12-2023

Designing an overseas experiential course in data science

Hua Leong FWA
Singapore Management University, hlfwa@smu.edu.sg

Graham NG
Singapore Management University

Designing An Overseas Experiential Course in Data Science

1st Given Name Surname
dept. name of organization (of Aff.)
name of organization (of Aff.)
City, Country
email address or ORCID

2nd Given Name Surname
dept. name of organization (of Aff.)
name of organization (of Aff.)
City, Country
email address or ORCID

3rd Given Name Surname
dept. name of organization (of Aff.)
name of organization (of Aff.)
City, Country
email address or ORCID

4th Given Name Surname
dept. name of organization (of Aff.)
name of organization (of Aff.)
City, Country
email address or ORCID

5th Given Name Surname
dept. name of organization (of Aff.)
name of organization (of Aff.)
City, Country
email address or ORCID

6th Given Name Surname
dept. name of organization (of Aff.)
name of organization (of Aff.)
City, Country
email address or ORCID

***Abstract*—Unprecedented demand for data science professionals in the industry has led to many educational institutions launching new data science courses. It is however imperative that students of data science programmes learn through execution of real-world, authentic projects on top of acquiring foundational knowledge on the basics of data science. In the process of working on authentic, real-world projects, students not only create new knowledge but also learn to solve open, sophisticated, and ill-structured problems in an inter-disciplinary fashion. In this paper, we detailed our approach to design a data science curriculum premised on learners solving authentic data science problems sourced from an overseas project sponsor. The course consists of a local component in which students acquire requisite knowledge on data analytics through classroom-based learning and an overseas component where students were on-site with the overseas project sponsor. To evaluate the design of our course curriculum and uncover its potential benefits, we surveyed the project sponsor representatives and tasked our students with the submission of a personal reflection. Quantitative and qualitative analysis of the survey and reflections respectively revealed the sponsors' high satisfaction with the students' output and the students' perception that the course has enriched their learning. The results also indicated that the students attributed teamwork, communications, mentoring and data processing skills as key factors integral to successful project outcomes. In all, we have developed an experiential, project based and applied data science program in an overseas context offered to students from different disciplines which yields promising student learning outcomes.**

***Index Terms*—curriculum design, experiential learning, analytics, data science, machine learning, multi-disciplinary.**

I. Introduction

Data science has enjoyed phenomenal growth in the last decade. The widespread adoption of data science in both industry and government has led to unprecedented demand for data science professionals. Recognizing that data can provide them with a competitive advantage over their competitors, businesses are creating, transforming, and storing any available data in digital forms. This proliferation of complex and rich data in turn demands a mastery of skills stretching beyond the traditional domains of statistics, mathematics, and computer science. Coupled with the fact that data science encompasses the iterative process of data collecting, processing, exploring followed by defining hypotheses, performing analyses, and communicating the results, modern data science is thus inherently multi-disciplinary and experiential in nature; both a practiced art and developed skill [1]. It is thus imperative that students of data science programmes learn through execution of real-world, authentic projects on top of acquiring knowledge on foundational data wrangling, modelling, and algorithms [2], [3].

Fuelled by the increasing demand for data science professional, many educational institutions have either launched new data science courses or accommodated higher enrolment into their existing data science courses. Some educators have proposed curricula changes to integrate data science into their existing courses e.g. [4] while other educators have also recognized the need to incorporate experiential learning into data science curricula [2], [5]. Experiential learning is the process by which knowledge is created through transformation of experience and it comprises of concrete experience, reflective observation, abstract conceptualization and active experimentation [6]. In many data science courses, students learn various data science concepts e.g. construction and application of various analytics models by working on open source datasets from providers such as Kaggle[1]. One shortcoming of this approach [7] is that students' creativity is constrained as there are already many published solutions to these 'engineered' problems. By incorporating experiential learning into the learning process, students will then get to work on real world realistic dataset and acquire non-technical skills such as teamwork, communications and innovation in the process [8].

Project based learning (PBL) [9] is one such teaching

[1]http://kaggle.com

method that assimilates well into this process of experiential learning. PBL is a student-centred pedagogy which involves the learners solving an authentic challenge and is postulated to lead to deeper knowledge and higher engagement of learners [10], [11]. In the process of working on authentic, real-world projects, learners not only create new knowledge but also learn to solve open, sophisticated and ill-structured problems in an inter-disciplinary fashion. In many data science curricula, researchers have advocated for PBL through the incorporation of capstone projects or client-sponsored authentic projects to bridge the applied knowledge gaps of learners.

In the 21st century, the world has become more interconnected with diminishing physical and communication barriers. The learners of the 21st century is expected to be eventually equipped with the capability to solve problems which may not only be situated in their own community but in other global communities as well. The world becomes the classroom where learners are expected to be critical thinkers, problem solvers and global citizens beyond their own shores. To equip their learners with a global perspective, many universities have offered programs ranging from service learning, innovation [12] and capstone projects [13], overseas internships to overseas exchange programs.

To address the abovementioned challenges, in this paper, we detailed our approach in the subsequent sections to design a data science curriculum that is premised on learners solving authentic data science problems sourced from an overseas project sponsor.

II. Related Work

For many academic programs, the years of education usually culminates into a final year or capstone project which challenges learners to demonstrate their academic knowledge in an experiential way. Genevera I. Allen [5] reported on the design and structuring of capstone projects in their data science curriculum to create an experiential learning experience for their learners – executing the client sponsored projects allows their learners to both "practice the art" and "develop the skills" of data science. As posited by the author, a key factor that underpins the success of the client-sponsored capstone projects is the careful selection of the projects. The projects were framed to encompass all the aspects of data science and with well-defined objectives but yet not too specific as to inhibit the learners' open-ended inquiry. Similarly, we also sieved through our potential clients' proposed projects and provided consultancy to our overseas project clients on how to frame the scope of the project prior to the start of the course. As much as we want to have a variety of different problem statements to challenge the students, we would have to balance that with practical factors such as data availability and size of projects (as students have only a term of study to complete it alongside their other academic load).

The study by [14] describes how Worcester Polytechnic Institute (WPI) provided their engineering students with design experiences in an international context within their curriculum. The Major Qualifying Project (MQP) or a capstone project which is part of the required curricula for their engineering students entails all their graduating students working on an open-ended, real-world problem sponsored by external organizations. To facilitate the administration of MQP, WPI set up overseas project centres typically housed within a partner academic institution. Our study differs in that our course is offered as an elective on an opt-in basis and with a small enrolment (typically less than 50). The setting up of project centres would require substantial resources to be invested which may not be feasible for courses with smaller enrolments.

In 2015, my university pioneered an experiential, interdisciplinary project-based learning pedagogy – UNI-X [15] [16]. It is characterized by experiential learning (via solving a real-world problem identified by our partner organizations) and active learner mentoring by both our faculty and the industry representatives. It is also envisaged that the interdisciplinary and multi-perspective approach of UNI-X allows students to better appreciate the applicability and relevance of the course to their future careers. We further expanded this initiative to include projects sourced from overseas project clients and designed the curriculum to accommodate a full-time overseas segment where our students are on-site with the overseas project clients. Through the UNI-XO ("O" for overseas) initiative, we hope that our students acquire not only critical thinking and problem-solving but also cultural empathy skills and better understand the impact of data science within a global context. The course detailed in this paper is the first UNI-XO course designed with a data science agenda within our institution and where enrolment is on elective basis.

III. UNI-XO Experiential Overseas Data Science Project

We seek to develop a course which allows our students to work as a team to solve cross-disciplinary problems situated outside our local context using data science techniques. We expect that our students would also demonstrate how the solutions built by them for our overseas partner companies would allow the latter to effectively and efficiently apply data analytics and machine learning to improve their competitiveness and business efficiency. This course is offered on an opt-in basis once a year. The demographics of the 23 students for this first run is shown in Table I.

TABLE I
STUDENT DEMOGRAPHICS

Gender	Male	43.5%
	Female	56.5%
Course of Study	Information Systems	75%
	Business	12.5%
	Economics	8.3%
	Social Science	4.2%

A. Objectives

As we offer this course to students from disciplines of study other than information systems, to ensure that learners who undertake this course are equipped with basic programming and data analytics skills, we require them to have completed a data analytics foundational course as a pre-requisite. In the data analytics foundation course, students are taught data cleansing, preparation, visualization and exploratory skills as well as building of basic predictive models such as linear regression and decision trees using Python scikit-learn library [17]. With students coming from different disciplines of study, we hope that they can leverage on each other's strengths and diversity to create more holistic solutions.

B. Project

One of the key mandates for this course is for it to have an Asian focus and Thailand is identified as the country to focus on. Before the start of the course, we sourced for potential projects from our Thai contacts. These contacts are Thai companies or their associates who have previously worked with us on previous projects. We also had help from our internal department working on the university's internationalization efforts for the sourcing of the projects. We evaluated the list of projects that came in based on the viability of problem statements, availability of data and level of sponsor support. As students are concurrently taking other courses within the same semester, a key consideration is whether the problem statements can be feasibly completed by the students with ideally some challenging elements to stretch them. Data is also integral to data science projects and thus sufficient and adequate data must be available for students to work on possible solutions. Lastly, as we require our sponsors to have representatives to guide and mentor our students together with the faculty in charge, the level of sponsor support is also integral to the learning of our students. From the evaluation, we finally selected a project sponsor company who manufactures and distributes energy drinks. The project sponsor provided a total of 3 problem statements which we assessed to be scoped adequately for the students to work on over an academic term. Given that it was the first run of the course, we were also mindful of the communication overheads associated with multiple project sponsors coming from different companies.

For this first run of the course, we had a total enrolment of 23 students. The 23 students formed up 4-5 members team on their own, making up a total of 5 teams. We have a main faculty in charge of the entire course, from initial planning to final delivery and another adjunct faculty (with more than 7 years of industry experience in data science) to mentor the students on the project. The adjunct faculty have meet-ups with the teams on a fortnightly basis, advising and ensuring that the teams are working towards and meeting their project objectives. Having worked on data science projects in the industry, the adjunct faculty can also advise the students on the industry best practices. The student teams also schedule meetups with the project sponsor on a need to basis. To optimize the project sponsor's time, the adjunct faculty consolidates student queries and schedules virtual meetings with the project sponsor (the questions are massed up to sufficient volume for a discussion session). The entire list of student queries is also sent via email to the project sponsor before the virtual meetings to ensure efficient use of the meeting time.

C. Course Structure

Our curriculum is divided into a mix of instructional face-to-face and virtual project consultation sessions. For the face-to-face instructional portion, we covered applied data science topics using a mix of case studies and hands-on exercises. Each session lasts for a duration of 3 hours. The topics covered are listed in Table III.

TABLE II
SCHEDULE OF TOPICS COVERED

Session	Topics
1	Introduction to Machine Learning (ML) - What is ML? - ML Workflow - Supervised vs Unsupervised Learning - Features and Labels
2	Feature Engineering - Exploratory analysis - Data cleansing and preparation - Formulating Features
3	Research and Sharing of Thai Culture - Cultural practices - Corporate practices - Traditions and Festivals Project Briefing by Client
4	Translating client requirements part I - Data gathering Data exploration and understanding - Data engineering - Setup infrastructure and tools
5	Designing of ML models - Introduction to ML modelling - Mapping of ML models to project outcomes.
6	Evaluating and refining ML model - Overfitting and underfitting - Train-Test split and k-fold cross validation - Feature Selection - Hyperparameter optimization
7	Implementing and visualizing ML model - Deploying ML model - Batch versus online scoring - Visualizing model output
8	2 weeks of project immersion in Thailand.

A consideration in this course is whether to include deep learning into the curriculum. After gathering the project problem statements from our potential project sponsors, we find that most of the problem statements do not relate to computer vision or natural language processing use cases which necessitate use of deep learning techniques. Incidentally, most of the project sponsors are either just starting to set up their data infrastructure or just starting to explore the use of data and predictive analytics for enhancing business process efficiencies. Thus, we decided not to introduce deep learning but to focus on imparting applied data analytics skills to the students instead. The early data science maturity stage of the project sponsors is also beneficial as it allows much room for students to innovate and explore different data techniques to fulfil the project objectives. Another consideration is the students have learnt basic data preparation and analytics skills including the construction of basic models using Python with the help of scikit-learn but have not really integrated and applied the acquired knowledge into an end-to-end data science project and thus the need to have an applied focus for this course.

This applied focus is reflected in the design of the case studies and hands on exercises. The case studies are either created by the faculty or curated from online sources to expose students to various data science applications in different business domains. The hands-on exercises are in turn designed to build up to an end-to-end data science pipeline from data preparation, feature engineering, modelling to evaluation and final deployment of the models, taking into consideration different business considerations and constraints. To illustrate, for model deployment, students are taught different methods of deploying the model e.g. online mode if the business needs to generate the prediction on-the-fly when new data is available or batch mode if new arriving data can be batched up and prediction is only required once a day.

Within the course, we catered half a session (about 1.5 hours) for students to research and present on Thai culture. This is important as we anticipated that cultural empathy would aid our students in their frequent communication with our project sponsors and during their interactions with Thai locals in the 2 weeks full time segment. We reserved the other half session for our project sponsor to brief the students on his objectives and scope for the project. During the briefing session, the students are also encouraged to clarify with the project sponsor should they have queries on the project objectives and scope.

D. Assessment

The assessment is divided into both individual and group components. The students are required to submit an individual report detailing their key contributions to the project and their personal reflections from working on the project. Together with the class participation and the peer evaluation scores (derived from the peer evaluation that each student will have to submit), these constitute the individual assessment component for the student. A consolidated score is computed from each team's progress report sessions (every fortnight) with the adjunct faculty. The course culminates in a final presentation by each team to the project sponsors. Both the faculty and project sponsor representative will award a score to each team based on a set of rubrics with the faculty's assessment having a higher weight. The scores for the progress report and final presentation will add up to the final group score for the team.

IV. Evaluation

To evaluate the effectiveness of our course design and uncover its potential benefits, we surveyed the project sponsor representatives and did a qualitative analysis of our students' personal reflections. The survey had a total of 5 questions with each questions having 5 options following a Likert scale ranging from 1 – Strongly Disagree to 5 – Strongly Agree. A total of 5 project sponsor representatives responded to the survey. The survey result is shown in Fig. 1. From the survey results, all the project sponsor representatives either strongly agree or agree that

- The work of the students is professionally done.
- They were impressed with the outcomes of the project.
- They were convinced that this mode of collaboration with the university benefited the company.

All the project sponsor representatives were satisfied with the outcomes of this program. One of the respondent was however neutral as to whether they would be working with future batches of students. During our subsequent conversation with the project sponsor representatives, they highlighted that they need time to plan and consider how the students' work may be incorporated into their current work processes before they would consider hosting another batch of students.

As part of the course assessment, the students were required to submit individual reflections that preferably answers the following prompting questions.

- What have you learnt from working on a real-life project?
- What were some of the top challenges you encountered?
- What were some of the top successes you achieved?
- What would you have done differently if you could do this project again?

Qualitative analysis of all 23 students' reflections is performed using the open coding approach. De-identified copies of the students' reflections were first loaded into Taguette[2] (an open-source web-based document tagging tool for qualitative analysis). The first author read through each reflection and coded discrete portions of the text in the reflections as themes. This set of themes were then shared with the second author who independently analyzed the reflections. The 2 coders then met up to discuss and resolve any coding differences. The number of mentions for each theme is derived from the average of the 2 coders' count of mentions for each theme. The result of the qualitative analysis (top 5 themes with the highest number of mentions) is shown in Table III.

TABLE III
SCHEDULE OF TOPICS COVERED

Theme	No. reflections mentioning theme
Course experience	23 (100%)
Dataset	12 (52.2%)
Teamwork	12 (52.2%)
Instructor mentoring	10 (43.5%)
Sponsor comms	10 (43.5%)

A. Course experience

Overall, all the students rated the course positively in terms of their learning experience. This is evident from some of the salient quotes provided by the students:

"Although each experience is different on its own, perhaps the most transformative one is my overseas project experience in Thailand where I have a chance to work with industrial client and broaden my cultural understanding of countries within Asia"

[2]http://app.taguette.org

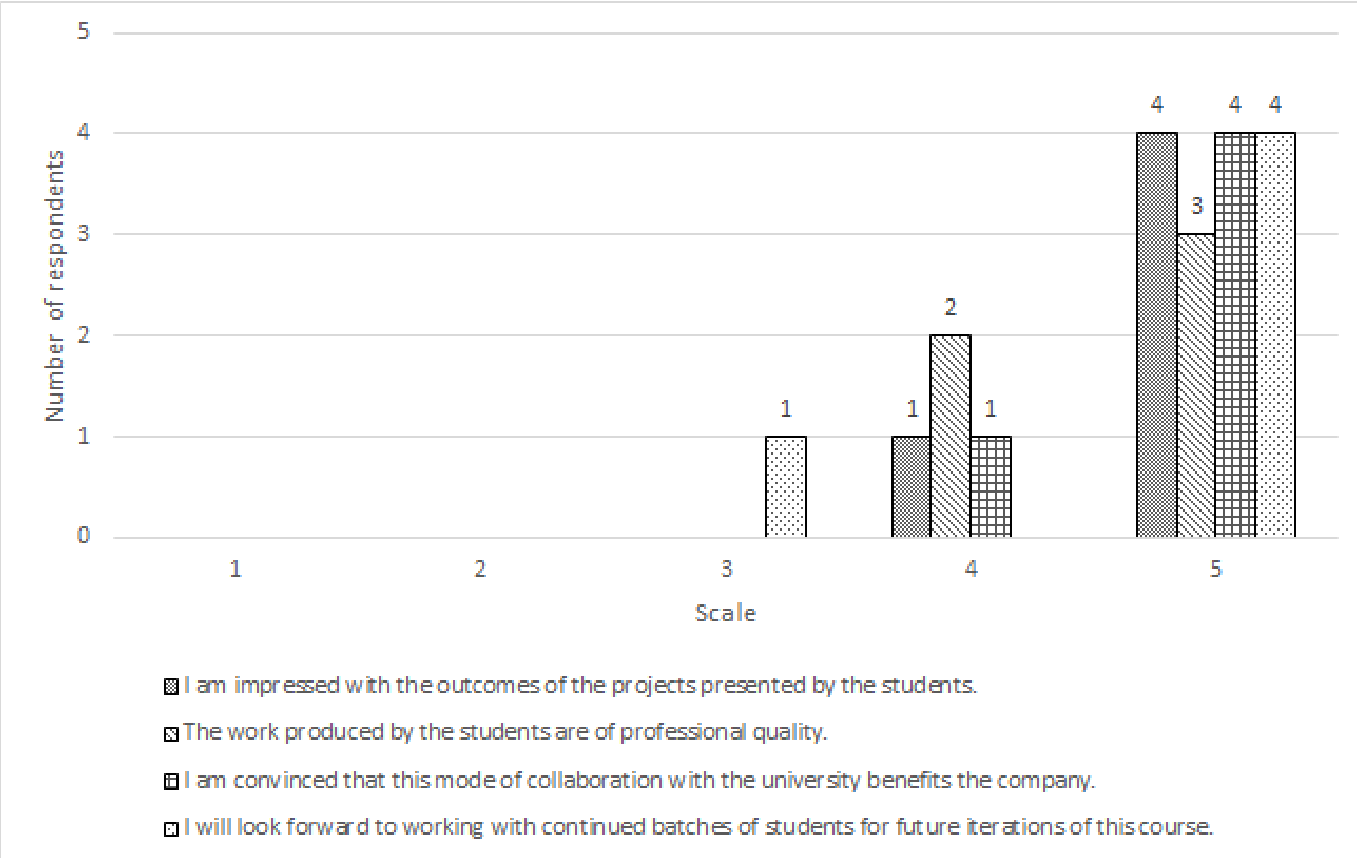

Fig. 1. Project Sponsor Survey Results

"This course provided a very enriching and eye-opening experience behind the logistics and supply chain of one of the biggest beverage companies in Thailand"

"This entire overseas project experience was beyond my expectation"

B. Dataset

More than half of the students (52.2%) reflected on the challenges that they had working on the large dataset provided by the sponsor as evident from some of the quotes below. The quality of the dataset (incorrect and missing data) was also highlighted by some students.

"One of the biggest challenges that we encountered as a team is dealing with big and ambiguous data"

"The project team faced numerous challenges dealing with large datasets"

"Our team found out that the provided data is filled with incorrect and missing data"

C. Teamwork

More than half of the students (52.2%) highlighted teamwork as one of key success criteria for the project. Some teams encountered issues such as low team motivation while others reflected that their cohesive team is integral to their positive project outcomes.

"Some of our team mates were troubled and frustrated that the project was incomplete as the deadline neared"

"I enjoy being a team player and working hand in hand with my project mates much more than doing individual work"

"Overall, I liked our team dynamic. We each knew what we had to do"

D. Instructor mentoring

Mentoring by the instructors were attributed as one of the factors that contribute to the success of their projects by close to half of the students (43.5%).

"The structure of having bi-weekly meetings with our course professors was paramount in the progression of my team's project."

"The fortnightly project consultation with professor was very insightful as he will always think in the client's perspective and then give us his professional view on how to improve our project"

E. Sponsor communication

Effective communication with project sponsor is highlighted as one of the key success factors for their project by the students. Some expressed their desire for faster response from the project sponsor while others valued the face-to-face meet ups with the project sponsor representatives during the overseas segment.

"The sponsors may take a long time to reply"

"I appreciated the importance of speaking with all stakeholders to understand the whole picture"

"Discovered the importance of asking people, on the ground, about their experiences and perspectives when we interviewed the van salesmen"

V. Discussions

From the quantitative analysis of the survey, the project sponsor rated the students' output highly and were satisfied with the outcomes of the program. Continuity with the same project sponsor for future iterations of the course may not be feasible as the project sponsor need time to plan and integrate elements of the student's project into their current work processes. Thus, new project sponsors would likely have to be identified for future runs of the course.

The students rated the course highly as evidenced by qualitative analysis of their personal reflections. The students also commented on the quality of the provided data set and reflected that it constitutes a major part of the challenges that they encountered. In our view however, part of the learning objectives of this course is for students to deal with ambiguity and uncertainty. At the start of the course, the instructors have assessed that the provided dataset is in a state that is adequate for the students to work on, albeit with the need for further clarification with the project sponsor and further data processing. This constitutes part of the learning in which the students would have to undergo to uncover data inadequacies and resolve them (e.g. with data pre-processing techniques) on their own.

Teamwork is also highlighted by the students as one of the critical success factors for their project. Some teams were cohesive while others had issues with motivation and efficient allocation of tasks among the team members. This again coincides with our course objective of students learning to work as a team which includes resolving team conflicts and differences.

Part of the course design involves the instructors acting as project mentors to the students. In our case, the instructors had considerable years of experience working with data analytics projects in the industry and could thus advise the students on where to focus on, whether they are progressing well and the potential pitfalls to note. The students' reflections corroborates that the mentors are instrumental to the success of their projects.

Finally, communications with the project sponsor was highlighted by some students as a bottleneck to the progress of their projects. This is understandably so as the main project sponsor representative had to fulfil his daily work responsibilities on top of answering queries from the students. One suggestion is possibly to work out with the project sponsor company on an 'undisturbed' time slot reserved for the main sproject sponsor representative to answer the students' queries. The students do however value the opportunity to be on the ground interacting with the project sponsor and understanding their work processes during the overseas segment.

VI. Conclusion

In this paper, we detailed the design and implementation of our experiential overseas data science project course which aims to equip our students to work as a team to solve cross-disciplinary problems situated outside our local context using data science techniques. To evaluate the design of our course curriculum and uncover its potential benefits, we surveyed the project sponsor representatives and tasked our students with the submission of a personal reflection. We then tallied the results of the survey and qualitatively evaluated the students' reflections. The results showed that the sponsors were highly satisfied with the students' output and the outcomes of the program. Analysis of the students' reflections also showed the high rating of the course by all students and the fulfilment of the course objectives and benefits e.g. learning to work as a team and effective communications on top of the hard skillsets of data science. We also realize the value of instructors pre-vetting the list of sponsors and their projects, helping to frame the projects and associated problem statements and assessing the adequacy of the data provided in ensuring the successful course and project outcomes. Overall, we have developed an experiential, project based and applied data science program in an overseas context offered to students from different disciplines which yields promising student learning outcomes.

References

[1] R. D. De Veaux, M. Agarwal, M. Averett, B. S. Baumer, A. Bray, T. C. Bressoud, L. Bryant, L. Z. Cheng, A. Francis, R. Gould *et al.*, "Curriculum guidelines for undergraduate programs in data science," *Annu Rev Stat Appl*, vol. 4, pp. 15–30, 2017.

[2] P. Anderson, J. Bowring, R. McCauley, G. Pothering, and C. Starr, "An undergraduate degree in data science: curriculum and a decade of implementation experience," in *Proceedings of the 45th ACM technical symposium on Computer science education*, 2014, pp. 145–150.

[3] N. A. of Sciences, Engineering and Medicine, *Data science for undergraduates: Opportunities and options*. National Academies Press, 2018.

[4] J. Hardin, R. Hoerl, N. J. Horton, D. Nolan, B. Baumer, O. Hall-Holt, P. Murrell, R. Peng, P. Roback, D. Temple Lang *et al.*, "Data science in statistics curricula: Preparing students to "think with data"," *The American Statistician*, vol. 69, no. 4, pp. 343–353, 2015.

[5] G. I. Allen, "Experiential learning in data science: Developing an interdisciplinary, client-sponsored capstone program," in *Proceedings of the 52nd ACM Technical Symposium on Computer Science Education*, 2021, pp. 516–522.

[6] D. A. Kolb, *Experiential learning: Experience as the source of learning and development*. FT press, 2014.

[7] E. Serrano, M. Molina, D. Manrique, and L. Baumela, "Experiential learning in data science: From the dataset repository to the platform of experiences," in *Intelligent Environments 2017*. IOS Press, 2017, pp. 122–130.

[8] A. Ayob, A. Hussain, M. M. Mustafa, and M. F. A. S. Shaarani, "Nurturing creativity and innovative thinking through experiential learning," *Procedia-Social and Behavioral Sciences*, vol. 18, pp. 247–254, 2011.

[9] P. C. Blumenfeld, E. Soloway, R. W. Marx, J. S. Krajcik, M. Guzdial, and A. Palincsar, "Motivating project-based learning: Sustaining the doing, supporting the learning," *Educational psychologist*, vol. 26, no. 3-4, pp. 369–398, 1991.

[10] S. Bell, "Project-based learning for the 21st century: Skills for the future," *The clearing house*, vol. 83, no. 2, pp. 39–43, 2010.

[11] N. Capon and D. Kuhn, "What's so good about problem-based learning?" *Cognition and instruction*, vol. 22, no. 1, pp. 61–79, 2004.

[12] L.-M. Hero and E. Lindfors, "Students' learning experience in a multidisciplinary innovation project," *Education+ Training*, vol. 61, no. 4, pp. 500–522, 2019.

[13] G. Ngai and S. C. Chan, "How much impact can be made in a week? designing effective international service learning projects for computing," in *Proceedings of the 46th ACM Technical Symposium on Computer Science Education*, 2015, pp. 645–650.

[14] R. F. Vaz and P. C. Pedersen, "Experiential learning with a global perspective: Overseas senior design projects," in *32nd Annual Frontiers in Education*, vol. 3. IEEE, 2002, pp. S3B–S3B.

[15] P.-S. Seow, G. Pan, and G. Koh, "Examining an experiential learning approach to prepare students for the volatile, uncertain, complex and ambiguous (vuca) work environment," *The International Journal of Management Education*, vol. 17, no. 1, pp. 62–76, 2019.

[16] G. Pan, P.-S. Seow, V. Shankararaman, and K. Koh, "Essence of partnership management in project-based learning: Insights from a university's global project programme," *Journal of International Education in Business*, 2020.

[17] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay, "Scikit-learn: Machine learning in Python," *Journal of Machine Learning Research*, vol. 12, pp. 2825–2830, 2011.