifying appropriate dissemination strategies for the instruments. While many similar survey instruments are shared via traditional publication routes, an important part of the motivation behind the instrument development was to make EM and EML research more accessible. In that case, traditional publishing routes were likely to place the instruments at a disadvantage [10],[11]. In addition to publishing accessibility, the research team wished to find a way to ensure that performing the research was, in itself, as accessible as possible—that anyone, regardless of experience level, could pick up the instruments and use them not only to observe their own classroom, but to publish their work.

This paper details the development and testing of a statistical toolkit intended to accompany the validated 3Cs instruments—the 5DC and the 2Cs instruments developed by the OSU research team.

Background:

The following section will give a more detailed overview of the instruments the statistical toolkit builds on, as well as more depth in the need the statistical toolkit will fulfill.

The Validated 3Cs Instruments

KEEN's 3Cs framework describes the entrepreneurial mindset as being composed of three key competencies: Curiosity, Connections, and Creating Value. The research team's goal was to identify or develop a validated instrument to measure each of these elements independently.

Curiosity is defined by KEEN as the "insatiable" drive to "investigate a rapidly changing world" [2]. As a psychological concept, curiosity has been thoroughly explored; in 2018, a team of psychologists conducted a series of studies to reconcile existing theoretical frameworks of curiosity, leading to the formation of the Five-Dimensional Curiosity Scale (5DC) [7]. This instrument has been used in many contexts, including engineering education, and was determined by the research team to appropriately capture curiosity within KEEN's definition [8]. The instrument consists of 25 prompts across 5 factors, each of which is scored by the student on a 7-point Likert scale.

The remaining 2 Cs—Connections and Creating Value—could not be measured using existing validated instruments. KEEN defines “connections” as a behavior wherein students “habitually pursue knowledge and integrate it with their own discoveries,” and “creating value” as the ability to "persistently anticipate and meet the needs of a changing world” [2]. The research team developed prompts using the KEEN framework [2], as well as learning objective statements developed from KEEN materials [12]. The final validated surveys use a 7-point Likert scale (aligning with the 5DC’s design). The “connections” instrument includes 18 items across 4 factors, and the “creating value” instrument includes 21 items across 3 factors.

Statistical Competency and Educational Research

While the details of the statistical toolkit’s design and goals will be further described in the methods, it is important to contextualize the tool as a potential introduction to statistics for educational research.

Statistical literacy has been defined as “the ability to understand and critically evaluate statistical results that permeate our daily lives,” [13, p. 1] as well as the ability to use statistics to communicate [14], both important components of reading and publishing research. Despite this, instilling statistical literacy in undergraduate and graduate students remains an ongoing challenge [15],[16]. Research has repeatedly demonstrated low statistical literacy amongst pre-service teachers [17], indicating a lack of statistical education at the undergraduate level. Undergraduate and graduate engineering students have also been found to have low exposure to statistics [18],[19]. Previous research warns that “collective quantitative proficiency” is waning even in doctoral-level research education [20],[21]. Therefore, it cannot be assumed that those who may be interested in improving their classrooms through quantitative research have the statistical foundation necessary to do so in a critical, rigorous manner.

Bridging this gap is challenging. It is not the assumption of this research team that such a gap can be spanned by this statistical toolkit alone. However, the toolkit aims to accurately and completely educate independent researchers on some basics of statistical analysis, while hopefully providing a vocabulary for such individuals to seek further statistical education.

Methods

The following section describes the overall statistical toolkit “outline” that was developed in conjunction with the survey validation team, as well as the process for testing the statistical toolkit with its intended audience.

Statistical Toolkit Criteria and Constraints

As with any engineering design project, it is critical to outline the criteria (what the product must do) and constraints (what the product must not do). The toolkit criteria were developed as a joint effort between the instrument validation team and the toolkit development team. The instrument validation team outlined the following criteria for a “version 1.0” of the toolkit:

I. The toolkit must use all three instruments (the existing 5DC and the newly-validated 2Cs)
II. The toolkit must, in some way, specify to the user the specific context the surveys were validated in (first-year engineering undergraduate classes) and encourage users to apply the instruments in this context
III. The toolkit must at least be capable of performing a pre-post t-test

While the final version of the toolkit will include more statistical tests, this was considered a “bare minimum” for the first iteration.

The constraints for the toolkit were developed from a variety of sources, including conversations with teaching faculty and published recommendations for statistical education [22]-[25]. They are as follows:

I. The toolkit must be built in free-to-use software to ensure the greatest accessibility
II. The toolkit must have a clear linear progression through the data analysis process, rather than displaying all results at once
III. The toolkit must be accompanied by example data that can be used to practice the process [22]
IV. The toolkit must explain the conceptual meaning behind its methods, results, and terminology [22]
V. The toolkit must require users to make at least some selections, so as to actively engage them in the analysis process [22]
VI. The toolkit must highlight the figures necessary for proper statistical reporting (p-value, effect size, power, etc.) [23],[24]
VII. The toolkit must emphasize that statistical testing *cannot* identify causal relationships between variables [25]

With these in mind, a rough outline of the toolkit’s various modules was developed in Excel. This “low-res” prototype was then tested with the target audience.

Process for Testing

A think-aloud protocol was used to explore the toolkit with a graduate student lecturer at the Ohio State University. Think-aloud protocols are common in both educational research [26] and engineering design [27]; it requires a participant (in this case, the graduate student lecturer) to verbalize their thought process while completing a task or set of tasks (in this case, completing a statistical analysis), allowing an observer to understand the source of mistakes and miscommunications. Here, the think-aloud protocol allowed the researcher to identify areas where more explanation might be needed, or where incorrect inferences were made about the meaning of various statistical terms and outcomes.

The graduate student lecturer had minimal experience with statistics for educational research. They had deployed the 3Cs instrument twice in an electrical and computer engineering course for undergraduate juniors, seniors, and graduate students—once before an EM-related intervention, and once after.

Results and Discussion

The following section details the result of the think-aloud protocols, the edits made following these think-alouds, and next steps for continuing development of the toolkit.

First Think-Aloud

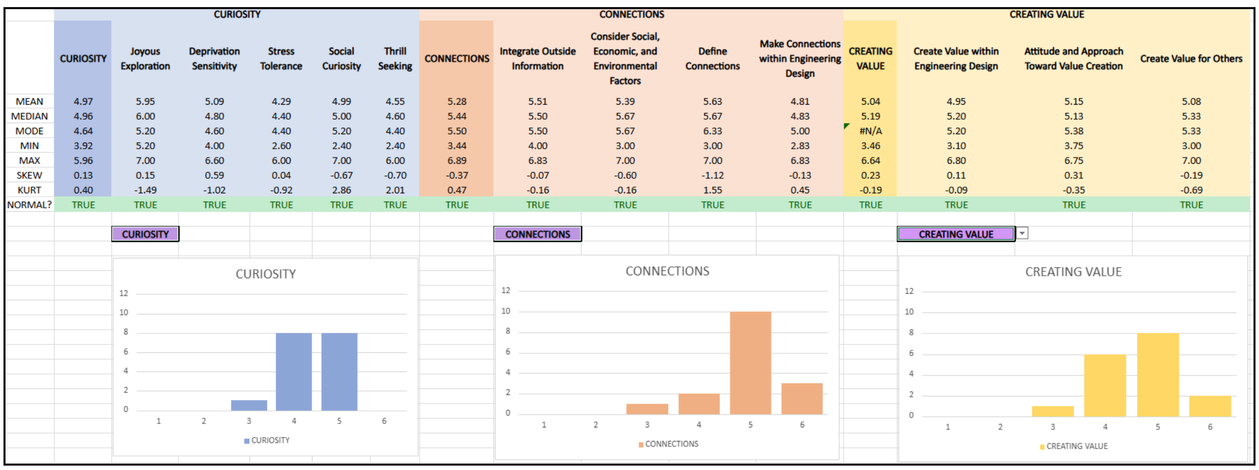
As previously discussed, the first think-aloud protocol centered around an “outline” of the tool’s various modules. It was not capable of actually performing any statistical analysis in this form. Rather, the participant was provided a list of steps, and the researcher acted in place of the tool, performing the requested statistical tests on demand in Excel. This process took approximately 90 minutes.

The purpose of this think-aloud protocol was to understand exactly what needed to be done, in what order, and how much extraneous information was requested by the user. Beyond these primary goals, three major pieces of feedback were identified:

I. The tool could not stand alone—enough additional information was requested by the participant that it was determined the interactive Excel sheet must be accompanied by a document with supporting information and direction.
II. This documentation needed to be explicit in its explanation of what could and could not be concluded from the statistical analysis.
III. The participant notes that some students had not completed the survey instrument, and asked whether these students should be removed from the

set. The final toolkit's supporting documentation would need to include information on data cleaning procedures.

From this information, a preliminary, independently-operating version of the toolkit was created. It was composed of six tabs: a blank tab titled "Raw Data" where users could paste their collected data; a static tab labeled "Cleaned Data" which could sense repeated survey-takers and organize data into "time 1" and "time2", as well as removing extraneous information; a suppressed tab labelled "Filtering" which hosts background procedures; two tabs labelled "Descriptive t1" and "Descriptive t2" displaying descriptive statistics for before and after the intervention; and, finally, an interactive tab labelled "t-test", where users could explore the results of the pre-post t-test. A screen capture of the original descriptive statistics tab can be seen in figure 1.



	CURIOSITY					
	CURIOSITY	Joyous Exploration	Deprivation Sensitivity	Stress Tolerance	Social Curiosity	Thrill Seeking
MEAN	4.97	5.95	5.09	4.29	4.99	4.55
MEDIAN	4.96	6.00	4.80	4.40	5.00	4.60
MODE	4.64	5.20	4.60	4.40	5.20	4.40
MIN	3.92	5.20	4.00	2.60	2.40	2.40
MAX	5.96	7.00	6.60	6.00	7.00	6.00
SKEW	0.13	0.15	0.59	0.04	-0.67	-0.70
KURT	0.40	-1.49	-1.02	-0.92	2.86	2.01
NORMAL?	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE

	CONNECTIONS				
	CONNECTIONS	Integrate Outside Information	Consider Social, Economic, and Environmental Factors	Define Connections	Make Connections within Engineering Design
MEAN	5.28	5.51	5.39	5.63	4.81
MEDIAN	5.44	5.50	5.67	5.67	4.83
MODE	5.50	5.50	5.67	6.33	5.00
MIN	3.44	4.00	3.00	3.00	2.83
MAX	6.89	6.83	7.00	7.00	6.83
SKEW	-0.37	-0.07	-0.60	-1.12	-0.13
KURT	0.47	-0.16	-0.16	1.55	0.45
NORMAL?	TRUE	TRUE	TRUE	TRUE	TRUE

	CREATING VALUE			
	CREATING VALUE	Create Value within Engineering Design	Attitude and Approach Toward Value Creation	Create Value for Others
MEAN	5.04	4.95	5.15	5.08
MEDIAN	5.19	5.20	5.13	5.33
MODE	#N/A	5.20	5.38	5.33
MIN	3.46	3.10	3.75	3.00
MAX	6.64	6.80	6.75	7.00
SKEW	0.23	0.11	0.31	-0.19
KURT	-0.19	-0.09	-0.35	-0.69
NORMAL?	TRUE	TRUE	TRUE	TRUE

Figure 1: Descriptive statistics tab from version 1.1 of the toolkit. Note the drop-down menus (highlighted in purple) which would allow users to view descriptive statistics for subfactors of each C.

A document was also developed to accompany the toolkit. Given the current state of the tool, the document focused on the data analysis only, not the set-up or proper application of the survey instruments in a classroom environment. These sections will be developed later when the toolkit is tested with new participants.

Version 1.1 of the document contained the following sections: Descriptive Statistics, which reviewed interpreting measures of centrality (mean, median, and mode), determining normality (using skewness and kurtosis), and reading histograms; Running a t-Test, which reviewed the selection the user needed to make in the tool, as well as the data the tool provided in the output (pp-values, effect sizes, and power); and Reporting the Results, which provided a template for writing up the results the toolkit gave, referencing specific cells where information could be located. The template was written as follows, with additional marginalia to explain how the bracketed statements should be filled in:

> *"A paired-samples t-test was conducted to compare [C / factor] before and after the intervention. Normality of the continuous outcome variable was checked and found to be within range. There was [a statistically significant difference / a difference approaching statistical significance] in [construct] before and after the intervention (p=[p value]). The effect size for this analysis was η^2 = [effect size], indicating that __% of the difference in [construct] can be explained by the intervention. This is a [small / medium / large] effect size. Post hoc power analysis suggests the test was [adequately powered / underpowered] (1-□ = [power])."*

With this version finalized, the development team moved to the second think-aloud protocol.

Second Think-Aloud

During the second think-aloud protocol, the researcher took a backseat and allowed the participant to self-direct through the analysis process using the toolkit and documentation. This process took approximately 30 minutes, less than half the time of the initial think-aloud.

The purpose of this think-aloud was to gather more granular feedback: understanding pain and pinch points, hearing where the user's eye was drawn when viewing the sheet, observing where the user felt they should be able to interact and where information became too overwhelming, etc. From this process, three additional major pieces of feedback were identified:

1. Lack of familiarity with histograms: While histograms are the standard for visualization data spread in statistics, they were not immediately intuitive to the user. Presenting multiple methods for visualizing/holistically understanding the data could improve this [28].
2. Data Handling "Safe Mode": while the supporting documentation described the assumptions for running a t-test and when to discard data, the tool would still run a t-test with data that did not meet these assumptions. The tool should be edited to check assumptions automatically, and to suppress results for tests run on non-normal data.
3. An "At a Glance" Tab: The t-test tab in version 1.1 only displayed the results of one t-test at a time. While this effectively cut down visual clutter and kept the tab focused, the user complained that running all of the t-tests (a total of 15—3 for Cs overall and 12 for the 3Cs' subfactors) became repetitive. A new tab should be added that will display all t-test results in brief, reducing necessary clicks to fully understand the data.

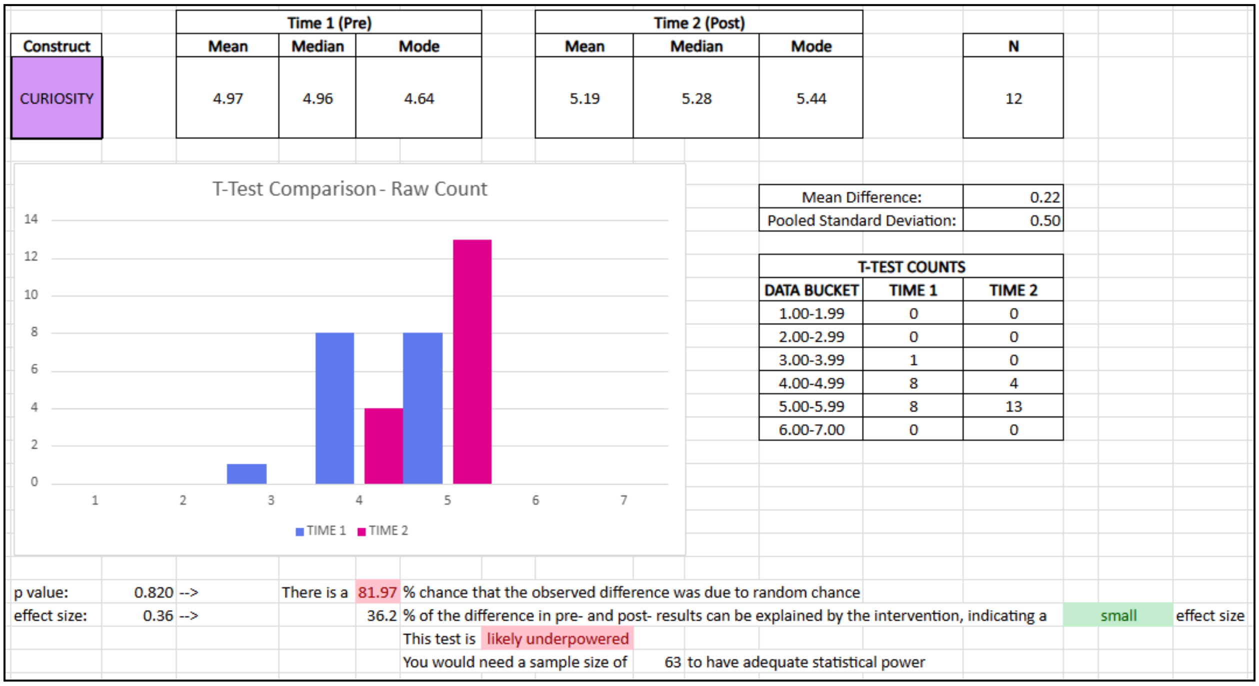
These feedback points were processed into version 1.2 of the toolkit. Feedback point 1 was addressed by adding tables of values accompanying all histogram visualizations. This supported the visual in such a way that made it easy for the user to understand the histogram's visualization, effectively teaching a user to read histograms at a glance. Feedback point 2 was addressed by coding in assumption checks—if the data does not meet an assumption (for example, the data was non-normal), the sheet displays an error message specific to the assumption that was not met (for example, "Error: Data is Non-Normal"). These error messages are explained further in the supporting documentation. Feedback point 3 was addressed by adding a new tab to the tool which displays all possible pre-post t-test results in a table, as seen in figure 2.

CONSTRUCT	Pre-MEAN	Post-MEAN	MEAN DIFFERENCE	POOLED STANDARD DEVIATION	p-VALUE	EFFECT SIZE	POWER
CURIOSITY	4.97	5.19	0.22	0.51	0.820	0.36	likely underpowered
Joyous Exploration	5.95	5.92	-0.04	0.75	0.412	0.13	likely underpowered
Deprivation Sensitivity	5.09	5.20	0.11	0.98	0.624	0.05	likely underpowered
Stress Tolerance	4.29	4.44	0.15	1.36	0.713	0.13	likely underpowered
Social Curiosity	4.99	5.32	0.33	1.00	0.511	0.25	likely underpowered
Thrill Seeking	4.55	5.09	0.54	0.87	0.832	0.62	likely underpowered
CONNECTIONS	5.28	5.46	0.18	0.83	0.823	0.14	likely underpowered
Integrate Outside Information	5.51	5.75	0.24	0.75	0.585	0.25	likely underpowered
Consider Social, Economic, and Environmental Factors	5.39	5.26	-0.13	1.16	0.089	0.14	likely underpowered
Define Connections	5.63	5.74	0.11	1.00	0.710	0.04	likely underpowered
Make Connections within Engineering Design	4.81	5.12	0.31	0.95	0.919	0.24	likely underpowered

Figure 2: A segment of the "At a Glance" t-test results table.

The table employs color-coded cells to indicate mean difference (where a red cell indicates a negative change after the intervention, and a green cell indicates a positive change). It also highlights the cells in the three rightmost columns to indicate statistical significance, practical significance, and power; orange cells represent low significance/power, yellow cells indicate medium significance/power, and green cells (not pictured) indicate high significance/power, while gray cells indicate a result which did not meet established significance/power thresholds [30]. All data in the sample used for the think-aloud protocols met t-test assumptions, and so no results were suppressed.

Figure 3 shows the detailed t-test results tab. Note the inclusion of the table of values beside the histogram, as well as the complete sentences (with color-coding) the sheet generates to aid in proper interpretation of the results.



Construct	Time 1 (Pre) Mean	Time 1 (Pre) Median	Time 1 (Pre) Mode	Time 2 (Post) Mean	Time 2 (Post) Median	Time 2 (Post) Mode	N
CURIOSITY	4.97	4.96	4.64	5.19	5.28	5.44	12

Mean Difference:	0.22
Pooled Standard Deviation:	0.50

T-TEST COUNTS		
DATA BUCKET	TIME 1	TIME 2
1.00-1.99	0	0
2.00-2.99	0	0
3.00-3.99	1	0
4.00-4.99	8	4
5.00-5.99	8	13
6.00-7.00	0	0

p value: 0.820 --> There is a 81.97 % chance that the observed difference was due to random chance

effect size: 0.36 --> 36.2 % of the difference in pre- and post- results can be explained by the intervention, indicating a small effect size

This test is likely underpowered

You would need a sample size of 63 to have adequate statistical power

Figure 3: The detailed t-test output tab for version 1.2 of the toolkit. Note the drop-down menu (highlighted in purple) which allows the user to select which C/factor they are interested in analyzing.

At this point, version 1.2 of the toolkit is ready for another round of think-aloud feedback.

Next Steps

An immediate next step for the toolkit is to recruit other users to test it in think-aloud settings. This would ideally involve users gathering data in their own classrooms, exclusively using guidance from the toolkit documentation to do so; however, to ensure that data is collected properly, the research team has determined it necessary to train undergraduate and graduate teaching assistants in proper use of the survey instrument. This way, a teaching assistant is able to catch and correct any issues in real time, ensuring data is always collected.

Version 2.0 of the toolkit will include an additional statistical test: the ANCOVA (analysis of covariances), as requested by the survey validation team. An ANCOVA analysis is capable of capturing greater nuance in student learning gains [30],[31], and would therefore require more in-depth educational resources to explain proper use [32].

Finally, to ensure dissemination of the toolkit is accessible as possible, the development team aims to distribute it via a faculty development workshop. To properly prepare the toolkit for this audience, a lesson plan and accompanying slide deck will need to be developed. This will hopefully see use beyond workshop settings, and could even be turned into an educational video.

Conclusions

The goal of this project was to develop a toolkit that could support independently-led research projects by technical engineering faculty with little to no experience in statistics. By addressing common barriers such as a lack of time, incentives, and introductory tools, this toolkit serves as a user-friendly gateway into the world of quantitative educational research, specifically through the lens of entrepreneurially-minded learning (EML). While the toolkit is still in its earliest stages of development, the development process has been illuminating; it has revealed the knowledge gaps between those formally trained in educational research and those not formally trained. During the think-aloud protocols, the design team observed the following consumer needs: (1) documentation explicitly guiding users through the process and demonstrating proper vocabulary to aid in interpreting outcomes; (2) methodological safeguards preventing the toolkit from generating results for statistical tests which do not meet assumptions; and (3), intuitive visuals and other ways for users to understand the data “at a glance”.

The current version of the toolkit (version 1.2) provides a foundation for performing descriptive statistics and paired-sample t-tests using validated instruments for the 3Cs. Moving forward, the team will focus on creating version 2.0, which aims to incorporate ANCOVA models for a more nuanced measurement of student learning gains. Future work also includes broader dissemination through faculty development workshops, lesson plans, and educational videos. Ultimately, it is the hope of the design team that this tool will provide faculty with a non-intimidating introduction to statistical analysis, empowering them to independently pursue further educational research and improve student outcomes beyond technical skill acquisition.

References

[1] A. Y. Huang‑Saad, C. S. Morton, & J. C. Libarkin, “Entrepreneurship assessment in higher education: A research review for engineering education researchers,” *Journal of Engineering Education*, vol. 107, no. 2, pp. 263-290, 2018.

[2] KEEN, “*The Framework for Entrepreneurially Minded Learning*,”, Engineering Unleashed, https://engineeringunleashed.com/framework, accessed 4 January 2026.

[3] J. S. London, J. M. Bekki, S. R. Brunhaver, A. R. Carberry, & A. F. McKenna, “A Framework for Entrepreneurial Mindsets and Behaviors in Undergraduate Engineering

Students: Operationalizing the Kern Family Foundation's" 3Cs"", *Advances in Engineering Education*, vol. 7, no. 1, 2018.

[4] D. R. Riley, H. M. Shuster, C. A. LeMasney, C. E. Silvestri, & K. E. Mallouk, "First-year engineering students' conceptualization of entrepreneurial mindset", *Entrepreneurship Education and Pedagogy*, vol. 6, no. 1, pp. 87-109, 20023.

[5] J. L. Gorlewicz, & S. Jayaram, S, "Instilling curiosity, connections, and creating value in entrepreneurial minded engineering: Concepts for a course sequence in dynamics and controls", *Entrepreneurship Education and Pedagogy*, vol. 3, no. 1, pp. 60-85, 2020.

[6] S. Kulturel-Konak, A. Konak, P. A. Alexander, K. Mehta, & D. R. Schneider, "Measuring transformative learning in STEM: validating the transformative learning scale for the innovation mindset (TLSIM) in experiential programs", *Entrepreneurship Education*, pp. 1-30, 2025.

[7] T. B. Kashdan, M. C. Stiksma, D. J. Disabato, P. E. McKnight, J. Bekier, J. Kaji, & R. Lazarus, "The five-dimensional curiosity scale: Capturing the bandwidth of curiosity and identifying four unique subgroups of curious people", *Journal of Research in Personality*, vol. 73, pp. 130-149, 2018.

[8] M. E. Ita, M. E. West, & R. L. Kajfez, "Development of Survey Instruments to Measure Undergraduate Engineering Students' Entrepreneurial Mindset: Connections and Creating Value", *International Journal of Engineering Education*, vol. 39, no. 4, pp. 811–822, 2023.

[9] B. Kotla, T. J. Stump, C. Jenkins, A. D. K. Motlagh, K. M. Kecskemety, & R. L. Kajfez, "WIP: Confirmatory Factor Analysis of Instruments to Measure Connections and Creating Value in First-Year Engineering Students", In *2025 ASEE Annual Conference & Exposition*, 2025.

[10] A. A. Espinosa, "Who owns knowledge?: The fight for open educational research", *Issues in Educational Research*, vol. 35, no. 1, pp. viii-xi, 2025.

[11] J. L. Fleming, S. E. Wilson, S. A. Hart, W. J. Therrien, B. G. Cook, "Open Accessibility in Education Research: Enhancing the Credibility, Equity, Impact, and Efficiency of Research", *Educational Psychology*, vol. 56, no. 2, pp. 110-121, 2021.

[12] L. Rumreich, M. E. West, K. M. Kecskemety, R. L. Kajfez, R. Desing and C. Wallwey, "*EM Learning Outcomes and Direct Assessment Rubrics*", Engineering Unleashed, https://engineeringunleashed.com/card/2441, accessed 4 January 2026.

[13] K. K. Wallman, "Enhancing statistical literacy: Enriching our society", *Journal of the American Statistical Association*, vol. 88, no. 421, pp. 1-8, 1993.

[14] I. Gal, "Statistical literacy: Meanings, components, responsibilities, the challenge of developing statistical literacy", *Reasoning and Thinking*, pp. 47-78, 2004.

[15] A. M. Aziz & R. Rosli, "A systematic literature review on developing students' statistical literacy skills", In *Journal of Physics: Conference Series*, vol. 1806, no. 1, pp. 012102, 2021.

[16] R. Utari, & R. I. I. Putri, “Advancing statistical literacy for prospective teachers: A systematic review of contemporary approaches and best practices”, *Jurnal Elemen*, vol. 11, no. 2, pp. 390-407, 2025.
[17] D. Friansah, R. I. I. Putri, Y. Hartono, & S. Somakim, “The importance of statistical literacy: A systematic literature review”, In *AIP Conference Proceedings*, vol. 3333, no. 1, pp. 020047, 2025.
[18] A. Lateh, “Exploring Statistical Literacy Deficiencies and Causal Associations among Master and Doctoral Graduates in Thailand”, *Pakistan Journal of Life & Social Sciences*, vol. 22, no. 2, 2024.
[19] M. A. Hjalmarson, G. L. Herman, & K. A. Douglas, “Considerations for Engineering Education Research Using Quantitative Methods”, In *International Handbook of Engineering Education Research*, pp. 638-659, Routledge, 2023.
[20] R. K. Henson, D. M. Hull, & C. S. Williams, “Methodology in our education research culture: Toward a stronger collective quantitative proficiency. Educational researcher”, vol. 39, no. 3, pp. 229-240, 2010.
[21] M. A. Earley, “A synthesis of the literature on research methods education. Teaching in Higher Education”, vol. 19, no. 3, pp. 242-253, 2014.
[22] R. Raman, J. Utts, A. I. Cohen, & M. J. Hayat, “Integrating ethics into the guidelines for assessment and instruction in statistics education (GAISE)”, *The American Statistician*, vol. 77, no. 3, pp. 323-330, 2023.
[23] R. K. Henson, D. M. Hull, & C. S. Williams, “Methodology in our education research culture: Toward a stronger collective quantitative proficiency”, *Educational Researcher*, vol. 39, no. 3, pp. 229-240, 2010.
[24] K. C. Mittag & B. Thompson, “A national survey of AERA members’ perceptions of statistical significance tests and other statistical issues”, *Educational Researcher*, vol. 29, no. 4, pp. 14–20, 2000.
[25] D. H. Robinson, J. R. Levin, G. D. Thomas, K. A. Pituch, & S. Vaughn, “The incidence of “causal” statements in teaching-and-learning research journals”, *American Educational Research Journal*, vol. 44, pp. 400–413, 2007.
[26] E. Charters, “The use of think-aloud methods in qualitative research an introduction to think-aloud methods”, *Brock Education Journal*, vol. 12, no. 2, 2003.
[27] G. Raju & S. A. Sorby, “Exploring the Impact of Think-Aloud Protocol in Engineering Design Problem Solving,” In *2025 ASEE Annual Conference & Exposition*, 2025.
[28] C. Rogers-Shaw, D. J. Carr-Chellman, & J. Choi, “Universal design for learning: Guidelines for accessible online instruction”, *Adult Learning*, vol. 29, no. 1, pp. 20-31, 2018.
[29] G. J. Privitera, *Research methods for the behavioral sciences*, Sage Publications, 2024.
[30] Wright, D. B. (2006). Comparing groups in a before–after design: When t test and ANCOVA produce different results. British journal of educational psychology, 76(3), 663-675.

[31] A. Gjana & R. Kosova, “Traditional Class, and Online Class Teaching. Comparing the Students Performance Using ANCOVA”, *Journal of Multidisciplinary Engineering Science and Technology*, pp. 14806-14811, 2021.
[32] J. H. Vargas & J. Z. Peet, The first primer for the QuantCrit‐curious critical race theorist or psychologist: On intersectionality theory, interaction effects, and AN (C) OVA/regression models”, *Journal of Social Issues*, vol. 80, no. 1, pp. 168-217, 2024.