TLEF Project – Final Report

Report Completion Date: (2021/04/01)

1. PROJECT OVERVIEW

1.1. General Information

Project Title:	Experiential Data Science for Undergraduate Cross-disciplinary Education (EDUCE)		
Principal Investigator:	Steven Hallam		
Report Submitted By:	Stephan Koenig		
Project Initiation Date:	2017/04/01	**Project Completion Date:**	2022/03/01
Project Type:	☒ Large Transformation ☐ Small Innovation ☐ Flexible Learning ☐ Other: [please specify]		

1.2. Project Focus Areas – *Please select all the areas that describe your project.*

☒ Resource development (e.g. learning materials, media)

☐ Infrastructure development (e.g. management tools, repositories, learning spaces)

☐ Pedagogies for student learning and/or engagement (e.g. active learning)

☐ Innovative assessments (e.g. two-stage exams, student peer-assessment)

☒ Teaching roles and training (e.g. teaching practice development, TA roles)

☒ Curriculum (e.g. program development/implementation, learning communities)

☐ Student experience outside the classroom (e.g. wellbeing, social inclusion)

☐ Experiential and work-integrated learning (e.g. co-op, community service learning)

☐ Indigenous-focused curricula and ways of knowing

☐ Diversity and inclusion in teaching and learning contexts

☒ Open educational resources

☐ Other: [please specify]

1.3. Final Project Summary

The Experiential Data Science for Undergraduate Cross-disciplinary Education (EDUCE) project has made exceptional progress over the past 4 years (3 years relevant TLEF funding) in developing a uniform, cross-disciplinary, collaborative, experiential learning framework to equip undergraduate students in the life sciences with basic competency and literacy in data science. EDUCE has reached key milestones and exceeded expectations while developing a community of practice focused on data science education that presents a new model for integrated teaching in higher education. EDUCE has produced numerous open educational resources and disseminated results at national conferences and one publication.

We have successfully developed a modular and progressive data science curriculum for our original target courses (MICB 301, 405, 425) as well as added several additional courses (MICB 322, 323, 421, 447) based on interactions with faculty in the Microbiology & Immunology Department (MBIM). Our curricular modules have been fully integrated with co-curricular workshops sponsored by ECOSCOPE, an NSERC industrial stream Collaborative Research and Training Experience (CREATE) training program. We have witnessed increased undergraduate enrollment in these workshops correlating in time with the delivery of classroom learning modules. Workshops also acted as effective testing grounds for new curricular modules enabling EDUCE to work iteratively with new faculty interested in bringing more data science into existing courses without disrupting established course structures or developing entirely new untested courses. In addition, during 2020/2021, we were part of the core group to pivot MICB421/447 from a lab-based to a bioinformatics course using existing and newly created EDUCE modules. We also implemented EDUCE modules at BCIT in the 2nd and 3rd year UBC/BCIT Biotechnology program to harmonize the data science content across institutions.

Koenig (and before Dill-McFarland) and EDUCE TAs in collaborations with ASDa and ECOSCOPE led the improvement and integration of ECOSCOPE workshops with EDUCE curricula. Integration of six existing R workshops is complete, including using microbiologically relevant data, timing workshops to coincide with course modules, and aligning workshop learning objectives to overlap with and build upon classroom instruction. In addition, new workshops in phylogenetic analyses, visualization of multi-omics data and amplicon analysis in R was developed with ECOSCOPE.

Impact metrics (to-date Feb 2022):
- Number of students exposed to EDUCE modules: 1,637
- Number of class instructional hours: 148
- Number of workshop hours: 244

From the standpoint of integrated teaching models, our interactions with numerous departments in the Faculty of Science (FoS) have enabled us to cultivate a diverse pool of highly qualified teaching assistants (TAs) that span STEM fields. These interdisciplinary TAs were fully integrated into the development and deployment of curricular modules and participated as peers in the teaching process. Unlike conventional TAs, EDUCE TAs were not assigned to individual courses. Instead, they worked as a team across the full range of teaching modules and co-curricular activities supported by the EDUCE community of practice.

As testimony to this alignment and long-term trajectory of EDUCE, Hallam, Dill-McFarland (former TLF), and Koenig (current TLF) have been active participants in the FoS Data Science Curriculum Committee administered by Sara Harris, Associate Dean Academic in FoS. This committee has been developing a plan for improving data science education at the undergraduate level, including integrating the EDUCE model within the larger FoS strategic plan for data science education. EDUCE provides a life science perspective and vital context information

for the development of data science course modules calibrated with co-curricular activities within a community of practice model that emphasizes collaborations at the individual, departmental and faculty levels of organization. This model has since informed the development of course-based undergraduate research experiences (CUREs) in the Department of Microbiology and Immunology leading to the development of new active learning courses with a data science focus and reformulation of core curriculum relevant to second, third and fourth year courses.

1.4. Team Members – *Please fill in the following table and include **students**, undergraduate and/or graduate, who participated in your project.*

Name	Title/Affiliation	Responsibilities/Roles
Steven Hallam	Professor, UBC	Director of ECOSCOPE MICB 405 and 425 instructor
Kim Dill-McFarland	PostDoc, UBC	Teaching and Learning Fellow (TLF)
Stephan Koenig	PostDoc, UBC	TLF
Jeff Miller	CTLT, UBC	TLEF program consultant
Ashley Welsh	Skylight, UBC	Assessment consultant
Eric Jacindu	Skylight, UBC	Assessment consultant
Gaby Cohen Freue	Professor, UBC	Statistics consultant
Lindsay Eltis	Professor, UBC	MICB 301 instructor
William Mohn	Professor, UBC	MICB 301 instructor
Jennifer Sibley	Lecturer, UBC	MICB 322 instructor
Marcia Graves	Professor, UBC	MICB 323 instructor
Martin Hirst	Professor, UBC	MICB 405 instructor
Sean Crowe	Professor, UBC	MICB 425 instructor
David Oliver	Professor, UBC	MICB 421 and 447 instructor
Evelyn Sun	PostDoc, UBC	MICB 421 instructor
Tiffany Timbers	Professor, UBC	Consultant
Patrick Walsh	Professor, UBC	Consultant
Florent Mazel	PostDoc, Botany, UBC	Workshop Coordinator
Lisa McEwen	PhD student, Medical Genetics, UBC	Teaching Assistant (TA) 2017 (2 terms)
Kris Hong	BSc student, Statistics, UBC	TA 2017 (2 terms)
Julia Beni	PhD student, Environmental Engineering, University of Minnesota—Twin Cities	TA 2017 (2 terms)
David Yin	BSc student, Computer Science and Statistics, UBC	TA 2018 (1 term)
Nolan Shelley	PhD student, Botany, UBC	TA 2018 (1 term)
Yue Liu	MSc student, Applied Mathematics, UBC	TA 2018 (2 terms)

Andrew Wilson	BSc student, Microbiology and Immunology, UBC	TA 2019 (2 terms)
Julia Anstett	PhD student, Microbiology and Immunology, UBC	TA 2019 (2 terms)
Michelle Kang	MSc student, Bioinformatics, UBC	TA 2019–20 (3 terms)
Gil Henriques	PhD student, Zoology, UBC	TA 2020–21 (3 terms) Workshop coordinator
Avery Noonan	PhD student, Microbiology and Immunology, UBC	TA 2020 (1 term)
Ryan Karimi	BSc student, Chemistry, UBC	TA 2020 (1 term)
Cathy Yan	BSc student, Microbiology and Immunology, UBC	TA 2020 (2 terms)
Pranav Sampara	PhD student, Civil Engineering, UBC	TA 2020 (1 term)
Resmi Radhamony	MSc student, Forestry, UBC	TA 2020 (1 term)
Jonah Lin	BSc, Computer Science	Directed studies—Course compiler
Mohammad Najjarzadegan	PhD student, Electrical Engineering, UBC	Center for the Integration of Research, Teaching, and Learning (CIRTL) Teaching-as-Research (TAR) student
Andrew Li	MSc student, Psychology, UBC	TA 2021–22 (2 terms) Workshop coordinator

1.5. Courses Reached – *Please fill in the following table with **past**, **current**, and **future** courses and sections (e.g. HIST 101, 002, 2017/2018, Sep) that have been/will be reached by your project, including courses not included in your original proposal (you may adapt this section to the context of your project as necessary).*

Course	Section	Academic Year	Term (Summer/Fall/Winter)
MICB 301	1	2017/18	Sep
MICB 405	1	2017/18	Sep
MICB 421	1	2017/18	Jan
MICB 425	1	2017/18	Jan
MICB 301	1	2018/19	Sep
MICB 322	1	2018/19	Sep
MICB 405	1	2018/19	Sep
MICB 447	1	2018/19	Sep
BIOL 436	1	2018/19	Sep
MICB 425	1	2018/19	Jan
MICB 301	1	2019/20	Sep
MICB 405	1	2019/20	Sep
MICB 421	1	2019/20	Jan
MICB 425	1	2019/20	Jan

MICB 301	1	2020/21	Sep
MICB 322	1	2020/21	Sep
MICB 405	1	2020/21	Sep
MICB 447	1	2020/21	Sep
MICB 421	1	2020/21	Jan
MICB 425	1	2020/21	Jan
MICB 322	1	2021/22	Sep
MICB 405	1	2021/22	Sep
MICB 447	1	2021/22	Sep

2. OUTPUTS AND/OR PRODUCTS

2.1. *Please **list** project outputs and/or products (e.g. resources, infrastructure, new courses/programs). Indicate the current location of such products and provide a URL if applicable.*

Product(s)/Achievement(s):	Location:
Workshops material	https://educe-ubc.github.io/workshops/
EDUCE website	https://educe-ubc.github.io
Course compiler for EDUCE material	https://educe-ubc.github.io/compiler.html
Data manipulator to explore microbiome sequence data	https://educe-ubc.github.io/manipulator.html
Videos of fundamental concepts and command line tools	https://educe-ubc.github.io/setup_video.html https://educe-ubc.github.io/cmd_line_video.html
R package with interactive learning materials	https://github.com/EDUCE-UBC/educer
Course websites	https://educe-ubc.github.io/MICB405 https://educe-ubc.github.io/MICB425

2.2. Item(s) Not Met – *Please list intended project outputs and/or products that were not attained and the reason(s) for this.*

Item(s) Not Met:	Reason:

3. PROJECT IMPACT

3.1. Project Impact Areas – *Please select all the areas where your project made an impact.*

☒ Student learning and knowledge

☒ Student engagement and attitudes

☒ Instructional team-teaching practice and satisfaction

☐ Student wellbeing, social inclusion

☐ Awareness and capacity around strategic areas (indigenous, equity and diversity)

☐ Unit operations and processes

☐ Other: [please specify]

3.2. What were you hoping to change or where were you hoping to see an impact with this project? – *Please describe the intended benefits of the project for students, TAs, instructors and/or community members.*

The EDUCE project aimed to:

- update curriculum to provide logical and progressive development of computational and analytical confidence and competency for students in 3rd and 4th-year MICB courses.
- design modular content for easy customization to different learning environments.
- Provide content as open education resources.
- offer optional co-curricular workshops for students to deepen their data science skills.
- fully integrate TAs in the EDUCE community of practice, including training in curriculum design and implementation, assessment methods, and communication with diverse audiences, particularly novice students.
- provide TAs with opportunities to lead introductory R workshops, thereby obtaining hands-on teaching experience with guided feedback from the TLF and students.
- establish co-teaching in courses receptive to EDUCE modules, thus exposing research and teaching faculty to the integrated EDUCE model and making co-instructors part of the EDUCE community of practice.
- Promote data science education within MBIM aligned with the UBC Strategic Plan to promote project sustainability.

3.3. Were these changes/impacts achieved? How do you know they occurred? – *How did you measure changes/impacts? (e.g. collected survey data, conducted focus groups/interviews, learning analytics, etc.) Describe what was learned from this process. You are encouraged to include copies of data collection tools (e.g. surveys and interview protocols) as well as graphical representations of data and/or scenarios or quotes to represent and illustrate key themes.*

Course modules were being assessed by pre- and post-course surveys, created by Dill-McFarland. These surveys assessed student:

- understanding of the terms “data science” and “bioinformatics.”

- interest and previous experience in data-science programs (i.e., command line, R) and related fields (i.e. bioinformatics, statistics, computer science).

- interest and intentions to pursue further data science courses and/or co-curriculars.

848 pre-course responses and 808 post-course responses have been collected. In-depth analysis of the first EDUCE course (MICB 301) for academic years 2017 and 2018 (see Dill-McFarland et al. 2021) indicated limited familiarity with data science among MBIM undergraduate students prior to EDUCE modules. After course modules, students showed increased interest and experience in data science topics like “bioinformatics” and “computer science” (see publication for further details). Below is the updated figure for the three core EDUCE courses (MICB 301, 405 and 425) for academic years 2017–2021. Overall, students reported an increase in experience in all three domains (bioinformatics, computer science and statistics) but no change in interest. Thus, EDUCE demonstrated achievement of one of our fundamental goals: to expose undergraduate students to introductory data science within the context of current MICB courses and curriculum.

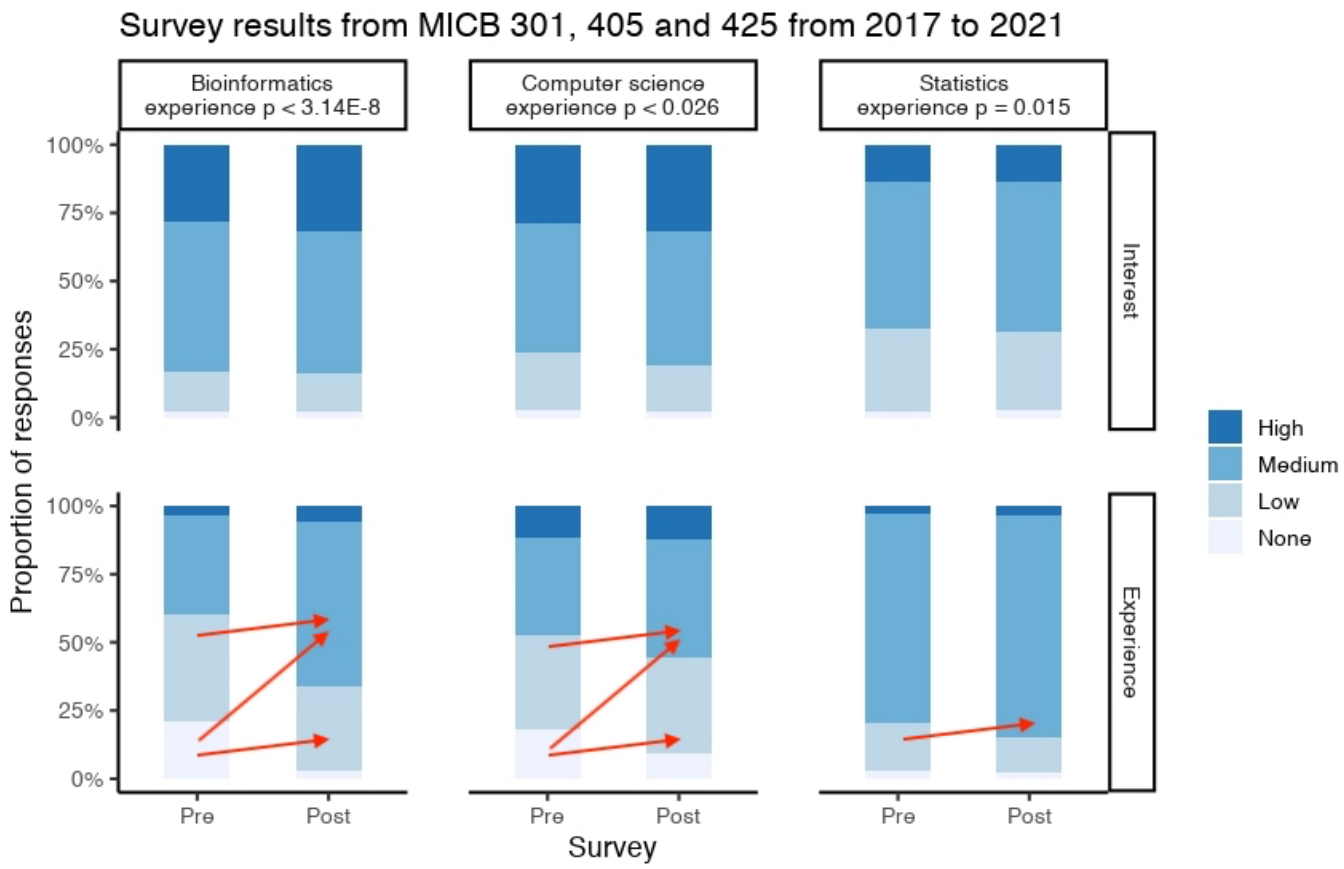

Figure 1 Student interest and experience in data science. Student responses from matched pre- and post-surveys in MICB 301, 405 and 425 indicating self-reported interest (N = 513–520) and experience (N = 459–502) in data science areas including bioinformatics and computer science. Due to changes in survey format, numerical responses from 2018 to 2021 were converted to match categorical groups from 2017 to 2018 as none (0), low (1–3), medium (4–7), and high (8–10). Arrows and *p*-values indicate significant response changes by Monte Carlo symmetry tests for paired contingency tables.

In an effort to increase undergraduate participation in workshops, they were advertised to undergraduates by announcements, hand-outs, flyers, and course website posts (i.e. Canvas). In addition, undergraduate student registration was subsidized by ECOSCOPE. Our efforts have increased undergraduate participation in workshops from 0% (2016/17) to 7% (2017/18), 23% (2018/19), 22% (2019/20), and 33% (2020/21) of participants. Course survey responses further indicate that more than 75% of MBIM undergraduates are interested in co-curricular workshops.

3.4. Dissemination – *Please provide a list of **past** and **upcoming** scholarly activities (e.g. publications, presentations, invited talks, etc.) in which you or anyone from your team have shared information regarding this project. Be sure to include author names, presentation title, date, and presentation forum (e.g., journal, conference name, event).*

Presentations by Dill-McFarland
(Unless otherwise noted, available at https://github.com/EDUCE-UBC/presentations):
- 2018 TLEF Showcase (poster)
- 2018 CTLT Spring Institute (talk)
- 2018 American Society for Microbiology Microbe Meeting (Atlanta, GA, USA) (poster)
- 2018 UBC Jupyter Day, hosted by MATH (talk) https://github.com/patrickwalls/jupyterday2018

- 2018 M&I Faculty retreat (talk)
- 2019 M&I seminar series (talk)
- 2019 American Society for Microbiology Microbe Meeting (San Francisco, CA, USA) (talk)
--- Chair symposium entitled "Incorporating Data Science and Reproducible Research Into Your Practice and Teaching"

Presentations by Koenig
- 2021 M&I symposium (talk)

Publications
Dill-McFarland KA, König SG, Mazel F, Oliver DC, McEwen LM, Hong KY, Hallam SJ. 2021. An integrated, modular approach to data science education in microbiology. Plos Comput Biol 17:e1008661.

4. TEACHING PRACTICES – *Please indicate if **your** teaching practices or those of **others** have changed as a result of your project. If so, in what ways. Do you see these changes as sustainable over time? Why or why not?*

The modular format of the EDUCE material reinforced best practices in curriculum design. Learning outcomes were defined around the project-based final assignments with modules providing the scaffolding in the form of hands-on tutorials, quizzes and smaller assignments.

Live coding, i.e., developing code and walking students through it during class, has been a valuable experience. Invariably, you made mistakes as you go along, offering excellent opportunities to show students how to troubleshoot problems, interpret error messages, look up documentation, and use internet searches. These steps demonstrate that learning is an active process and experience but require the instructor to be comfortable to sometimes struggle in front of a classroom.

Providing EDUCE material as open educational resources (OERs) has also promoted reflection on how to provide access to and structure learning materials to students best and make it easier for instructors and contributors to update the material.

The use of EDUCE material in a CURE course showed that its integration was more straightforward with an active learning-based than a lecture-based course. The vision of the EDUCE program to promote team-teaching was best realized in the CURE course, i.e., members were specialized in different domains but acted as a unit to guide groups of students in their projects. In this way, we created a shared community of practice among instructors, TAs and students.

5. PROJECT SUSTAINMENT – *Please describe the sustainment strategy for the project components. How will this be sustained and potentially expanded (e.g. over the next five years). What challenges do you foresee for achieving the expected long-term impacts listed above?*

The EDUCE modules were created as OERs available for adoption into courses by instructors. Most notably, during 2020/2021, we were part of the core group to pivot two CUREs (MICB 421/447) from a lab-based to a bioinformatics course using existing and newly created EDUCE modules. Two more sessions of a bioinformatic

CURE were offered in 2021/2022, and MIBM is in the process of creating a continuing course (MICB 475) using EDUCE materials.

In addition, EDUCE modules in MICB 425 have been fully integrated into the course, with plans to develop data science-related material further as needed. EDUCE material created for workshops will also continue to be offered in ECOSCOPE workshops.

EDUCE functions from a collaborative community of practice with a dedicated lead postdoc. We strongly recommend a full-time individual responsible for overall data science curriculum flow and coordination and teaching and organizing the material. This person also serves as the connection point and main communicator between faculty, TAs, students, and other members of a diverse community of practice. While EDUCE course modules are designed to be extensible, some modifications will be required before their implementation in a new course to accommodate individual teaching styles, course schedules, and students' prior knowledge. This process is made easier through collaboration across a community of practice and the use of open-source tools like GitHub. However, as with any new curriculum, instructor preparation time remains a barrier to implementation.

A dedicated data science lead reduces this barrier by removing the need for primary course instructors to develop or teach data science material they are unfamiliar with. In addition to this lead, integration within a larger curriculum committee context within the department and the Faculty of Science is needed. At the departmental level this helps ensure that curriculum integration occurs transparently and in communication with all stakeholders. At the faculty level, this helps ensure that curriculum development is built on common core foundations of DSC 100 (Introduction to Data Science) and other initiatives across departments.

Unfortunately, funding for a data science lead position in MBIM could not be secured despite the success of the EDUCE initiative. Moreover, there are no plans to transfer the teaching and coordination responsibilities above to current or future faculty. The current strategy of MBIM leadership is to provide data science skills to MBIM students by making DSCI 100 a program requirement and offering MICB 405 (Bioinformatics), MICB 405 (Microbial Ecological Genomics) as well as MICB 475 (see above). This creates numerous logistic challenges including gap years that remain to be addressed at the departmental level and could be the focus of future TLEF initiatives.

In conversation with MBIM tenure track teaching faculty after exposure to EDUCE modules, they expressed an increased interest in opportunities for personal development in data science skills and inquired about workshops. Unfortunately, the EDUCE program has had less success in generating buy-in from research faculty, making developing a coordinated curriculum strategy for data science in MBIM more challenging. Some research faculty reported a lack of integration of the EDUCE material into the existing course curriculum. While efforts were made by EDUCE team members to develop data science modules with integration in mind more iteration is required to achieve this objective. More involvement at the committee level would be useful in developing long-term integration strategies that engage all interested research faculty in the process of data science instruction. This is especially true with respect to more recent faculty hires who arrive with more data science competency and a willingness to develop curriculum materials within an active learning context. Teaching assignments are critical here and more emphasis could be placed on team teaching opportunities that pair motivated research faculty with tenure track teaching faculty to ensure implementation of the departments data science objectives.

In conclusion, EDUCE provided an innovative framework for teaching data science within the context of microbiology and immunology courses that was based on a progressive and modular teaching and learning paradigm within a community of practice that engaged different training levels for curriculum development and implementation. The initiative reached more than a fifteen hundred undergraduates in for credit courses and hundreds more trainees in co-curricular workshops that were calibrated to reinforce topics covered in the classroom environment. Many of the lessons learned from the EDUCE experience have informed development of CUREs courses within the department and energized tenure track teaching faculty to continue building a truly integrated data science curriculum spanning four years of undergraduate course work. The pedagogical foundations of EDUCE were published in PLoS computational biology and the EDUCE teaching team continues to develop manuscripts for peer-reviewed journals in collaboration with tenure track instructors developing CUREs.