ARTICLE
The Spine Lab: A Short-Duration, Fully-Remote Course-Based Undergraduate Research Experience

Robert J. Wickham[1], Elizabeth A. Genné-Bacon[2], & Michele H. Jacob[3]
[1]Psychology Department, Elizabethtown College, Elizabethtown, PA 17022; [2] Center for Science Education, Department of Medical Education, Tufts University School of Medicine, Boston, MA 02111, [3]Department of Neuroscience, Sackler Biomedical Graduate School, Tufts University School of Medicine, Boston, MA 02111.

Course-based undergraduate research experiences (CUREs) are increasingly common approaches to provide students with authentic laboratory experiences. Typically, CUREs are semester-long, in-person experiences that can be financially and time prohibitive for some institutions, faculty, and students. Here, we developed a short-duration, fully-online CURE, the Spine Lab, to provide an opportunity for students to conduct original research. In this CURE, we focused on synaptic spines in the mammalian brain; synapses are the unit structure that functions in rapid information processing. The students worked together in pairs and as a class to analyze cortical neuron spine density and structural morphology changes between a mouse line with learning impairments (forebrain-specific β-catenin knockouts [β-cat cKOs]) and control (Ctl) littermates. The students showed their results in an online poster presentation. Their findings show that spine density is significantly reduced, while spine structural maturation is unaltered in the β-cat cKO. Defining pathophysiological changes caused by CTNNB1/β-catenin loss-of-function provides important insights relevant to human disorders caused by disruptive mutations in this gene. To assess the benefits of this CURE, students completed a pre- and post-test assessment including a content quiz, STEM identity survey, and a standardized CURE survey. Participation in the Spine Lab correlated with improved content and STEM identity scores, and decreased negative attitudes about science. Moreover, direct comparison to the CURE database reveals that the Spine Lab produces comparable benefits to traditional CUREs. This work as a whole suggests that short-duration, fully-online CUREs can provide benefit to students and could be an inclusive tool to improve student outcomes.

    Key words: CURE; β-catenin; CTNNB1; remote-learning

Course-based undergraduate research experiences (CUREs) are a teaching technique in which students have the opportunity to participate in an authentic research experience within the confines of a formal laboratory course. CUREs are generally defined as projects encompassing elements of discovery (outcome of the experiments will be unknown to both students and the instructor) and broader impact (the experiments are done with the intent to contribute to an actual scientific study; (Auchincloss et al., 2014; Spell et al., 2014). Simply put, CUREs are authentic research experiences that differ from traditional laboratory courses in which experimental outcomes are known. Research shows that CUREs offer many of the benefits of traditional faculty-mentored research experiences, such as increased understanding of content, an increased sense of ownership of student work, increased STEM identity, and increased retention in STEM majors (Bangera and Brownell, 2014).

CUREs can promote inclusive science for a few reasons (Bangera and Brownell, 2014). First, they are a more student-centered teaching practice (which promotes equity in classrooms; Secker, 2002). Second, they help foster a sense of science identity and confidence, which minoritized students often lack compared to non-minoritized peers (Rainey et al., 2018). Third, they can introduce students to research much earlier in their academic careers at a larger scale (all students can do it) compared to traditional mentored research (limited by availability of research mentors). In addition, institutions that do not normally have many research opportunities for students (such as community colleges- where a greater share of minority students are enrolled; American Association of Community Colleges [AACC], 2021; National Center for Education Statistics [NCES], 2021) can still do CUREs.

During the COVID-19 pandemic, most institutions switched to a fully-online or hybrid format, causing significant impediments for access to both faculty-mentored research experiences and laboratory courses. The majority of existing CUREs are designed to be semester-long laboratory experiences, and thus lack the flexibility to be converted for online learning. Importantly, since students of color are disproportionately harmed by the effects of COVID-19 (Tai et al., 2021), more inclusive research experiences are needed.

There is great need for flexible CUREs that are readily accessible and can be used in remote and hybrid courses (Genné-Bacon and Bascom-Slack, 2018). To address this need we have created a short-duration, fully online CURE called the Spine Lab. Because little is known about the efficacy of online CUREs, in addition to assessing student learning, we also sought to test gains in student STEM identity, attitudes about science, experience in various academic skills, and research skills using previously developed instruments (Lopatto and Jaworski, 2018; McDonald et al., 2019). Using these previously developed instruments, we can make direct comparisons of the effectiveness of this short-duration, fully online CURE to

CUREs that have run for much longer periods of time and in-person (e.g., physically in the same room).

Research Question Background

Disruptive mutations in the human gene *CTNNB1,* that codes for β-catenin, cause intellectual disability (ID, IQ < 70; (Dubruc et al., 2014; Wickham et al., 2019). Previous work has demonstrated that mice with genetic manipulations that delete β-catenin (β-cat cKO) display learning impairments compared to littermate controls (Ctl) (Wickham et al., 2019). At the molecular level, β-catenin serves as driver of Wnt-target gene expression, which is important for activity-dependent plasticity and learning(Mulligan and Cheyette, 2012). Additionally, β-catenin links the synaptic adhesion protein N-cadherin to α-N-catenin, helping form a connection to the actin cytoskeleton which aids in synapse stability and plasticity(Yu and Malenka, 2004). While Wnt-gene expression was unexpectedly found to be normal in the β-cat cKO (likely due to compensation by its closely related homologue, γ-catenin), levels of both N-cadherin and α-N-catenin were reduced. In line with reduction of synaptic adhesion protein levels, spine density in layer II/III medial prefrontal cortical pyramidal neurons and CA1 dorsal hippocampus was reduced in the β-cat cKO, relative to control littermates. Thus, the working model is that loss of β-catenin can lead to reductions in synaptic adhesion proteins, which reduces synaptic spine density, causing ID. It should be noted that heterozygous mice might serve as a better model for ID since humans are typically heterozygous for loss-of-function β-catenin mutations.

To provide an opportunity for students to analyze synaptic spines and expand this data, we analyzed images of brain slices in the β-cat cKO and Ctl medial prefrontal cortex to independently confirm the changes in spine density, and importantly, add a new analysis by measuring spine structural morphology . Layer II/III medial prefrontal pyramidal cortical neurons were chosen due to the fact that they are largely glutamatergic and the CAMKII Cre-driver used to remove β-catenin is primarily glutamatergically expressed. In addition, these neurons are mediators of cortico-cortical communication, which is likely important for learning. Synaptic spines display a range of morphology, from immature thin-filopodia to mature mushroom-shaped, with the latter indicative of larger synapses strengthened by activity (Nikonenko et al., 2002; Ackermann and Matus, 2003). Spine morphology can be assessed by counting the number of spines of each category (thin, stubby, mushroom, branched) and dividing it over the total number of spines, yielding the spine fraction for each morphological subtype (Risher et al., 2014). Spine size correlates with both synaptic size (such as the postsynaptic density) and synaptic strength which aids in learning and memory formation(Chen et al., 2007). It should be noted that recent work demonstrates that spine morphology exists more or less on a continuum (Berry and Nedivi, 2017; Ofer et al., 2021), so categorical approaches to classification may be less precise in capturing the dynamic changes in spine morphology. However, categorical approaches may be better suited to the skill and competency level of

introduction to neuroscience students who are early in their academic career.

Here, we describe how we implemented a fully online CURE centering around analysis of spine density and spine maturation in the cortex of Ctl and β-cat cKOs. The outcomes of the CURE were overwhelmingly positive. Students found that spine density was reduced in the β-cat cKO, validating our previous analysis (Wickham et al., 2019). Importantly, students generated new data that demonstrates that spine morphology is unaltered by β-catenin loss. We showed that students increased their STEM identity, improved their mastery of the content (including quantitative analysis and tests for statistical significance, with wide applicability to future work by the students), reduced negative opinions on STEM, and showed similar learning gains across various skills compared to the CURE survey database which is largely comprised of in-person, semester-long experiences. Our data suggests that brief, and even distance-learning, CUREs may also have benefits akin to traditionally delivered CUREs.

Learning Objectives
1. Classify spines based on their morphology (Q1, see **Supplementary Material 2** for questions and grading scheme)
2. Calculate proportion of spine subtypes (spine fraction) and spine density (Q2)
3. Explain the relationship between spine morphology and long-term potentiation and long-term depression (Q3-8)
4. Use spine morphology data to make hypotheses about behavioral phenotypes (Q9,10,13)
5. Reflect on unexpected experimental results and determine nature of error/troubleshoot (Q11, 12)

MATERIALS AND METHODS
Institution and Course Structure
Elizabethtown College is a 4-year primarily undergraduate institution that is 86.2% White, 4.8% Hispanic/Latino, 2.6% Asian, 2.2% Black/African-American, 1.5% multi-racial, 0.1% Native Hawaiian or Pacific Islander, and 1.7% race/ethnicity unknown. Approximately 63% of Elizabethtown College students identify as female and 37% identity as male. Approximately 20% of students are Pell-grant eligible, while 97% of all students that are eligible for some form of financial aid were merit, need, or scholarship-based. Introduction to Neuroscience is a 4-credit course that is required for the Psychology and Neuroscience major, but it is often taken to fulfill a core physical sciences course or counts towards the Psychology minor.

This course met twice a week synchronously via Zoom for 75 minutes per meeting. The instructor was on the Zoom call for the duration of each class and incorporated breakout rooms for group work. The Spine Lab is a required part of the course and counted for 20% of the course grade. This course has no formal laboratory associated with it, so this was the only laboratory activity for the course. Contact hours for this CURE were

approximately 6 hours, with an expected 8 hours of outside of class work (watching pre-recorded lecture (30 minutes), reading lab manual (1 hr), individual annotation (1 hr), group annotation (1 hr), poster work (4 hrs) and poster recording (0.5 hrs). Topics covered prior to the CURE include neuroanatomy, membrane and action potential, synaptic neurotransmission, introduction to pharmacodynamics, motor systems including pyramidal and basal ganglia circuitry, somatosensory system, visual system, hunger and homeostasis, sleep, neural basis of learning and memory, and neurodevelopment. Given the content within the Spine Lab, it was important students had some background in synaptic plasticity (the last two classes). The CURE was offered during weeks 10, 11, and 12 out of a 15-week semester.

This CURE involved a collaboration between teaching and research faculty. The instructor of the course (RJW) taught the course while collecting the data in collaboration with the research faculty (MHJ).

Participants
This pedagogical study was reviewed and approved by the Elizabethtown College Institutional Review Board (IRB) and was qualified for IRB exemption. Student data was collected over one semester (Fall 2020, during the COVID-19 pandemic) of two sections (A and B) of a fully online, partially flipped introduction neuroscience course (NEU 125: Introduction to Neuroscience) taught by the same instructor (RJW). In most cases, classes met in person twice per week, with one of the meetings requiring students to watch a pre-recorded series of mini-lectures (~10 minutes in length) and the subsequent meeting reserved for practicing and discussion of problem sets.

There are no pre-requisites for this course. Each section had 40 students and was comprised of a mixture of majors in occupational therapy (32%), psychology (16%), education (10%), undeclared (7%), communications (7%), biological sciences (5%), music therapy (5%), and other (18%, major not represented more than twice). A majority of students were second-year students (57%, first-years, 16%; third-years,14%, fourth-years, 13%). Over 90% of students identified as White. 77% of the students self-identified their gender as woman, 18% as man, and 4% as non-binary, genderqueer, or gender non-conforming. These racial and gender demographics of students in the class were similar for those who fully completed the CURE survey and were a little more female (77 versus 63) compared to the rest of the institution.

Materials
Materials required are images of synaptic spines with good resolution (see below), Image J software (an open source image processing software), and software to annotate these images. Images used in this study were acquired by transcardially perfusing β-cat cKO and control littermate mice with 4% paraformaldehyde. The brains were processed for spine density analysis via gene gun delivery of micro-tungsten particles coated with the lipophilic fluorescent dye DiI and super-resolution laser-scanning confocal microscopy. Images used in this module (80 x 10 x 100 µm, l x w x d) are freely available upon request, but if one wishes to stain and image for other purposes there are several published protocols available (Risher et al., 2014). The Spine Lab is intended to be a generalizable CURE using similar kinds of spine data, thus alternative approaches for collecting data can be employed as long as there is sufficient resolution and number of spines for analysis. It is common to use a pixel resolution of 1024 x 1024 (Srivastava et al., 2011) for visualization of individual synaptic spines. Typically, images are good enough quality for spine density and classification analyses if the signal to noise ratio is at least 2:1 (Parker et al., 2020); a good rule of thumb is that if you can distinctly see branched spines, including the neck and both heads, then you have achieved the minimum resolution. In the instructor's experience, students become adept at classification of spines after approximately 30-40 spines, so typically images that have at least 30-40 spines are sufficient. In the present study, each student classified around 80-100 spines per dendrite (with one dendrite per image), one from a control mouse and one from a β-cat cKO. ImageJ can be used to adjust contrast to visually aid students in their analysis and for annotation should one wish. For annotation/analysis of the images, Microsoft PowerPoint or Microsoft Paint is sufficient.

General Procedure
Two days prior to the laboratory module, students were instructed to read the laboratory manual (Supplementary Material 1) and watched two brief pre-recorded lectures, both of which outlined the biology behind β-catenin function and a related developmental disorder CTNNB1 Syndrome. Students knew that that they were going to be assessed on their mastery of the concepts in the videos and laboratory manual. Enough detail was provided in these items such that students, in theory, would have sufficient background to answer the questions in the content test (see below, Supplementary Material 2). The lecture and laboratory manual identified that the research the class was doing included both confirmatory and novel aspects that had not been conducted previously (i.e., original research) and emphasized that this work was valuable in understanding the pathophysiology underlying intellectual disabilities caused by CTNNB1/β-catenin inactivating mutations. Afterwards, students conducted the pre-test, STEM identity assessment, and CURE pre-questionnaire. Then, students worked in pairs to classify and count synaptic spines on a roughly 100 µm section of the second dendritic branch of layer II/III medial prefrontal cortex neurons in Ctl and β-cat cKO mice. They created and presented a poster on their findings and reviewed peer presentations. Afterwards, students conducted the post-test, STEM identity assessment, and CURE post-questionnaire. Importantly, this entire experience occurred online, with classes being held synchronously over Zoom.

Assessments- Development and Analyses
Before and after completion of the spine laboratory module, students were asked to complete an online survey in class

	Pre test			Post test		
	Difficulty index	Discrimination index	Point biserial correlation	Difficulty index	Discrimination index	Point biserial correlation
Goal range:	0.3-0.9	≥0.3	≥0.2	0.3-0.9	≥0.3	≥0.2
Item name						
Q1	0.80	0.33	0.35	0.95	0.11	0.17†
Q2	0.20	0.19	0.19†	0.56	0.07	0.13†
Q3	0.67	0.63	0.63	0.73	0.52	0.65
Q4	0.64	0.33	0.42	0.87	0.26	0.56
Q5	0.49	0.89	0.84	0.73	0.41	0.62
Q6	0.64	0.59	0.75	0.78	0.26	0.56
Q7	0.65	0.33	0.54	0.87	0.19	0.56
Q8	0.40	0.52	0.63	0.65	0.48	0.64
Q9	0.38	0.56	0.60	0.49	0.56	0.59
Q10	0.55	0.70	0.70	0.67	0.56	0.67
Q11	0.67	0.48	0.51	0.78	0.26	0.49
Q12	0.87	0.11	0.29	0.96	0.07	0.40
Q13	0.20	0.07	0.19†	0.18	0.04	0.05†
†p-value >0.05						

Table 1: Difficulty index, discrimination index, and point biserial correlation values for pre and post-test items. Note that Q2 and Q13 were removed from the final score (Figure 1) due to low PBS both pre and post-test, low discrimination index, and low difficulty index.

via Microsoft Forms. The survey was voluntary and anonymous, with pre and post scores matched using a unique self-generated code. There were three major elements of these surveys. First, students completed a content test to assess their proficiency in the learning objectives (development of this content test described below). Second, students completed the STEM Professional Identity Overlap measure (STEM-PIO-1; (McDonald et al., 2019). STEM-PIO-1 is a single item assessment designed to estimate STEM identity by asking students how much they feel their personal self-perception overlaps with that of a STEM professional (McDonald et al., 2019). Previous research demonstrates this single-item assessment displays convergent, discriminant, and criterion validity as well as moderate test-retest reliability (r = 0.76) (McDonald et al., 2019). Lastly, students completed a modified version of the Classroom Undergraduate Research Experience (CURE) survey (Lopatto, 2004, 2007; Lopatto and Jaworski, 2018). This is a flexible survey designed to measure student perception of research-like experiences, as well as positive and negative attitudes toward science. We omitted some elements of the survey that were not relevant to our hypotheses, such as questions centered around one's learning style, as well as utilized an abbreviated attitudes survey that included questions that have been shown to factor well with each other (Hoskins et al., 2011).

We included an aggregate data set of 18,062 matched pre/post responses from students at multiple institutions between 2015-2018 that is available to make comparisons between individual CUREs and national benchmarks. Seventy percent of the eighty students (n= 56) completed both the Spine Lab pre and post-lab survey.

Content Test Item Analysis
Test items were generated to align with the learning goals.

Classical test theory was used to assess the validity of the content test. Item analysis on both pre- and post-tests included index of difficulty, discrimination index, and point-biserial correlations for each item (Table 1). For these analyses, item responses were considered correct if students received greater then 50% of possible points, and incorrect if students received less than or equal to 50% of possible points (for example, on a question worth 2 points, responses were considered incorrect if students received a score of 1 point or lower). Total scores used in these analyses were based on the unmodified grades (i.e., on total points received, not number of questions correct).

The index of difficulty measures question difficulty and was calculated for each item by dividing the number of students who got the item correct by the total number of students. A score closer to 0 indicates that the question was challenging, while a score closer to 1 indicates the question was relatively easy. The discrimination index and point-biserial correlation coefficients are both measures of whether high-scoring students are more likely to get an item correct. The discrimination index measures how well each item differentiated between high-scoring students and low-scoring students, with a positive score indicating that more high-scoring students got a question right compared to low-scoring students. To calculate this index, student responses were sorted from lowest to highest total score in Excel and divided into two equal groups of high-scorers and low-scorers (because of the odd number of student responses, one middle-scoring student was left out of this analysis). Discrimination index for each item was calculated by subtracting the number of low-scoring students who got the item correct from the number of high scoring students who got the item correct, and dividing this by the number of students in each group. A positive value means higher-scoring students were more likely to get an item correct than lower-scoring students. Point-biserial
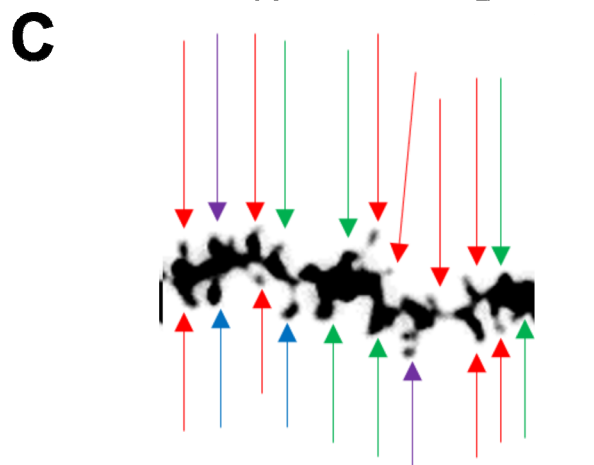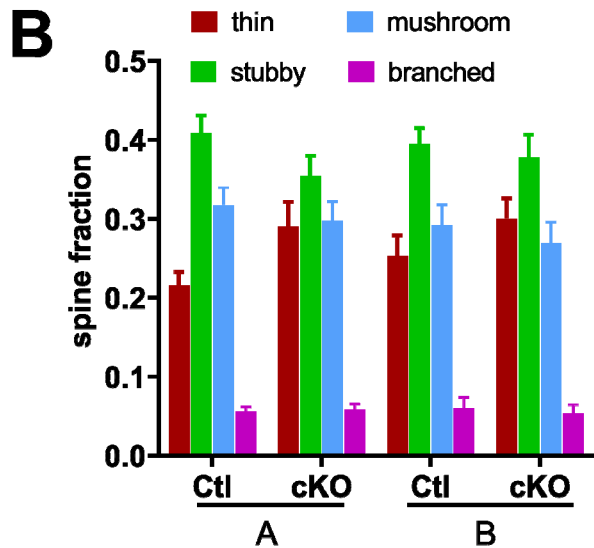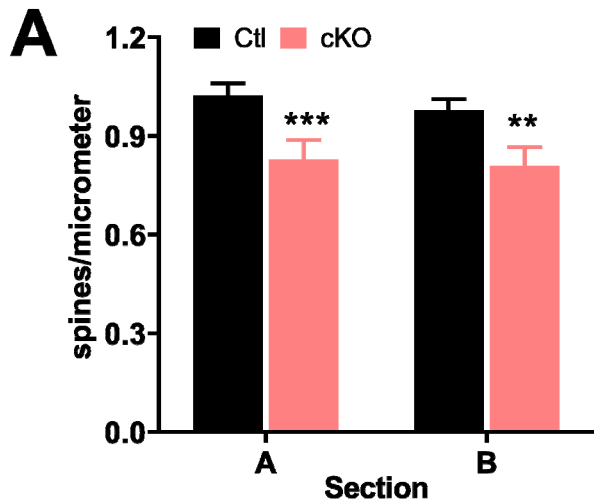
Figure 2. *(A)* Students who took the pre and post survey improved their test scores*)*, had higher STEM identity *(B)*, and had lower negative attitude about science *(C)*. **** $p < 0.0001$; ** $p < 0.01$; Wilcoxin matched-pairs signed rank tests, error bars are standard errors of the mean.

correlations were carried out in SPSS (IBM SPSS Statistics Version 27, Armonk, NY) using a score of 1 for correct and 0 for incorrect for each item (p-values reflect a two-tailed test). Point-biserial coefficients can range from -1.0 to 1.0, with coefficients greater than 0.2 generally considered acceptable. Based on these analyses, items Q2 and Q13 were excluded from further analysis (Supplementary Material 2). Internal reliability of the test was assessed as Cronbach's alpha using SPSS. Alpha was 0.807 for the pre-test and 0.764 for the post-test, before removing problematic items. After removing Q2 and Q13 alpha rose to 0.829 and 0.806 for pre-and post-tests, respectively, indicating a good level of internal reliability.

Pre/Post Test and STEM Identity Statistical Analysis
Scores on the pre and post tests were graded blindly and summed (**Supplementary Material 2**). Items removed from the content test item analysis were removed from this calculation. Since scores did not fall under a normal distribution, a Wilcoxin matched-pairs signed ranks test was performed to compare pre versus post-test performance on the test and STEM identity scores.

Spine Analysis Statistical Analysis
Independent Student's t-tests were used to compare spine density between the Ctl and cKO mice. For spine maturation, a Two-Way mixed-model (within subject variable = spine morphology; between subject variable = genotype) analysis of variance (ANOVA) was employed using SPSS.

Timeline
- Pre-Class 1: Two days prior to Class 1, students were instructed to read a detailed manual as well as view a recorded lecture on CTNNB1 syndrome, behavioral characteristics of the β-catenin cKO

Figure 1. *(A)* Spine density was shown to be reduced in the β-cat cKO in both Section A and B while spine morphology was found to be relatively similar in both sections *(B)*. *(C)* Representative student annotation of image of a 20μm segment of dendrite. Arrows: red = thins, green = stubby, blue = mushroom, purple = branched *** $p < 0.001$; ** $p < 0.01$; Wilcoxin matched-pairs signed rank tests, error bars are standard errors of the mean.
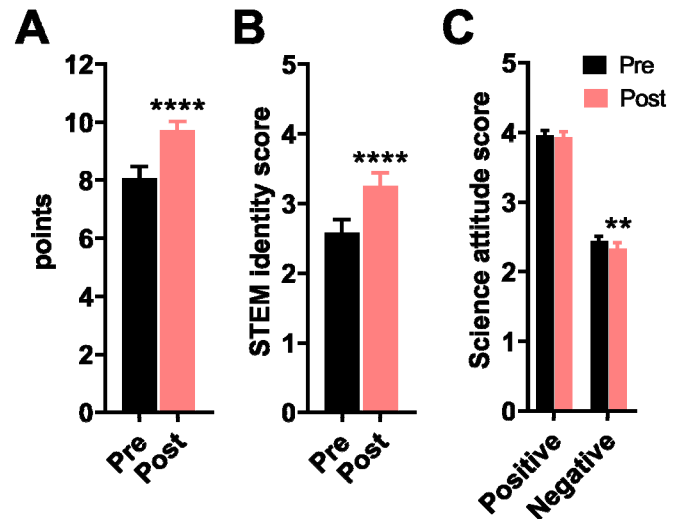
mouse, and an overview of β-catenin functions. Specifically, the lecture and laboratory manual provided an overview of CTNNB1 syndrome, introduced the β-catenin cKO mouse as a model of human CTNNB1 loss of function, and explained the two major functions of β-catenin that might be malfunctioning in the individuals with disruptive CTNNB1 gene mutations: cadherin-based synaptic adhesion and Wnt-target gene expression. Students at this stage had learned about long-term potentiation, an electrophysiological correlate of learning, and that synapses could enlarge (strengthen) or shrink (weaken) through experience and learning. Additionally, the lecture walked students through the general approach our class would use to assess changes in synaptic density and structural morphology. There was content overlap between the manual and the video to help reinforce concepts. The purpose of these items was to prepare students for the content and skills they would need to classify, count, and analyze spines and to help use this data from the cKOs to suggest potential mechanisms underlying CTNNB1 Syndrome.

- Class 1: In class, students conducted the pre-content test, STEM identity assessment, and CURE survey pre-questionnaire. Students individually formed hypotheses about how changes in spine density and/or morphology could lead to learning impairments in the β-cat cKO mouse.

- Class 2: Students were provided a pair of images (labeled A and B) and were instructed to classify and count the number of spines. Spine density was calculated by counting the number of spines on each dendrite for each image and by dividing by the length of the dendritic segment (provided to student). The spine fraction was calculated by counting the number of each spine subtype (thin, mushroom, stubby, or branched shape), and dividing each subtype by the total number of spines. Students then calculated spine density and fraction of mature spines. At the beginning of class up until 30 minutes into class, students were allowed to anonymously submit annotated images via Canvas using a PowerPoint file to seek feedback. Typically, students would place a "?" over the spine. The instructor collected these images and made a separate PowerPoint deck. In the last 15 minutes of class, the instructor then went over each spine to provide both real-time iterative feedback to inquiring student but also extra practice for other students.

- Class 3: Students met with their pre-assigned partners (who had the same set of images) to compare and contrast their analysis. Students re-analyzed their data together, spine by spine compared classification and came to consensus.

Once completed, students submitted their data using a shared Google Sheets document.

- Between Class 3 and 4: The instructor (RJW) averaged spine density and spine fraction (the proportion of mushroom, thin, stubby, across images), performed statistical analyses, and provided students a graph of their results. Analysis was conducted separately for the two sections of the course. Each section had identical image sets, allowing for comparison across sections of the course.

- Class 4: Students were provided a guided handout that allowed for discussion in small groups (groups of four, two pairs of partners) and full class discussion afterwards.

- Between Class 4 and 5: Students created a poster and poster presentation (5-7 minutes) via Zoom and uploaded it to Canvas.

- Class 5: Students watched and commented on two of their peer's videos. The post-test, STEM identity assessment, and CURE pre-questionnaire was taken.

RESULTS

Spine Density

Identical image sets were provided to each section since there were not enough images for each pair to receive unique image sets. We have showed the data from both Section A and B to illustrate replicability across sections. For section A, a total of 2,023 spines across 19 dendrites were analyzed in the Ctl mouse while in section B a total of 1,585 spines across 19 dendrites were analyzed in the cKO mouse. Section A showed reductions in spine density in the cKO compared to the Ctl (Ctl, $M = 1.02$, $SEM = 0.04$; cKO, $M = 0.83$, $SEM = 0.06$; $t(36) = 2.75$, $p = 0.009$; Cohen's $d = 0.89$, Figure 1A) Similarly, Section B showed similar reductions in spine density in the cKO compared to the Ctl (Ctl, $M = 0.98$, $SEM = 0.03$; cKO, $M = 0.81$, $SEM = 0.06$; $t(36) = 2.56$, $p = 0.01$; Cohen's $d = 0.84$, Figure 1B). This data is highly in line with the previous findings of cortical spine density in the Ctl ($M = 1.11$, $SEM = 0.04$) and β-cat cKO ($M = 0.92$, $SEM = 0.02$) from previous work (Wickham et al., 2019).

Spine Morphology

Spine morphology analysis was performed on the same dataset as spine density analysis. For Section A's data (**Figure 1B**), a mixed-model ANOVA (genotype: between subjects factor; spine, within subjects factor) revealed no main effect of genotype ($F(1,36) = 1.38$, $p = 0.25$), while section B (**Figure 1B)** showed similar effects, with no main effect of genotype ($F(1,36) = 1.54$, $p = 0.22$), a significant main effect of spine type ($F(3,108) = 63.9$, $p < 0.0001$), with a trending but non-significant genotype x spine type interaction ($F(3,108) = 2.45$, $p = 0.07$). Section effect of spine type ($F(3,108) = 53.2$, $p < 0.0001$), and a non-significant genotype x spine type interaction ($F(3,108) =0.75$, $p = 0.20$). Since branched spines were not

Item	Spine Lab Mean (SD)	CURE DB (SD)	t-ratio	p-value
Clarification of a career path	2.37 (1.22)	3.13 (1.22)	0.0003	0.0003
Skill in the interpretation of results	3.57 (0.82)	3.7 (0.98)	0.8548	0.8548
Tolerance for obstacles faced in the research process	3.35 (0.87)	3.67 (1.0)	0.8939	0.8939
Readiness for more demanding research	3.30 (1.00)	3.59 (1.04)	0.4207	0.4207
Understanding how knowledge is constructed	3.52 (0.88)	3.6 (1.01)	0.8939	0.8939
Understanding of the research process in your field	3.20 (1.12)	3.66 (1.06)	0.0283	0.0283
Ability to integrate theory and practice	3.37 (1.00)	3.61 (1.01)	0.6343	0.6343
Understanding of how scientists work on real problems	3.64 (0.89)	3.75 (1.00)	0.8704	0.8704
Understanding that scientific assertions require supporting evidence	3.54 (0.99)	3.78 (1.01)	0.6343	0.6343
Ability to analyze data and other information	3.80 (0.87)	3.86 (0.96)	0.8939	0.8939
Understanding science	3.58 (0.83)	3.77 (0.99)	0.8086	0.8086
Learning ethical conduct in your field	2.98 (1.15)	3.38 (1.18)	0.2132	0.2132
Learning laboratory techniques	3.32 (1.00)	3.86 (1.05)	0.0036	0.0036
Confidence in my potential to be a teacher of science	2.69 (1.07)	3.11 (1.27)	0.2873	0.2873
Skill in how to give an effective oral presentation	3.13 (1.06)	3.36 (1.23)	0.8086	0.8086
Skill in science writing	3.13 (1.06)	3.35 (1.13)	0.8086	0.8086
Self-confidence	3.06 (1.16)	3.34 (1.17)	0.6343	0.6343
Understanding of how scientists think	3.27 (0.85)	3.6 (1.05)	0.2873	0.2873
Learning to work independently	3.29 (1.08)	3.48 (1.13)	0.8168	0.8168
becoming part of a learning community	3.44 (0.95)	3.61 (1.09)	0.8423	0.8423

Table 2. Comparison of CURE results on research skills from the Spine Lab to the CURE Database (DB). Items in black indicate no change while items in red indicate lower Spine Lab CURE questionnaire scores relative to the CURE DB.

Item	Spine Lab Mean (SD)	CURE DB Mean (SD)	t-ratio	p-value
a scripted lab or project in which the students know the expected outcome	3.00 (1.08)	3.48 (0.9)	3.80	0.0017
a lab or project in which only the instructor knows the outcome	3.38 (1.07)	3.31 (0.89)	0.60	0.9577
a lab or project where no one knows the outcome	3.55 (1.15)	2.48 (1.09)	6.99	0.0001
at least one project that is assigned and structured by the instructor	3.72 (0.90)	3.66 (0.88)	0.52	0.9577
a project in which students have some input into the research process and/or what is being studied	3.75 (0.93)	2.98 (1.03)	5.50	0.0001
work individually	3.56 (0.96)	3.62 (1.02)	0.41	0.9577
work as a whole class	3.53 (1.00)	3.14 (1.02)	2.81	0.0389
work in small groups	3.73 (1.01)	3.88 (0.75)	1.51	0.5081
become responsible for a part of the project	3.83 (1.09)	3.83 (0.84)	0.03	0.9768
read primary scientific literature	3.37 (0.98)	3.13 (1.06)	1.66	0.4548
collect data	3.91 (0.90)	3.68 (0.87)	1.92	0.3282
analyze data	4.16 (0.83)	3.59 (0.87)	4.88	0.0001
present results orally	3.56 (1.05)	3.15 (1.04)	2.94	0.0288
present posters	3.57 (1.04)	2.9 (1.11)	4.46	0.0001
critique the work of other students	3.40 (0.97)	2.94 (1.03)	3.31	0.0094
listen to lectures	3.67 (1.10)	4.09 (0.83)	3.74	0.0021
read a textbook	2.87 (1.05)	4.07 (0.81)	10.69	0.0001
work on problem sets	3.33 (1.08)	3.88 (0.9)	4.45	0.0001

Table 3. Comparison of CURE results on classroom skills from the Spine Lab to the CURE DB. Items in black indicate no change, items in red indicate lower, while items in green indicate higher Spine Lab CURE questionnaire scores relative to the CURE DB.

distributed normally (but thins, stubbies, and mushrooms were), a Mann-Whitney U test showed no differences between Ctl and cKO mice (Section A: U = 180, p = 0.99; Section B: U = 177, p = 0.94). These data suggest that β-catenin loss does not significantly alter structural maturation of spines, but may be important in the formation of new spines or retention of existing spines.

Pre/Post Content Test
Students had higher test scores in the post-test (*mdn* = 10.0) compared to the pre-test (*mdn* = 8.5) and this difference was statistically significant via a Wilcoxin matched-pairs signed rank test (z = -4.3, $p < 0.0001$, r = 0.41), indicating the laboratory experience increased knowledge of the content (Figure 2A).

STEM Identity

Students had higher STEM identity scores in the post-survey (*mdn = 3)* compared to the pre-survey (*mdn = 2)* and this difference was statistically significant via a Wilcoxin matched-pairs signed rank test ($z = -4.2$, *p* < *0.0001, r = 0.40)*, indicating the laboratory experience helped improve their STEM identity (Figure 2B).

Science Attitudes (CURE Survey)

Positive attitudes about science did not change between the pre-survey (*mdn = 4, SEM = 0.07)* and the post-survey (*mdn = 4, SEM = 0.07, z = -0.16, p = 0.82, r = 0.01)*. In contrast, negative attitudes about science decreased from the pre-survey (*mdn = 2.5, SEM = 0.07)* to the post-survey (*mdn = 2.3, SEM = 0.08, z = -2.72, p = 0.007, r = 0.26)* (Figure 2C).

Comparison of Present CURE to CURE Database

Comparisons of the CURE questionnaire data from the Spine Lab to the CURE database indicate specific areas in which the Spine Lab outperformed, underperformed, and similarly performed relative to traditional CUREs (Tables 2 and 3). The CURE database which is compared to the Spine Lab uses identical scoring methods. For Question 25 (pages 41 and 42), which assess classroom skill gains and Question 26 (pages 43 and 44) which assess research skills, the following scoring method was used: 1 = no gain; 2 = small gain; 3 = moderate gain; 4 = large gain; 5 = very large gain.

DISCUSSION

In this CURE, we provided students with a scientific question, which had real-world implications, in a distance-learning modality over a period of 2-3 weeks. We used a modified survey to assess changes in self-reported class experiences and scientific skills in order to compare the effectiveness of a short-duration, fully-remote CURE. By and large, the self-reported gains in class experiences and research skills suggest that short-duration, fully-remote CUREs might be used to provide benefit to students where traditional semester-long in-person CUREs are not possible. In order to compare the effectiveness of the Spine Lab CURE to more traditional, longer, and in-person CUREs, we examined a repository of CURE data collected from the CURE database (Lopatto and Jaworski, 2018). By comparing the self-reported scores in the Spine Lab CURE to the aggregate dataset, we can get a sense for which areas the Spine Lab under or overperformed against a national benchmark, as well as compare the effectiveness of a short-duration, fully-remote CURE against more traditional, longer and in-person CUREs.

We found it interesting to note where our students made larger and smaller improvements compared to traditional CUREs. Larger improvements in "critiquing the work of other students" is in line with one of the major activities of our CURE, where students have to work in teams to establish the best criteria for spine morphology and density measurement. Having this discussion likely improved this item's score. Other items such as "present posters" and "present results orally" fits with the group

poster activity students completed. We also observed increases in "data analysis," "work as a whole class," and "a project in which students have some input into the research process and/or what is being studied" items/ This increase is consistent with the activities in the CURE where students required input on how data analysis should be carried out as well as the analysis itself from the entire class in order to proceed to the statistical analysis stage. The data from the CURE also supports the idea that students did not know the outcome of the experiment since they scored higher on "a lab or project where no one knows the outcome" while scoring lower on "a scripted lab or project in which the students know the expected results" compared to traditional CUREs. Some of the scores that were lower than traditional CUREs also fit with the way in which this CURE was delivered. For example, students showed smaller gains in "listen to lecture", "work on problem sets", and "read a textbook", which were not major features of this CURE. Lower scores in "learning laboratory techniques" also is in line with this CURE, since the technique employed here, visual quantification and qualitative estimation of morphology, may have been viewed as not a strictly "laboratory technique" since it can be done outside of the laboratory. We also observed lower scores in "understanding of research process in your field," which is likely due to limited attention to the research process as this course is at the introductory level and those topics are typically reserved for research methods courses. Additionally, we also observed lower scores on "clarification of a career path", which might reflect the short-duration of the CURE (e.g., students might not make career decisions based on a 2-3 week module) or that many of the students coming into the class already had their career path established (plurality of students are in the occupational therapy program). One limitation of comparing the Spine Lab to the CURE Database is that the demographics of students in the Spine Lab skew overwhelmingly white and female relative to the demographics of students in the CURE Database which is more diverse and less female-biased.

While the benefit of CUREs is undisputed, the precise parameters that define CUREs are not always agreed upon. A recent classification of CUREs suggests five key components of a CURE, all of which are elements of the Spine Lab CURE (Auchincloss et al., 2014). First, students should be engaged in the scientific process. In the Spine Lab, students developed and tested hypotheses using appropriate scientific methods (spine analysis) and scored higher than traditional CURES on the "a project in which students have some input into the research process and/or what is being studied" element. Second, the outcome of the experiment is unknown. In the Spine Lab, students were blinded to their datasets and the outcome of the experiment was also unknown. Moreover, students scored higher than traditional CURES on "a lab or project where no one knows the outcome". Third, the work has broader impact and utility outside of the classroom. This criterion was met since the data generated by students was both novel and provides valuable disease relevant data. Fourth, there is some form of collaboration either with students or

with the instructor. The Spine Lab was designed for students to work together to build scientific collaboration skills, and the Spine Lab scored higher than the CURE database on "work as a whole class" and "critique the work of other students" while performing similarly to the CURE database on "work in small groups". And fifth, there are elements of iteration. Students in the Spine Lab repeated the analysis in collaboration with a peer as a way to improve reliability but also to mirror the scientific process — engaging in reproducibility and independent verification of data.

In addition to observing these student gains, we also assessed STEM identity. Measures of STEM identity can vary in whether they emphasize identity centrality or identity typicality (McDonald et al., 2019). Centrality is thought to be more of a measure of how something is part of their sense of self. For example, "Science is an important part of who I am" is an example of centrality. In contrast, typicality is thought to be more of how people self-stereotype. For example, "Science is a masculine discipline and therefore women are not meant for it" is an extreme example of negative STEM typicality. For different examples of centrality versus typicality see McDonald et al. (2019). The single-item measure we employed is a better measure for STEM typicality than centrality. Thus, we interpret the increase in STEM identity scores as reductions in negative stereotypes about who is and who is not a scientist. While we do not have sufficient power to separate out STEM identity scores by gender, it is possible that since the vast majority of our students identify as female, they viewed women as scientists as more typical after the CURE. Indeed, having a picture of Dr. Michele Jacob in the laboratory manual was intentional, as this was a way to signal to students that women, too, are influential scientists. Importantly, we wanted to signal that it is not uncommon for women to fit this description.

We also observed decreases in STEM negative attitudes while observing no changes in positive STEM attitudes. It is likely that the lack of increase in STEM positive attitudes results from a ceiling effect. Students reported, on average, a 4/5 on the positive attitude index prior to the CURE. In contrast, students reported an average of 2.5/5 on the negative attitudes index. Since the maximum and minimum scores on these indices are 5 and 1, respectively, there was more room for negative attitudes to decrease than there was room for positive attitudes to increase. An alternative explanation may be that there could be nonlinearity in the scale. For example, it might be harder for a student to go from a 4 to a 5 than a 3 to a 4. It is likely that longer exposure to research might drive scores even higher. Decreases in STEM negative attitude were small (0.2 pts on a 1-5 scale, with a small effect size, $r = 0.26$), which may be a reflection that science attitudes might not change over a short-period of time. Longer CUREs may have an advantage in having more time for attitudes to change to a greater extent (Shaffer et al., 2014).

Students also made significant learning gains in content. We intentionally wanted to provide students an opportunity to have learned this material independently (through the lab manual and pre-recorded video) to directly assess the learning gains made from the CURE itself. For example, it wouldn't mean much if we gave the pre-test at the beginning of the course as students may have picked up other skills or content mastery between the pre-test and CURE. The CURE was in the later third of the semester, so it was important that the CURE start immediately after the pre-test. Thus, in theory, some students could have received perfect scores on the content pre-test. Indeed, this did happen on one occasion.

Two questions in the content test were omitted from the final scoring due to low pre/post discrimination score and high difficulty level. Inclusion of these scores do not change the outcome of increased performance on the test (Supplementary Material 3, Supplementary Figure 1). Question 2 was omitted due to its poor PBS, difficulty, and discrimination scores, but analysis at the item level shows that students did dramatically increase their performance on this question (pre: $M = 0.35$, $SEM = 0.05$; post: $M = 0.74$, $SEM. = 0.05$, $p < 0.0001$, Supplementary Material 3, Supplementary Figure 2), suggesting that learning objective 2 was met. Question 13 was also omitted due to its PBS, difficulty, and discrimination scores. The reason for these low scores is likely due to the poor wording of the question. For example, many of the answers refer to "they", but it is unclear if "they" refers to NMDA antagonist injected mice or not (Supplementary Material 2). Question 13 was initially used to assess learning objective 4. Students did increase their overall scores in this learning objective (Question 9 and 10, pooled) although this was only trending ($p = 0.09$; Supplementary Material 3, Supplementary Figure 2).

The Spine Lab CURE differs from traditional CUREs in duration and via a distance-learning delivery. However, the Spine Lab also differs in how content is typically delivered compared to traditional CUREs. For example, traditional CUREs are often facilitated in standalone laboratories that may or may not be accompanied by a lecture component. By contrast, the Spine Lab is "embedded" into the lecture component of a class and occurs roughly two-thirds into the lecture as opposed to starting at the beginning of the course. It is possible that student completion of course content prior to the CURE influenced our outcome measures. For the test, STEM identity, and science attitude assessments, baseline outcome measures could be influenced by class content in the first two-thirds of the course. For example, knowledge of concepts taught prior to the Spine Lab, such as long-term potentiation, spine morphology, and neuroanatomy, likely influenced test scores at baseline. However, we tested students a second time after the Spine Lab on these measures which allowed for us to capture these outcome gains from their research experience. CURE gains in research and classroom skills might be more sensitive to the Spine Lab's embedded nature, since these scores are collected as perceived gains and do not have a corresponding pre-CURE baseline (Lopatto, 2004, 2007; Lopatto and Jaworski, 2018). For the most part, the gains made in research and classroom skills in the Spine Lab seem to be specific to activities in the Spine Lab that did

not occur in the first two-thirds of the class. For example, all of the gains from the Spine Lab that were higher than the CURE database scores were competencies that were not focused on during the first two-thirds of the class (e.g., poster presentations, critiquing work of other students, etc.), suggesting that these gains are likely Spine Lab specific.

Holistically, we believe our data suggest that distance learning CUREs, and even brief experiences, can lead to improvements in scientific knowledge, higher STEM identity, and reductions in negative STEM attitudes. Students, indeed, not only discovered a new finding (spine morphology and maturation is not dependent on β-catenin levels) but validated the reductions in spine density previously found within this dataset. Having students replicate a prior analysis was by design: if the results were similar then the validity of the unknown outcome (spine morphology) can be better trusted.

Short-duration, online CUREs like the Spine Lab may further the goals of inclusive science education, by increasing equity in access to research opportunities. The Spine Lab itself is not unusually inclusive or promoting equity by itself, but by creating and disseminating an easy to use, online/hybrid ready CURE, we are creating the opportunity for more students to access CUREs, which overall promotes equity and inclusivity in science. This would be a very easy CURE to implement at schools that have avoided utilizing CUREs due to cost or resource access.

There are many ways to adapt this CURE for individual needs. Datasets like ours (free to use upon request via e-mail to wickhamr@etown.edu), are likely relatively common within most neuroscience research programs, especially R1 research institutions. Further, this CURE could be modified to include skills in image processing, such as using ImageJ to measure intensity of staining for selected proteins. Moreover, the investigators measured the length of the dendrite for each student so that the focus of the analysis was more on classification and identification of dendritic spines. However, the rigor of the research experience can be increased by having students become more familiar with imaging software (such as ImageJ) to measure the dendritic length themselves. Forming collaborations between the research faculty and the teaching faculty can be mutually beneficial by providing the former with new analyses and the latter with new questions to bring to students.

REFERENCES

Ackermann M, Matus A (2003) Activity-induced targeting of profilin and stabilization of dendritic spine morphology. Nat Neurosci 6:1194-1200.

American Association of Community Colleges (2021) Fast Facts. Washington, DC: AACC. Available at https://www.aacc.nche.edu/research-trends/fast-facts/.

Auchincloss LC, Laursen SL, Branchaw JL, Eagan K, Graham M, Hanauer DI, Lawrie G, McLinn CM, Pelaez N, Rowland S, Towns M, Trautmann NM, Varma-Nelson P, Weston TJ, Dolan EL (2014) Assessment of course-based undergraduate research experiences: a meeting report. CBE Life Sci Educ 13:29-40.

Bangera G, Brownell SE (2014) Course-Based Undergraduate Research Experiences Can Make Scientific Research More Inclusive. CBE—Life Sciences Education 13:602-606.

Berry KP, Nedivi E (2017) Spine Dynamics: Are They All the Same? Neuron 96:43-55.

Chen LY, Rex CS, Casale MS, Gall CM, Lynch G (2007) Changes in synaptic morphology accompany actin signaling during LTP. J Neurosci 27:5363-5372.

Dubruc E, Putoux A, Labalme A, Rougeot C, Sanlaville D, Edery P (2014) A new intellectual disability syndrome caused by CTNNB1 haploinsufficiency. Am J Med Genet A 164A:1571-1575.

Genné-Bacon EA, Bascom-Slack CA (2018) The PARE Project: A Short Course-Based Research Project for National Surveillance of Antibiotic-Resistant Microbes in Environmental Samples. Journal of Microbiology and Biology Education 19(3):19.3.97.

Hoskins S, Lapatto D, Stevens L (2011) The C.R.E.A.T.E. Approach to Primary Literature Shifts Undergraduates' Self-Assessed Ability to Read and Analyze Journal Articles, Attitudes about Science, and Epistemological Beliefs. CBE Life Sci Educ 10:11.

Lopatto D (2004) Survey of Undergraduate Research Experiences (SURE): First Findings. Cell Biology Education 3:270-277.

Lopatto D (2007) Undergraduate Research Experiences Support Science Career Decisions and Active Learning. CBE—Life Sciences Education 6:297-306.

Lopatto D, Jaworski L (2018) CURE Benchmarks 2015-2018. Grinnell, IA: Grinnell College. Available at: https://www.grinnell.edu/sites/default/files/docs/2019-07/CUREBenchmarkStatistics2015-2108.pdf.

McDonald MM, Zeigler-Hill V, Vrabel JK, Escobar M (2019) A Single-Item Measure for Assessing STEM Identity. Frontiers in Education 4(78).

Mulligan KA, Cheyette BN (2012) Wnt signaling in vertebrate neural development and function. J Neuroimmune Pharmacol 7:774-787.

National Center for Education Statistics (2021) Characteristics of Postsecondary Students. Washington, DC: Institute of Education Sciences and U.S. Department of Education. Available at https://nces.ed.gov/programs/coe/indicator/csb.

Nikonenko I, Jourdain P, Alberi S, Toni N, Muller D (2002) Activity-induced changes of spine morphology. Hippocampus 12:585-591.

Ofer N, Berger DR, Kasthuri N, Lichtman JW, Yuste R (2021) Ultrastructural analysis of dendritic spine necks reveals a continuum of spine morphologies. Dev Neurobiol.81(5):746-757.

Parker EM, Kindja NL, Cheetham CEJ, Sweet RA (2020) Sex differences in dendritic spine density and morphology in auditory and visual cortices in adolescence and adulthood. Sci Rep 10:9442.

Rainey K, Dancy M, Mickelson R, Stearns E, Moller S (2018) Race and gender differences in how sense of belonging influences decisions to major in STEM. International Journal of STEM Education 5:10.

Risher WC, Ustunkaya T, Singh Alvarado J, Eroglu C (2014) Rapid Golgi analysis method for efficient and unbiased classification of dendritic spines. PLoS One 9:e107591.

Secker Cv (2002) Effects of Inquiry-Based Teacher Practices on Science Excellence and Equity. The Journal of Educational Research 95:151-160.

Shaffer CD et al. (2014) A Course-Based Research Experience: How Benefits Change with Increased Investment in Instructional Time. CBE Life Sci Educ 13:111-130.

Spell RM, Guinan JA, Miller KR, Beck CW (2014) Redefining Authentic Research Experiences in Introductory Biology

Laboratories and Barriers to Their Implementation. CBE Life Sci Educ 13:9.

Srivastava DP, Woolfrey KM, Penzes P (2011) Analysis of dendritic spine morphology in cultured CNS neurons. J Vis Exp:e2794.

Tai DBG, Shah A, Doubeni CA, Sia IG, Wieland ML (2021) The Disproportionate Impact of COVID-19 on Racial and Ethnic Minorities in the United States. Clin Infect Dis 72:703-706.

Wickham RJ, Alexander JM, Eden LW, Valencia-Yang M, Llamas J, Aubrey JR, Jacob MH (2019) Learning impairments and molecular changes in the brain caused by beta-catenin loss. Hum Mol Genet 28:2965-2975.

Yu X, Malenka RC (2004) Multiple functions for the cadherin/catenin complex during neuronal development. Neuropharmacology 47:779-786.

Address Correspondence to: Robert J. Wickham, Ph.D., Psychology Department, Elizabethtown College, Elizabethtown, PA 17022. Email: wickhamr@etown.edu