Module and Kemp Instructional Design Approaches to Integrate STEM Issues and Public Policy into Data Science Curricula at a Native Hawaiian and Pacific Islander-serving Institution

Dr. Rylan C. Chong, Chaminade University

Dr. Rylan Chong is a data scientist in the Division of Natural Sciences and Mathematics at Chaminade University of Honolulu. He earned his bachelor's degree in Computer Science from Chaminade University of Honolulu. Dr. Chong has a master's degree from Purdue University in Information Security. He specialized in biometric systems and human factors during his master's degree. Dr. Chong completed his Doctor of Philosophy degree in Information Security and a Graduate Certificate in Information Security Policy at Purdue University. His dissertation work investigated the relationships of social cognitive career theory factors and cybersecurity research self-efficacy of former and current college students.

Dr. Helen Turner, Chaminade University

Helen Turner is the Dean of Natural Sciences and Mathematics and VP for Innovation at Chaminade University.

Dr. Mark Speck, Chaminade University

Module and Kemp instructional design approaches to integrate STEM issues and public policy into Data Science curricula at a Native Hawaiian and Pacific Island serving institution

Abstract

There is a need for public policy to be integrated in science, technology, engineering, and mathematics (STEM) education as it brings policy and social relevance into STEM classes with the potential to increase engagement and success of students with STEM content. From the perspective of developing future engineers and scientists, an early emphasis on linkages to public policy and societal issues can promote student buy-in, and prepare for future policy and advocacy work that are, increasingly, a component of scientific and engineering careers. From the perspective of the science and engineering community, early exposure to the responsiveness and integration of policy into STEM will increase capacity to clearly articulate and demonstrate the value of science and engineering to society. As a Native Hawaiian and Pacific Island (NHPI) serving institution, there is an added imperative for us to incorporate social and policy content into classes that focus upon issues that are explicitly relevant to the Pacific context. This paper will extend the work of Ngambeki, Depew, Dark, and Chong by exploring how to introduce public policy into data science classes taught to students who are majority Pacific residents and with significant percentages of indigenous identities. Chaminade University of Honolulu National Science Foundation (NSF) I-USE and INCLUDES programs supported pilot deployment of data science modules into traditional STEM classes at Chaminade University between 2017-2019. Each module specifically incorporated data sets that derived from key policy related issues in the Pacific, ranging from engineering, environment, and health disparities. The case study used in this paper will present a model to design a module, and the Kemp instructional design framework to integrate the module and public policy in a data science class.

Introduction

Science, technology, and innovations (STI) are growing and advancing at a significant rate providing many benefits from cures, improving resources of supply and demand, being able to connect with love ones across the globe, and finding new STI. However, along with the benefits, there are safety challenges, from a pharmaceutical drug causing harm, rise in cyber threats, and infrastructure disasters. There is a calling from articles, reports, and an accreditations [1]–[8] that there is a need for public policy to be integrated in science, technology, engineering, and mathematics (STEM) education as students need to master the technical competencies, but be mindful that one's work could provide benefits and harm. More importantly, students of STEM fields, who one day will become the next generation of scientists, technologists, engineers, and mathematicians are important actors in the area of policy as they will fill the holes of STEM expertise that policymakers lack, assist in the development of policy, and advocate various causes [9], [10].

In Fall 2019, Chaminade University of Honolulu will be launching a four-year data science bachelor's degree program that will include a data science decision-theater center and a curriculum that will utilize an active learning approach in most classes where students will work in teams with a faculty or mentor on real and relevant data science problems. Data science is an area that utilizes competencies in computer science, statistics, research methods, and other domain focus areas (e.g., engineering, environment, and health) [11], [12]. Central to data science is being able to work with and handle any forms of data (i.e., small, big, clean, messy, simple, and complex) using a lifecycle process of planning, collecting, processing, analyzing, preserving, sharing, and determining the course [13]–[18]. Because of the calling to incorporate public policy in the curriculum, there are several methods to introduce policy in a STEM education that involves various ranges of time, resources, expertise, and effort which includes, "seminars, modules, workshops, exposure experience, focused course, survey course, certificate, and program [1]." Chaminade University considered the module approach that will be used for the first set of introduction data science classes during the Fall 2019 semester, because of its flexibility, "short treatment of a policy topic, can be used within a larger class [1]," and provides a blueprint and guide for students to follow as they progress through a class. The purpose of this paper is to take a case study approach that will discuss the process of designing a policy module, and integrating it within a data science class.

Module design

Chaminade University has been piloting various data science modules in traditional STEM classes the past few years for students to understand the social, economic, and political context of their work while using data science tools and techniques. Examples of modules, mapping petroleum spills, metagenomic study of ground oil, and Hawaiian fisheries. A case study framework approach was used to design these modules, which includes choosing a topic, provide a framework, determine scope, and choosing the module [4]. Mapping global petroleum spills data science module will be used as the case study example.

Choosing topic

Choosing a topic is the first stage in developing a module and is considered a critical stage that that sets the foundation for the remaining stages [4]. There are particular factors to consider when choosing a topic. The topic should relate to the desired learning outcomes, be relevant to students, show linkage between the STEM area and public policy, and connect to the technical material [4]. The topic chosen was environmental policy and oil spills, because oil spills present a persistent and reoccurring global environmental problem both on the land and offshore [19], [20].

Although the last oil spill in Hawaii occurred in 1977 [21], a major oil spill in Hawaii or anywhere in the Pacific would be catastrophic socially, economically, and environmentally as wild life and marine life, Hawaii coral reefs, those who rely on fishing as a source of food, businesses, tourism, agriculture, and human health and well-being will all be impacted. The goal for the topic and module is to A) expose students of the data science methods and techniques (i.e., collecting, storing, cleaning, analyzing, preserving, sharing, and determining the course of the data), B) use expression data from gene chips that focuses on different environmental samples, and C) familiarize the students to social, economic, policies, environmental science, and technologies behind the petroleum industry and oil spills.

Provide a framework

The second stage in designing the module is to provide a clear framework that will provide focus to the discussion and to structure students' understanding [4]. A framework is described as an outline, steps, structure, or a set of principles [4]. There were three frameworks this module used, which included the data lifecycle (i.e., planning, collecting, processing or cleaning, analyzing data, preserving, sharing, and determining the course of the data) [15], [17], [18], [22], social systems (i.e., social, political, and economic) [23], and four goals of public policy (i.e., equity, efficiency, security, and liberty) [24]. The data lifecycle provides students a guide and an overall picture on how to work with data. The social systems and four goals provide students a conceptual framing on how to think about the module, the problem, results and interpretation, and the social environment tradeoffs.

Determine scope

The next stage is to determine the scope of the module. The scope are the boundaries of the breadth and depth of the module [4]. The scope includes the petroleum company, location, dates, min and max Tonnes, and type of data science techniques and tools to use. In addition, the three frameworks will be used for discussions.

Choosing a module

Choosing a module is the culmination of the earlier stages and incorporation of five principles.

- ***Representativeness*** – "the instructor should consider whether the case under consideration is illustrative of other similar events or processes and how it illuminates them. Depending on the desired outcomes the selected case should be either highly representative to serve as a reasonable example from which the students can extrapolate; highly illustrate the conditions for or consequences of non-representative action, i.e. highlight an impactful success or failure; or serve as a point of comparison [4]." Petroleum spills are ongoing and persistent problem. Hawaii and the Pacific islands are particularly vulnerable to these disasters and ripe for modern preventative policies.

- ***Salience*** – "the case should illustrate a particularly important event that had or is still has a significant impact [4]." One particular case that is of major concern in Hawaii are the Red Hill oil tanks that the US Navy manage [25]. They are situated directly above an active aquifer and leaks have been identified in the chambers. There is an ongoing debate on what steps need to be taken if any at all to address this problem [25].

- ***Interest*** – "the case study chosen should be of potential interest to the students, such as its controversial nature, its continuing relevance, its familiarity etc. [4]." The petroleum spills module opens the discussion on the real dangers Hawaii could face if there is a catastrophic event. Hawaii and the Pacific depends on near and far shore fisheries, agriculture, being able to provide its own potable water, marine and wild life eco systems, tourism, and outdoor activities (e.g., swimming, training, and surfing). In addition, the discussion could further into the tradeoffs among social groups, environment, economics, and policies.

- ***Clarity*** – "the instructor should select a case, in which the facts are fairly clear and not in dispute, unless the goal is to illustrate the impact of such complexities on the policy process and STEM [4]." The module uses publicly available data that is not under debate. The spills represented in the module are not an exhaustive list, but are historically relevant as they were reported through multiple sources and not under scrutiny as to whether they occurred or not.

  Technically relevant – "the case in question should include an example of the area of STEM or a strong connection to the STEM area to help students connect technical and policy subjects [4]." The modules are using modern techniques and contemporary sources to look at a persistent environmental problem that requires policy level solutions as well as technical level remediation efforts.

Introductory data science class case study using Kemp framework

Chaminade University will be launching a four-year data science bachelor's degree program in Fall 2019. The program will offer 21 classes and a directed research class. One of the classes is an *Introduction to data structures, data analytics, and data lifecycle* (ISAL) class that will incorporate the global petroleum spill module to guide discussions, and will be used as the case study for this paper.

Kemp instructional design proposed framework

The Kemp framework was considered in designing the introduction data science class as it provided a holistic and detailed blueprint that guided the development of a class. This design process has a linear approach. However, this framework has the flexibility to become nonlinear. In other words, this framework includes elements instead of stages or levels, and provides flexibility for the designer to move to a specific element during the class design process [26]. There are nine elements:

- ***Instructional goals and problems*** – *Identify instructional problems, and specify goals for designing an instructional program.*
- ***Learner characteristics*** – *Examine learner characteristics that should receive attention during planning.*
- ***Task analysis*** – *Identify subject content, and analyze task components related to stated goals and purposes.*
- ***Instructional objectives*** – *State instructional objectives for the learner.*
- ***Content sequence*** – *Sequence content within each instructional unit for logical learning.*
- ***Instructional strategies*** – *Design instructional strategies so that each learner can master the objectives.*
- ***Designing and delivering the message*** – *Plan the instructional message and delivery.*
- ***Evaluation instruments*** – *Develop evaluation instruments to assess objectives.*
- ***Select resources*** – *Select resources to support instruction and learning activities.* [27]

Implementing instructional design framework to data science class

This section presents how the Kemp instructional design framework is used to develop the ISAL class and the integration of a piloted module. This section starts with discussing the instructional goals and problems. Next, discuss the learner characteristics, task analysis, instructional objectives, content sequence, instructional strategies, designing the message, and development of instructions. This section will end on evaluation instruments.

<u>Instructional goals and problems</u>

The first element of the Kemp model is identifying instructional goals and potential problems to be aware of [27]. The goals for ISAL are for the students to be able to identify and describe the stages of the data lifecycle and data security; communication; social systems; four public policy goals; and public policy solutions of their work. Some of the problems identified, the students will have a diverse range of data analytics and database experience (i.e., beginner vs. advanced); may have anxiety programming and with data analytics; some students might treat this class as "just another class to get through" or lack of motivation; some may have socioeconomic barriers; and some students who may be midcareer or changing career, have family commitments, or have a busy schedule and might not have enough time in their day to commit to the work.

<u>Learner characteristics</u>

The second element is learner characteristics, which are the factors about the students that defines who they are that could influence the learning process and outcomes [27]. Chaminade University's majority is minority. Chaminade University has an undergraduate student body of 1099, 244 declared STEM majors, and 41% of STEM majors were Native Hawaiian and Pacific Islanders (NHPI). Students enrolling in this ISAL class will have a diverse background. Since Chaminade University includes a large number of NHPIs, the values, culture, and traditions form around NHPI ways. Some students may have socioeconomic barriers, be midcareer, be changing career, have family commitments, have busy schedules, or enrolled to learn particular skills. Students will have diverse range of experiences in data analytics and databases, but primarily will have no competencies in these areas.

<u>Task analysis</u>

This element defines and analyzes the content and related tasks. Morris, Ross, and Kemp introduced three analysis techniques that include A) topic analysis, which is an outline of potential topics and subtopics; B) procedural analysis, which are steps to complete a task; and C) critical analysis, which are the competencies needed to accomplish the tasks and topics [27], [28]. The topic analysis, there are three primary topics that include the data lifecycle, social factors, and data security of data science as shown in Table 1.

Table 1 Topic Analysis

Topics	Subtopics
Data lifecycle	Plan, collect, process, analyze, preserve, share, and determine course of data

Social factors	Ethics, social systems, four public policy goals, public policy solutions
Data security	Threats, protection, response or recover, and social topics

Procedural analysis includes 11 short papers and a project. Unlike traditional classes where assignments are isolated from other works, each paper will add towards the students' projects. This way, students are not working on the project last-minute, and actively working towards a final project product. The last type of analysis is the critical analysis. This ISAL class is an introduction class to data science. Using Bloom's Taxonomy, students are minimally expected to be able to identify, describe, and explain each topic supported by the assignments.

Instructional objectives

These are the set of outcomes that are aligned with the class goals that needs to be learned and mastered [27]. The outcomes were designed and guided by Bloom's Taxonomy. The following are the class outcomes:

- Identify and describe data lifecycle.
- Identify and describe the data planning process and approaches.
- Identify and describe data security, ethics, and policies.
- Explain how to generate and collect data.
- Identify and describe the cleaning, formatting, and preparing data process.
- Identify and describe data storage and data structures.
- Identify and describe data management, analysis, visualization, and interpretation.
- Identify and describe communication and storytelling of data results
- Identify and describe the social, political, and economic systems and how data results could influence these systems.
- Identify and describe potential public policy solutions from data results.
- Explain the sharing, publishing, and preserving data process.

Content sequence

Content sequence is the logical order of the topics that will assist the students to understand, learn, and see the big picture [27]. This is primarily structured sequentially that is meant to help learners build on existing competencies. Table 2 illustrates the order of the content sequence.

Table 2 Content Sequence

Week	**Topics**
1	• Introduce data lifecycle.
2	• Designing study and research process. • Data management plan.
3	• Data security and ethics. • Policies for access, sharing, and re-use.
4	• Generating and collecting data.
5	• Cleaning, formatting, and preparing data.
6	• Data storage and data structures.
7	• Data management.

8	• Data analysis.
9	• Data visualization.
10	<Break>
11	• Data interpretation and influence on the social, political, and economic systems
12	• Communication and storytelling of data results to Hawaii community
13	• Sharing, publishing, and preserving data.
14	• Public policy solutions
15	Last day of classes
16	Finals week

Instructional strategies

This element utilizes strategies, models, frameworks, approaches, so forth that will enable the students to master the outcomes [27], [28]. Like traditional classes where students attend class, listens to a lecture, and do assignments and exams, this class and many other classes in the program will incorporate some these strategies. In addition, based on the previous elements, this class will embrace experiential learning, active learning, student/learner-centered, constructive alignment, and project-oriented strategies. From the start of this class, students will be working in a team with a faculty or mentor on a real and relevant problem for most of the semester. Throughout the class, faculties and mentors are guides to learning, a module will be used to demonstrate each topic, students will be creating knowledge, and to some extent, the students will set the pace of learning.

Designing and delivery of message

There are four parts from designing the message to delivery, which is meant to help students engage and recall the content. These parts include cueing, structure, visualization case, and delivery [28]. Cueing is used as a precursor to the topic that will be covered. Could be performed as a question towards the students or an activity a student may do that leads up to the topic [28]. For example, if the topic of the week is on public policy solutions. An instructor could pose a question to the students to think about policy opportunities and implications from their data analysis results. Next is the structure, is for the instructor to use appropriate headings that promotes the topic, idea, method, or theory [28]. An example is public policy and related solutions. The third part is using visualization from figures, charts, and pictures to illustrate heading [28]. The last part is delivery. Lectures is the traditional method of delivery of the message [28]. However, this class will use several delivery methods to facilitate individual work and group collaboration. The primary method of delivery will be to use the module that will be made available online and in class to guide the class sessions as the students will see the topic structure, relating examples, and visualization for a particular week, and cued for the next topic for the following week.

Evaluation instruments

The evaluation instrument element are methods and instruments that will be used in this class that will evaluate learner's mastery of each outcome [27]. This is primarily performed using some form of feedback mechanism, such as asking the students questions, responses from a discussion or presentation, the assignments, and the semester long project [28]. This class will also include a pretest survey and posttest survey study on self-efficacy, anxiety, and career development of the class. An interview may be used if it could help further explain or bring clarity to a response.

Select resources

This last element are the resources that will support the topics and activities [27]. Each resource selected should be short, clear, and easy to understand [29]. This class will utilize required readings. In addition, tutorials, online videos, and websites will be used to support the required readings.

Conclusion and next steps

There is a calling for public policy to be integrated into STEM education. A method proposed in this paper to introduce public policy was using a module approach. Chaminade University been piloting various data science modules for the past few years and is planning to integrate these modules in their new data science classes. The module used as a case for this paper was on mapping global petroleum spills, in which this paper discussed its design from choosing the topic, providing a framework, determining the scope, and choosing the module. Next, this paper introduced Kemp's instructional design approach to design the *Introduction to data structures, data analytics, and data lifecycle* class with the goal to integrate the topics of data science lifecycle process, data security, ethics, social systems, and public policy solutions. The next steps will be to develop a pretest survey and posttest survey study that will investigate the students' self-efficacy, anxiety, and career development.

Acknowledgements

This work was supported by the National Science Foundation grants of NSF I-USE (DUE1525884), NSF INCLUDES (HRD1744526), and NIH INBRE (P20 GM 103466) for the development of a culturally-responsive data science, analytics, and visualization undergraduate curriculum and summer immersion program at Chaminade University of Honolulu in partnership with Texas Advanced Computing Center (TACC). Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the NSF, NIH, Chaminade University of Honolulu, and Texas Advanced Computing Center.

The author thanks his co-authors Dr. Helen Turner (Chaminade University) and Dr. Mark Speck (Chaminade University), the Data Science Program team, and the Division of Natural Sciences and Mathematics at Chaminade University of Honolulu for their insights to the development of this article.

References

[1] I. B. Ngambeki, D. R. Depew, M. J. Dark, and R. C. Chong, "Approaches to integrating policy into engineering education," presented at the 120th ASEE Annual Conference and Exposition, Atlanta, GA, 2013, pp. 1–14.

[2] D. Grasso and D. Martinelli, "Holistic engineering: The dawn of a new era for the profession," in *Career development in bioengineering and biotechnology*, G. Madhavan, B. Oakley, and L. Kun, Eds. New York, NY: Springer, 2008, pp. 303–307.

[3] National Research Council of the National Academies, *Using science as evidence in public policy*. Washington, DC: The National Academies Press, 2012.

[4] R. Chong, D. Depew, I. Ngambeki, and M. Dark, "Teaching social topics in engineering: The case of energy policy and social goals," presented at the 120th ASEE Annual Conference and Exposition, Atlanta, GA, 2013, pp. 1–16.

[5] J. Mendoza-Garcia, I. B. Ngambeki, L. J. Behbehani, D. Evangelou, S. C. Rao, and M. F. Cox, "Defining the knowledge and skills that enable engineers to participate in public policy," presented at the American Society for Engineering Education, 2012, pp. 1–12.

[6] WASC Senior College and University Commission, "Handbook of accreditation 2013 revised quick reference guide." 2013.

[7] National Academy of Engineering, *Educating the engineer of 2020: Adapting engineering education to the new century*. Washington, DC: The National Academies Press, 2005.

[8] ABET, "Criteria for accrediting engineering prorams," ABET, Baltimore, MD, 2013.

[9] S. L. Otto, *Fool me twice: Fighting the assault on science in America*. New York, NY: Rodale Inc., 2011.

[10] C. F. Manski, *Public policy in an uncertain world: Analysis and decisions*. Cambridge, MA: Harvard University Press, 2013.

[11] J. Vanschoren *et al.*, "Towards a data science collaboratory," in *Advances in intelligent data analysis XIV*, France, 2015, pp. 1–13.

[12] M. A. Waller and S. E. Fawcett, "Data science, predictive analytics, and big data: A revolution that will transform supply chain design and management," *J. Bus. Logist.*, vol. 34, no. 2, pp. 77–84, Jun. 2013.

[13] J. Frankenfield, "Data science," *Investopedia*, 2018. [Online]. Available: https://www.investopedia.com/terms/d/data-science.asp. [Accessed: 17-Jan-2019].

[14] APEC Human Resource Development Working Group, "Data science and analytics skills shortage: Equipping the APEC workforce with the competencies demanded by employers," Asian-Pacific Economic Cooperation, Singapre, 2017.

[15] Microsoft, "What is the team data science process?," *Microsoft Azure*, 19-Oct-2017. [Online]. Available: https://docs.microsoft.com/en-us/azure/machine-learning/team-data-science-process/overview. [Accessed: 17-Sep-2018].

[16] T. H. Davenport and D. J. Patil, "Data scientist: The sexiest job of the 21st century," *Harvard Business Review*, 2012. [Online]. Available: https://hbr.org/2012/10/data-scientist-the-sexiest-job-of-the-21st-century. [Accessed: 17-Jan-2019].

[17] ITTStar Consulting LLC, "IT data life cycle," *ITT*, 2018. [Online]. Available: http://www.ittstar.com/it_data.html. [Accessed: 17-Jan-2019].

[18] University of Ottawa, "What is research data management," *uOttawa Library*. [Online]. Available: https://biblio.uottawa.ca/en/services/faculty/research-data-management/what-research-data-management. [Accessed: 17-Jan-2019].

[19] Associated Press, “Top 20 onshore U.S. oil and gas spills since 2010,” *USA Today*, 2017. [Online]. Available: https://www.usatoday.com/story/news/nation/2017/11/17/top-20-onshore-oil-and-gas-spills/876390001/. [Accessed: 26-Jan-2019].

[20] M. Roser, “Oil spills,” *Our World in Data*, 2019. [Online]. Available: https://ourworldindata.org/oil-spills. [Accessed: 26-Jan-2019].

[21] The New York Times, “Tanker explodes in 50-mile slick west of Honolulu,” 1977. [Online]. Available: https://www.nytimes.com/1977/02/25/archives/tanker-explodes-in-50mile-slick-west-of-honolulu.html. [Accessed: 26-Jan-2019].

[22] A. Ball, “Review of data management lifecycle models,” University of Bath, United Kingdom, Review redm1rep120110ab10, 2012.

[23] M. Dark, I. Ngambeki, D. Depew, and R. Chong, “Social engagement by the engineer,” in *Understanding the global energy crisis*, E. Coyle and R. A. Simmons, Eds. West Lafayette, IN: Purdue University Press, 2014.

[24] D. Stone, *Policy paradox: The art of political decision making*, 3rd ed. New York, NY: W. W. Norton and Company, 2012.

[25] R. Daysog, “Red Hill fuel spill poses threat to water supply, officials conclude,” *Hawaii News Now*, 2015. [Online]. Available: http://www.hawaiinewsnow.com/story/30399540/red-hill-water-contamination/. [Accessed: 17-Mar-2019].

[26] S. Kurt, “Kemp design model,” *Educational Technology*, 2016. [Online]. Available: https://educationaltechnology.net/kemp-design-model/. [Accessed: 17-Mar-2019].

[27] J. E. Kemp, G. R. Morrison, and S. M. Ross, *Designing effective instruction*, 2nd ed. Upper Saddle River, N.J: Merrill, 1998.

[28] S. N. J. Ahmad, “Instructional design and delivery: Jerrold Kemp instructional design model,” 2013.

[29] R. Chong, “Teaching social, political, and economic topics to student data scientists: The case of a data science course,” in *Arts, Humanities, Social Sciences, and Education*, Honolulu, HI, 2019, pp. 1–10.