

Attribution 3.0 Unported (CC BY 3.0 DEED)

<https://creativecommons.org/licenses/by/3.0/>

Citation: Chen, Lujie, Jamie Gillan, Matthew Decker, Egan Eteffa, Anjelica Marzan, Justin Thai, and Sarah Jewett. "Embedding Digital Data Storytelling in Introductory Data Science Course: An Inter-Institutional Transdisciplinary Pilot Study." *Journal of Problem Based Learning in Higher Education* 11, no. 2 (October 10, 2023): 126–52. <https://doi.org/10.54337/ojs.jpblhe.v11i2.7767>.

Access to this work was provided by the University of Maryland, Baltimore County (UMBC) ScholarWorks@UMBC digital repository on the Maryland Shared Open Access (MD-SOAR) platform.

Please provide feedback

Please support the ScholarWorks@UMBC repository by emailing scholarworks-group@umbc.edu and telling us what having access to this work means to you and why it's important to you. Thank you.

Embedding Digital Data Storytelling in Introductory Data Science Course: An Inter-Institutional Transdisciplinary Pilot Study

*Lujie Karen Chen, Jamie Gillan, Matthew Decker, Egan Eteffa, Anjelica Marzan,
Justin Thai, Sarah Jewett **

ABSTRACT

With the emergence of data science as an inherently multidisciplinary subject, there is increasing demand for graduates with well-rounded competence in computing, analytics, and communication skills. However, in conventional education systems, computing & quantitative, and communication skills are often taught in different disciplines. Data storytelling is constructing and presenting data stories to highlight the analytical insights to achieve the communication goals to a specific audience. Digital data storytelling leverages digital storytelling techniques and best practices in communication to deliver stories that can be shared in digital formats to a wide audience. In this paper, we describe and reflect on a semester-long project-based learning pilot using Digital Storytelling as a framework to allow students to explore topics themed around human flourishing and sustainability with the end goal of constructing data stories delivered in digital or video format (i.e., Digital Data Storytelling). The pilot work was conducted in an introductory data science course at a 4-year Minority Serving Institution in collaboration with students studying non-STEM disciplines at a partner community college. Our pilot demonstrates the potential benefit of this sustainability-aware Project-Based Learning design in raising students' awareness of sustainability issues, increasing confidence in cross-disciplinary communication competency, and at the same time

* Lujie Karen Chen, University of Maryland Baltimore County, United States
Email: lujie@umbc.edu
Jamie Gillan, Montgomery College, United States
Email: jamie.gillan@montgomerycollege.edu
Matthew Decker, Montgomery College, United States
Email: matthew.decker@montgomerycollege.edu
Egan Eteffa, University of Maryland College Park, United States
Email: eganeeteffa@gmail.com
Anjelica Marzan, Montgomery College, United States
Email: amarzan@montgomerycollege.edu
Justin Thai, University of Maryland Baltimore County, United States
Email: jthai.justin@gmail.com
Sarah Jewett, University of Maryland Baltimore County, United States
Email: sjewett@umbc.edu

deepening their understanding of data science concepts. We further reflect on the significant role of an effective program model as well as challenges and opportunities for building transdisciplinary communication competency to prepare for a diverse data science workforce.

Keywords: Data Science Education, Communication Competency, Data Storytelling, Digital Storytelling, Transdisciplinary Collaboration

INTRODUCTION: WEAVING SUSTAINABILITY INTO DATA SCIENCE EDUCATION

With the technology affordance and computational power for collecting, storing, and processing data, individuals with data skills are in high demand (Nicolaus Henke et al., 2016). A sustainable society needs a diverse and inclusive workforce equipped with the skills to extract insights from large and complex dataset and turn them into actionable narratives that inform the design of sustainable solutions to benefit humanity. Data science education is such a field that has the potential to answer the call. However, the current data science education in higher education typically focuses on a few narrowly defined technical competencies, such as programming and quantitative reasoning. Still, those skills alone are insufficient to support the vision toward a sustainable future. For data science education to integrate sustainability, we argue for leveraging student-centered active learning pedagogy, such as project-based learning to encourage students to be engaged in data science problem-solving with sustainability implications. Specifically, we propose three aspects of sustainability that can be woven into data science education by considering the domain aspects of data science (Domain Sustainability), the process of doing data science (Disciplinary Sustainability), and the diverse student population to be engaged in data science education (Educational Sustainability).

Domain Sustainability: Data science is an empirical science that solves data-driven problems in specific domains. One straightforward approach to embed sustainability into data science education is to motivate students to launch authentic inquiries into domains closely connected to sustainability issues. One overarching domain framework is human flourishing (VanderWeele, 2017), which concerns living well sustainably, both as individuals and collectively as communities. A flourishing society largely depends on the sustainability of the environment and human capacity, which are enabled by infrastructure in education and healthcare, and beyond. By engaging students in data-driven projects exploring datasets in those related domains, we begin to build a platform for students to launch data-driven inquiries and answer their own curious questions, thus

motivating them to hone in on data science skills while simultaneously deepening their understanding of the sustainability aspects that the data and problem provide.

Disciplinary Sustainability: A core competency of data scientists is communicating and collaborating confidently with people from diverse backgrounds and disciplines. A data science project team in the real world comprises domain experts and stakeholders with various degrees of data acumen. They practice distinct norms and speak the languages of their discipline. Sustainable data science education thus needs to be deliberate in training students to work in a multidisciplinary environment, developing intellectual humility to quickly adapt to the culture of the discipline where the problem is situated. Moreover, to encourage data science students to develop essential competency in communicating with data, i.e., data storytelling skills (Dykes, 2019; Knafllic, 2015), data science educators could benefit from methodology and approaches from other disciplines, such as communication and literacy, where theory and practice of narrative and storytelling are well developed (Bietti et al., 2019; Risam, 2018).

Educational Sustainability: Data science and analytical professionals, like many other STEM disciplines, will have the highest chance of being sustainable if students from underrepresented groups can actively participate and have their voices heard, and ultimately be well equipped to discover and solve problems emerging from their communities (Norman, 2023). A diverse workforce has more opportunities to create sustainable and inclusive products and services (Williams et al., 2019). However, so far, students from minority groups are disproportionately underrepresented in the data science workforce (National Science Foundation, 2021). Educational institutions must intentionally implement educational interventions to engage students from diverse backgrounds; this includes students from minority backgrounds, first-generation students, and students transferred from or currently enrolled in community colleges.

This framework is inspired by the principles of sustainable education, also known as Education for Sustainable Development (ESD), as outlined by The United Nations Educational, Scientific and Cultural Organization (UNESCO). ESD embodies educational practices that integrate key sustainable development themes into teaching and learning processes. UNESCO has identified five priority action areas to achieve these objectives¹. Of these, Priority Action Area 2: Transforming Learning Environments and Priority Action Area 6: Empowering and Mobilizing Youth, have particular relevance for undergraduate data science education, which is the primary focus of this paper.

While ESD provides an over-arching goal, we would like to operationalize in the context of data science education, thus providing educators with a framework, referred to as "Sustainable Data Science Education," to guide their strategic and practical pedagogical decisions. This framework encompasses three interconnected components of suitability

in domain, discipline, and education, which can be explicitly linked to most of the Sustainable Development Goals (SDGs) as illustrated in Figure 1.

Domain Sustainability encourages educators to present sustainability as a subject of study in data science education. At least nine SDGs can intersect with data science education, either through exploring datasets and problems related to these indicators or through in-depth investigations to inform policies that could enhance these indicators. Considering the complex and multidisciplinary nature of solving problems related to SDGs, students need to be challenged to equip themselves with the necessary knowledge and skills for problem-solving with multidisciplinary teams - a concept referred to as **Disciplinary Sustainability**. **Educational Sustainability** advocates for intentional efforts to promote inclusive participation from a diverse range of backgrounds, ensuring that all students can engage in data-driven inquiry, whether as professionals or as global citizens. This can be directly linked to four SDGs: Quality Education, Gender Equality, Decent Work and Economic Growth, and Reduced Inequality.

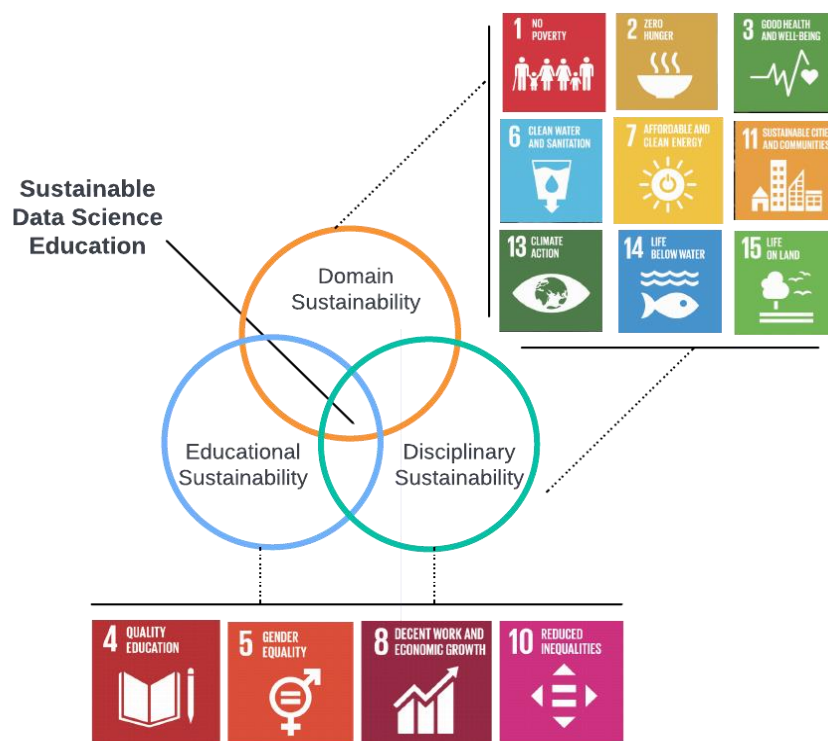


Figure 1. Illustration of three components defining Sustainable Data Science Education, including Domain Sustainability which is linked to nine of the Sustainable Development Goals or SDGs², and Educational Sustainability and Disciplinary Sustainability, as relevant to four of the SDGs.

In this paper, we describe a pilot study in an introductory data science course in a 4-year Minority Serving Institution (MSI) in collaboration with a Digital Storytelling Internship program from a community college with Hispanic Serving Institution (HSI) designation. This pilot operationalizes the three layers of sustainability through a project-based

learning approach. Along the dimension of **domain sustainability**, students are invited to launch a data storytelling project within an overarching theme of Human Flourishing, which is largely centered around environmental sustainability topics, such as climate change, and human society sustainability topics, such as happiness, well-being, and mental and physical health; meanwhile, it explores the **educational sustainability** by engaging students from a university and a community college college, with the majority from underrepresented groups. Moreover, to strengthen **disciplinary sustainability**, the project involves peer-assisted learning (Topping, 2005) activities through two cross-training workshops where students (STEM students from the university and Non-STEM students from the community college) learn skills from the discipline they are unfamiliar with while honing their skillset with relative strength.

In the rest of the paper, we give a brief overview of the literature to contextualize the contribution of this work. We will then provide the institutional, program, and course context, followed by a description of the pedagogical design around project-based learning and its implementation and the data collected. We will share findings from the study and conclude with reflections and discussions of the specific contribution of this pilot to the research and practice in project-based learning for sustainable data science education, as well as the implications and lessons learned from this pilot project for future research and practices.

LITERATURE BACKGROUND

Digital Storytelling, Data Storytelling, and Digital Data Storytelling

Digital Storytelling is a process where individuals, regardless of video editing experience, craft personal video narratives that emphasize the creator's voice. Digital stories often last one to five minutes and combine images, video clips, music, text, and voiceovers. Digital storytelling fosters higher-order thinking skills, develops digital literacy, amplifies one's creative voice, and helps students to develop their professional identity. The storytelling journey starts with an open-ended prompt, leading to idea selection, script creation (250-500 words), and video production. Crucial to this process are Story Circles (Lambert, 2018), spaces for deep listening, feedback, and co-creation (Liguori, 2020). During each Story Circle, storytellers learn more about their narrative and audience expectations and, ideally, continue to leave with a set of editing plans until they complete and publish their story.

Data Storytelling is about effectively communicating data-driven insights (Casola, 2020) and leveraging them for social change (Bhargava, 2017). It merges data analysis, visualization, and narrative to enhance comprehension, spark action, and promote collaboration (Kosara & Mackinlay, 2013). It goes beyond mere visualization and emphasizes clear rhetorical functions like persuasion and explanation. Starting with data

insights generated from the data science workflow, it involves refining these insights into coherent stories. This approach mirrors disciplines like journalism or filming, where raw materials are transformed into stories in various formats.

Digital data storytelling is an emerging concept that integrates the data storytelling process of generating data stories and the creative process of digital storytelling to culminate into data stories presented as short videos. Perhaps one of the most popular digital data stories is late Hans Rosling's 4-minute video that tells the data story of the development trajectory of 200 countries in 200 years (BBC, 2010). Digital data stories expand digital stories' focus on personal stories of qualitative and ethnographic nature to stories that are told in the form of data stories, often supported by quantitative analysis. While data stories can assume various formats, ranging from a single static visualization to a series of presentation slides to interactive webpages, digital data stories adopt a special presentation format rendered in video format.

Teaching Cross-disciplinary Communication Competency in Data Science

Despite the recognized importance of communication skills, the research on effective pedagogical strategy and tactics is sparse other than a few lines of work to explore teaching practices for data-driven written communication, for example, through a course (Nolan & Stoudt, 2021), workshop (Hildreth et al., 2022) or embedded Python Notebook (Willis et al., 2020). One of the contributions of this research is to explore a pedagogical approach that integrates communication skills training with data storytelling project-based learning where students work with an authentic task of creating public digital data stories by analyzing a real-world dataset. This setup is augmented with peer-assisted learning in a cross-training format that motivates students to reinforce their data analytical skills through teaching non-STEM majors while learning new storytelling strategies from students with strength in communication.

Teaching Social Responsibility in Data Science

Though the modern computational and mathematical tools from data science are powerful, sustainable data science education programs should foreground teaching students how to use the tool wisely by maximizing the positive societal impact while reducing negative consequences arising from issues such as algorithmic bias, privacy, and security (Kearns & Roth, 2019; O'Neil, 2017). This goal can be realized through two complementary strategies. The first strategy involves encouraging students to explore problems with societal impact, for example, those links to the United Nations Sustainable Development Goals (United Nations, 2015), as demonstrated by current research and development work in the field of AI for social good (Tomašev et al., 2020) or Data Science for Social Good (Ghani, 2018). The second strategy explores how to teach students to practice data science in a socially responsible manner. There is a growing body of literature that explores pedagogical approaches to the study of ethics in data

science and computing (Lewis & Stoyanovich, 2021; McDonald et al., 2022); some of the other work empowers students from under-represented groups to work on projects addressing concerns in their community (Fisler, 2021; Ryoo et al., 2021). By encouraging students from minority-serving institutions to explore real-world datasets within the framework of human flourishing, this project offers students opportunities to explore societal issues through data-driven, hands-on exploration. This work adds to the limited but growing body of work in incorporating social responsibility into data science in higher education, such as through volunteering work (Monroe-White et al., 2023) or service learning (Farahi & Stroud, 2018).

Project-based Learning in Data Science Education

It is common to incorporate projects in data science education as projects allow students to apply theoretical knowledge and skills to the application (Farahi & Stroud, 2018) while engaging students in authentic problem-solving. Some variation of PBL in data science involves experiential learning with real clients (G. I. Allen, 2021) or service learning that engages a community partner (Farahi & Stroud, 2018). However, a limited amount of work on PBL explicitly incorporates communication goals. One exception is the work reported by (Donoghue et al., 2021), which elaborates on a student-centered project with a communication and storytelling component comprised of lectures on data communication and a written report as a deliverable. Students were evaluated on their “effectiveness of written communication, clarity in their story, and careful design of their visualizations.” Our project takes a step further by introducing digital storytelling as the data story delivery format while integrating peer-assisted learning (Topping, 2005) on topics of societal concern. Digital storytelling expands data storytelling’s common delivery medium of written sources into video format, making data stories accessible to a wider audience.

Project-based Learning and Digital Storytelling in STEM Education

Digital storytelling, despite its adoption as an educational tool as early as in the 1990s (Wu & Chen, 2020), has not received much attention in STEM education in high education, and its intersection with Project-based Learning (PBL) is even more limited. A recent survey on Educational Digital Storytelling revealed that of the 57 studies analyzed, merely nine were centered around STEM disciplines, with just four being conducted within higher education contexts, while the remaining focused on primary and secondary school subjects like science, physics, computing, and technology. Among the four studies situated in higher education, two were designed as individual assignments aimed at enhancing digital literacy (Chan et al., 2017) and fostering intercultural awareness (P. M. Ribeiro, 2016) rather than focusing on a project-based learning objective. (Rambe & Mlambo, 2014) introduced the concept of a digital storytelling community platform called "Knowledge Audio Repository," where graduate students

can externalize their personal knowledge and reflect on their research process using the Digital Storytelling format. The study by (Brace et al., 2015) was only one of the few instances where digital storytelling was integrated into a project-based learning framework. The research involved 47 college students working on group-based projects around food systems, with a digital storytelling output intended to disseminate health messages to the public. The aim was to encourage healthier food systems and stimulate a shift in food production practices.

As noted in (Wu & Chen, 2020), the use of Digital Storytelling in STEM education has not been fully exploited, and there is a lack of educational projects that articulate learning objectives with a reconstructive orientation - the process of reconstructing the meaning of a given concept. They further advocated using Digital Storytelling to probe climate change issues through critical evaluations of "facts, views, and underlying assumptions" and "relating issues to personal contexts." This perspective aligns with our intent to incorporate Digital Data Storytelling with a focus on sustainability, thereby endorsing a reconstructive approach to STEM inquiries.

INSTITUTION AND COURSE CONTEXT

Institution Context Institution A is a minority-serving public university on the east coast of the United States with a student population of over 10,000, with more than half of the students from minority backgrounds. This institution hosts over 100 degrees and programs, and over 85% of the class with less than 50 students. Institution A is also one of the major destinations for transfer students from 2-year community colleges in this area. About a third to a half of newly enrolled students are transfer students. Institution B is a community college with campuses less than 40 miles away from institution A. Institution B is a minority-majority with a Hispanic Serving Institution (HSI) designation with 43,000 students from 155 countries; about two-thirds of the student population belongs to minority groups. It is worth noting that A is a key transfer-receiving institution for B, and B is a key transfer-sending institution for A.

Course Context The pilot was conducted as part of the introductory data science course in the Spring semester of 2022. Modeled after the Data 8 course from UC Berkeley (Adhikari et al., 2021), it requires no prior programming experience and only high school Algebra 1. The course teaches data science using a custom python package similar to Pandas. Institution A's adaptation centered on Digital Data Storytelling (DDST). This provided cross-disciplinary project-based learning opportunities, focusing on real-world sustainability-related datasets. Data science students and DST interns exchanged expertise, ranging from coding and data visualization to narrative creation and video editing, respectively. The class has 25 students, with over 90% from minority groups,

over 40% community college transfers, 30% first-generation college students, and 30% females. While most were from STEM backgrounds, including information systems and computer science, some from fields like psychology, sociology, and public health.

COMPLEMENTARY PROGRAM MODEL

The course served as a host site for an inter institutional collaboration in digital storytelling. Institutions A and B have broad and deep inter-institutional partnerships fostered by the rich and varied collaborations of faculty and staff over time. The digital storytelling collaboration is a powerful exemplar of the value of that partnership for both communities. First piloted in the Fall of 2019, institution B's Digital Storytelling Internship is a two-tiered student leadership opportunity offered through the Humanities Center. During their first semester, Level One interns learn about digital storytelling, create their own digital stories, and develop expertise in the storytelling process. They support faculty and students engaged in digital story projects through office hours or classroom attendance. If they choose to continue, Level Two interns leverage their newly-gained skills and expertise and take on leadership roles in special projects hosted by key partner institutions, including institution A, the internship's most active and robust collaborator. In the past three years, institution A's faculty and staff across nine departments and units have hosted 19 unique interns, with four returning for an additional semester.

At institution A, faculty members across various disciplines host an intern to support a digital storytelling assignment or project in one of their classes. Through online and/or in-person class visits, the intern explains the story process, facilitates story circles, teaches video editing, and supports emerging stories and storytellers. One of the most distinctive and important aspects of this experience is that these community college interns serve as experts in university classrooms. As the interns provide important expertise for their host classes, they also gain opportunities to explore institution A as a potential transfer institution. They have special status as visiting students on campus with access to institution A's physical and virtual resources. They concurrently enroll in a practicum through the Center in which they reflect on their developing practices as storytellers and leaders. In these sessions, they also network with faculty and staff while they learn about important resources for transfer students, unique disciplinary opportunities, and significant campus commitments.

This program model provided another important perspective on teaching and learning, as well as interdisciplinary learning. This brief reflection below, written by the interns, illustrated the significant implications of using this model to complement communication goals in the data science class.

In this applied experience, the cross-training experience offered opportunities for us as story interns to gain and deepen skills by enacting the roles of student and teacher. As inexperienced novices, we learned the basics of coding, while as emerging experts, we taught the basics of digital storytelling. For the coding session, we learned how to perform simple warm-up tasks using the Python programming language, such as getting the program to say “Hello [Name], how are you today?” as well as fixing tables that mirrored a restaurant order. For the storytelling session, we led the students through the story process using story prompts, Story Circles, and a digital story platform (WeVideo). For example, during the Story Circle, we opened with this prompt: “What were some of the first encounters with the special people in your life?” In a subsequent Story Circle activity, one learner recalled a story of a dear friend they made when they first came to the U.S. Another shared a memory from middle school when they met one of their current best friends. Not only did this activity allow the data science students to practice their storytelling skills, it was an opportunity to recall memories and share personal stories in a safe space. Throughout the session, we strove to ensure that we tapped into various learning styles and prioritized the development of visual, written, and hands-on components.

We also gained useful insights and perspectives during the cross-training experience. While the Python tasks wouldn’t be particularly taxing for those who are familiar with code, as digital storytellers with minimal STEM field crossover, we needed great assistance from the data science students, who very willingly walked us through each activity. The coding session helped us see in action just how meticulous the act can be, where even the slightest mistake of missing a comma or apostrophe would result in an error. In some ways, the meticulous parts of coding are similar to how detailed the editing portion of digital stories can be. While audiences may not recognize small errors in audio recording or be aware of the tightness of cuts between shots, these are elements that content creators are constantly mindful of, just as programmers are to coding. Before this experience, we would have been unlikely to see these shared dimensions or to be mindful of the ways we can effectively communicate messages across a variety of mediums regardless of the field of expertise.

Crossing traditional disciplinary boundaries opens up new possibilities for collaboration and communication. Now, just about a year later, we recognize the ways the experience helped us see how much we value interdisciplinary learning in our academic and professional settings. Collaborative learning environments like the cross-training session not only allowed us to share our perspectives as students involved in the world of humanities but also the time to listen to those who are versed in STEM. Overall, much of what we experienced with the cross-training sessions went beyond what we were a part of in the virtual classroom. This type of unified environment is something that we cherish

in our learning environments and recognize as incredibly resourceful in real-life situations. When those who work in the sciences do essential research, the humanities can bring inspirational ways to present such research or offer creative outlets to spread the main takeaways of the research, offering more collaborative solutions.

These reflections offered a nuanced perspective on the ways that the course and the program model worked together to broaden the impact of the experience beyond the students enrolled in the class.

RESEARCH QUESTIONS

RQ1: What are data science students' learning experience and outcomes from this transdisciplinary and sustainability-aware PBL pilots?

RQ2: What are the implications and lessons learned for implementing transdisciplinary and sustainability-aware PBL in data science education?

LEARNING DESIGN FOR DIGITAL DATA STORYTELLING PROJECTS

Course Overview

Table 1 shows the course's weekly schedule and how DDST PBL activities fit with course units that run for 14 weeks. The course covers data wrangling and visualization, followed by statistical reasoning and basic computing. By mid-term, students use these skills for projects, mirroring real-world data science processes. Unlike UC Berkeley's Data 8, our course emphasizes student-driven dataset selection and offers sustained inquiry, leading to concrete data stories. In addition to those defined PBL features, the project was augmented with a peer-assisted learning component through collaboration with two DST community college interns who participated in two cross-training workshops (Table 2) and provided ongoing consultation to project teams.

Week #	1	2	3	4	5	6	7	8	9	10	11	12	13	14
Course Content	Overview	Table Operations		Topic Selection	Data Visualization		Curious Qs	Probability		Statistical Inference				Review
Project Based Learning	Project Timeline	Interest Survey								Weekly Update				Presentation
	Cross-Training Workshop				Workshop #1				Workshop #2					
	Interns Activity	Meet and Greet	Workshop Planning		Workshop Participation	Workshop Planning		Workshop Participation		Consulting/Giving Feedback to Project Teams				

Table 1. The overall course schedule and the alignment of Digital Data Storytelling PBL Activity with the course content.

	Week No.	Data Science Students' Role	Digital Storytelling Interns' Role
Workshop #1	Week 5	Teach DST interns Table Functions	Teach DS students story construction process
Workshop #2	Week 9	Teach DST interns data visualization	Teach DS students digital story implementation techniques

Table 2. The specifics of the two cross-training workshops with data science students (DS students) and community college Digital Storytelling interns (DST interns).

Interest Survey: The project started with an interest survey in week 2 when students were asked to watch a video on human flourishing published by the Templeton Foundation³, reflect on what living a good life looks like and share thoughts on how to use data-driven technology to improve human flourishing. In addition, students identify areas of interest relevant to the empirical evidence of human flourishing (VanderWeele, 2017).

Topic Selection: In week 4, students were asked to finalize their topic selection and teaming. Students were given a list of public data sources which include websites such as kaggle.com and government websites hosting datasets. To encourage students to look into sustainability issues, we also listed climate-related datasets themed around global warming, polar region ice melting, sea-level change, natural disaster, etc.

Curious Questions: After allowing students to take a few weeks to explore potential data sources and zones in the dataset to work on, we asked them to pose a few curious questions based on background reading or some initial understanding of the dataset. Those questions may guide their deeper exploration in the coming weeks. We also offered a mental health data set and an option to design a quantified-self project in which they would collect and analyze their own data.

Please refer to Appendix for a summary of students' responses to the interest survey, topic selection, and curious questions.

Weekly update and feedback: After the mid-term, students were asked to submit a weekly update each with a few data insights they gleaned from analysis, presented in table summaries or plots. Students also need to provide a narrative to tell the stories of the data. The instructor, teaching assistants, and DST interns were asked to provide feedback on those weekly updates. During the next iteration, students were expected to revise their narrative in accordance with their feedback. To compose a complete data story, students must select relevant story slices or data insights to organize it into a logical

sequence and weave it into a compelling and coherent story. This process happens organically later in the process.

Final deliverable: The project culminates in a set of digital data story elements, which includes a set of story slices (in the format of presentation slides) and written scripts of the story narrative. For extra credits, students could produce a digital data story using the video editing tool WeVideo.

Engaging DST interns in Students' Digital Data Storytelling Projects

Both interns are non-STEM majors (communication and business, respectively), and neither has previous experience with data science. Their background brings a fresh perspective to data science students, who are typically STEM students who are trained to mainly think in terms of numbers, abstraction, and computing, often with little exposure to qualitative thinking, narrative, and communication through mediums such as storytelling. As such, interns are well-positioned to serve not only as experts who could coach and provide feedback for students but also as "sounding boards" for students to try out new materials and strategies.

The interns were introduced to the class early in the semester and interacted with faculty and students through the virtual platforms Discord, an online instant messaging and virtual meeting platform. Since over 40% of the students in the data science class are transfer students from community colleges, they helped to build a sense of community for the interns as they explored prospective transfer institutions.

Cross-training workshop

We conducted two cross-training workshops in weeks 5 and 9, respectively, as illustrated in Table 2. The objective of the cross-training workshop is for data science students and DST interns to share their relative strengths in data science and digital storytelling and discover the common ground between Digital Storytelling (deeply rooted in the humanities and social sciences) and Data Science (commonly perceived as the STEM discipline).

The workshop was conducted over Zoom; each session lasted three hours. In each 3-hour session, half of the time, data science students taught their newly learned skills to interns, and half of the time, interns taught data science students digital storytelling skills. The specific topic during the cross-training is outlined in Table 2. This design is motivated by the learning-by-teaching (Chase et al., 2009) empirical evidence by providing opportunities for data science students to teach non-STEM students. We hypothesize this experience will improve their learning and sharpen their cross-disciplinary communication skills when they are challenged to explain technical details to non-technical audiences.

METHODOLOGY: DATA COLLECTION AND ANALYSIS

The study was approved by Institution A's Institutional Research Board. Students' demographic information was collected as part of the pre-course survey. In addition, we also collected students' survey responses at the beginning of the project. Students' work products, including weekly updates and final products, were collected. We conducted semi-structured interviews with three data science students who participated in cross-training sessions. The main objective of the interviews was to gain deep insights into students' learning experiences both as "teachers" to DST interns as well as students learning digital storytelling techniques from DST interns. All interviews were recorded with the consent of the participants and later transcribed verbatim for analysis. The interview protocol consisted of open-ended questions relevant to the research questions.

The qualitative data, including students' responses to open-ended survey questions, interview transcripts were analyzed using thematic analysis, following the six-phase guide in (Braun & Clarke, 2006). This process involved familiarizing ourselves with the data, generating initial codes, searching for themes among codes relevant to research questions, reviewing themes, defining and naming themes, and producing the findings. Excel spreadsheets were utilized to facilitate the organization and analysis of the data.

FINDINGS

In this section, we summarize the findings pertaining to RQ1 on data science students' learning experience and outcomes from this transdisciplinary and sustainability-aware PBL pilot.

Data Science Students' Learning from the two Cross-training Sessions with DST interns

As Teachers of Data Science

Several themes emerged from analyzing the interview transcripts with three data science students.

The positive effect of the teaching experience.

Though with different prior knowledge and various teaching experiences, all students agree that the teaching experience via cross-training provided them with opportunities to sharpen their skills, absorb materials better, "cement understanding," and improve their self-efficacy by teaching others.

We noted two possible pathways leading to enhanced learning which resonates with the literature on learning-by-teaching or protege effects (Chase et al., 2009).

Pathway #1: Motivated to learn deeply to teach effectively.

As commented by Student A3,

"I honestly felt an immense pressure to really know my stuff because I am expected to teach others and sound like I know what I am doing. So naturally, I studied harder and took additional measures such as writing myself a script, creating a cheat sheet, and annotating my personal copy of my exercise question with supplementary information for myself. I did so much to create a safety net for myself, and it paid off because it helped cement my understanding and made me prepared." - Student A3

Likewise, being expected to be the master of the teaching topic motivated them to "absorb materials differently" (A1). For student A2, however, who comes with little teaching background and prior knowledge, the group-coaching format with multiple "teachers" in the break-out room is helpful as they could support each other.

Pathway #2: The real urge to communicate well about something they know.

A secondary pathway is the desire to communicate well about something they know. All students realize that learning alone is one thing, but the bar is much higher if they need to get the idea across. This could be even more challenging if the "students" are from different disciplines. Despite little formal training, students were able to develop their strategies and tactics to tackle this challenge.

For example, A2 uses the strategy of envisioning teaching a younger sibling, as he remarks,

"I tried my best to simplify code and the (main) point of the module, and I tried thinking of ways that I could show it to maybe my younger siblings for example. That's mostly my approach, I just wanted to make it accessible and easily understandable." - Student A2

A3's strategy is a "piecemeal approach of breaking down steps and explaining how they work together to manipulate data." A2 also mentions teaching allows him to develop empathetic thinking. As he noted,

"... So when you're learning for somebody else, you have to think about what type of, I guess, perspective that the other person has in terms of what you're learning." - Student A2

Likewise, A3 mentions that he strives to "make it cohesive to not just myself, but also someone without similar experience as me." This desire to be cohesive may trigger decisions about how to organize knowledge and structure learning (e.g., active learning), thus improving outcomes. Interestingly, students also reflected that in preparing lessons, they need to be ready to improvise and deal with unexpected questions or think on their

feet. However, some students expressed needing to work out more problems to prepare themselves better.

As Learners of Digital Storytelling

Takeaways from being trained in storytelling concepts and techniques.

Student A2 participated in the Story Circle breakout session led by one of the interns. Reflecting on his experience, he feels more confident in his ability to convey data through a data storytelling format, though being a first-timer participant, he felt it was fast-paced and challenging:

“...it got me to think about preparing information to share and the best way to articulate it in a way that is easy to understand and descriptive. .. it was pretty quick, I would say, in terms of the pacing. Because I’m not sure about how the timing worked, but we were taught what digital data storytelling is like, and then we moved into a prompt, and we shared with our groups what we prepared for the prompt. It was pretty (fast-paced) – like we didn’t necessarily have time to write anything down or anything,” - Student A2

Students A1 and A3, who participated in the WeVideo demo session, appreciated the interns’ well-planned, well-taught tutorial session and the excellent organization of the hand-out that helped to “connect the dots” so that they could use the tools to convey quantitative information. Overall, they feel the DST principle and techniques learned are “valuable addition to the knowledge base.”

Students’ Digital Data Stories Product

While all student team deliverables met the minimum requirements of a slide deck of a collection of selected data insights and story scripts, three project teams followed through to create a digital story video. Please refer to Appendix for two examples of digital data stories, one on ***World Happiness*** and another on ***Industrialization, Carbon Emissions, and Climate Change***. In both of the videos, students were able to present compelling and coherent stories backed by data analytics. Following a narrative arc, the stories began with an exposition of the issues and followed up with analysis, concluding with calls to action. Moreover, the application of digital storytelling techniques shared by interns, such as the usage of stock videos, seemed to be quite effective.

Students benefited from the ongoing support from interns; for example, on one occasion, the interns were asked to review a student’s script draft for their video regarding climate change. While the team had a very strong base in delivering the essential facts regarding the issue, the interns’ edits helped to establish a more urgent tone to emphasize the need to address climate change. This is where the artistic choices in digital storytelling complemented and enhanced the data analytics content. While it is important to share

critical information regarding a topic, presenting it compellingly maintains the audience's attention and has a better chance of achieving communication goals.

Students' End-of-course Reflections

Toward the end of the course, students were asked to reflect on takeaways from the courses. Among the 18 submitted reflections, 14 (78%) explicitly mentioned the data storytelling project. Five students believed those skills learned through the projects could be readily applied to support, for example, the sense-making of the data they will encounter in real life. As mentioned by one student, “before (the project), I would just skim over graphs, but now I would stop and try to understand what is being said in the data.” Yet another student speaks to the authenticity of the task by explaining that “the project had us break down datasets and get to apply code and do it the way data scientists do.” Four students saw the immediate value of those skills to their future job. One of the students gave a vivid example of how he could use data storytelling skills to communicate effectively with his co-workers to help to improve their performance. Six students expressed positive sentiment toward the project experience using words such as “enjoy” and “like,” and some are especially proud of the digital data stories they were able to make, as this required them to “break out of the comfort zone” of their discipline. It is worth mentioning that two of the students explicitly mentioned the value of this project in raising awareness of societal issues such as climate change and human well-being.

DISCUSSION

In this paper, we report a novel project-based learning Digital Data Storytelling pilot in an undergraduate introductory data science course. This pilot is built upon a cross-disciplinary and inter-institutional partnership between a university and community college serving diverse student populations. This pilot was designed to explore sustainability from three perspectives: *domain sustainability* (by tasking students with exploring real-world datasets with sustainability themes such as climate change or well-being), *disciplinary sustainability* (through the authentic task of creating public data stories as the result of multidisciplinary collaboration), and *educational sustainability* (by engaging students from a diverse background using an asset-based peer-assisted learning approach through cross-training sessions where data science students and Digital Storytelling interns learn from and teach each other the skills with relative strength). In this section, we summarize the implications and lessons learned for implementing transdisciplinary and sustainability-aware PBL in data science education.

Promises of Sustainability-Aware PBL Learning Design in Data Science Education

Featuring authentic tasks with clear communication goals and social impact, and exploration of peer-assisted learning through cross-training, this project adds to the

limited data science PBL literature in teaching cross-disciplinary communication competency and social responsibility. In particular, the peer-assisted learning format we explored involves students outside the course with different disciplinary majors. This setup presents an authentic communication scenario challenging data science students to hone their cross-disciplinary communication skills, which frequently arise in real-world contexts and are highly valued by employers (Halwani et al., 2021). From the reflections of the cross-training participants, it is evident that the students benefit from being a teacher of their domain when their newly acquired knowledge and skills are reinforced through the protege effect (Duran, 2017). Moreover, they develop the capacity to empathize with learners, which lays the foundation for effective communication across disciplines. Students also seem to benefit from increased self-efficacy in their domain. By presenting students with the opportunity to learn skills out of their comfort zone, we note that students appreciate the challenges and potential values of working with students from different disciplines. Though with only limited evidence, we observe that some students connect with the sustainability themes embedded in the data analysis task, leading to increased awareness about societal issues like climate change and human well-being. Future work could benefit from a more principled approach to measure students' learning outcomes in areas of sustainability competency.

Significance of an Inter-institutional Transdisciplinary Program Model

Our pilot is built on a mature collaborative inter-institutional transdisciplinary program model, which provides important infrastructure for enriching data science students' learning experience via the communication expertise brought by DST interns. Crossing disciplinary boundaries requires concerted efforts, and developing an effective inter-institutional collaboration like ours could take years. For institutes that do not yet have such infrastructure, small initiatives can make a difference. Inviting colleagues from other disciplines (e.g. communication, social science or humanity) to give guest lectures in data science classes, or organizing campus-wide hackathon events to encourage students to form multidisciplinary teams, could open avenues for future interdisciplinary collaboration.

Lesson Learned for Teaching Cross-disciplinary Data-driven Communication

Despite the awareness of its benefit, teaching cross-disciplinary data-driven communication is a relatively new area and can be challenging. This pilot study explored possible formats and collected an initial set of evidence to demonstrate its potential values. However, we identified several areas for future improvement from both teaching practice and research perspectives. Firstly, though it was planned for storytelling interns to provide feedback to data science students while they craft the stories, in reality, only a small number of the teams could take advantage of these resources. It is likely due to the short timeline of the project. Most teams spend most of their time wrestling with data

analysis to derive insights, so they have little time to construct stories. Moreover, we realize that students may benefit from the scaffolding of the feedback-giving and seeking process, for example, by providing a list of example questions that data science students could ask interns. Secondly, though there are theoretical explorations on the narrative structure of data storytelling (Cohn, 2013), it is yet to be operationalized in teaching practices. Students from both disciplines will likely benefit from explicit discussion to draw the parallelization between the story construction in digital storytelling and data storytelling. Story circles explored in the cross-training workshop could be adapted to facilitate feedback and reflection during the data stories construction process later in the project. Thirdly, though this study captures a few students' experiences through written reflections and interviews, a more structured approach to measure more students' learning outcomes in communication competency and growth of domain knowledge will be beneficial to guide future pedagogical decisions and provide more rigorous evidence of its effectiveness and areas of needed growth.

ACKNOWLEDGEMENT

This material is based upon work supported by the US National Science Foundation under Grant No. #1915714 and #2302795.

References

- Adhikari, A., DeNero, J., & Jordan, M. I. (2021). Interleaving Computational and Inferential Thinking: Data Science for Undergraduates at Berkeley. *Harvard Data Science Review*. <https://doi.org/10.1162/99608f92.cb0fa8d2>
- Allen, D. E., Donham, R. S., & Bernhardt, S. A. (2011). Problem-based learning. *New Directions for Teaching and Learning*, 2011(128), 21–29. <https://doi.org/10.1002/tl.465>
- Allen, G. I. (2021). Experiential Learning in Data Science: Developing an Interdisciplinary, Client-Sponsored Capstone Program. *Proceedings of the 52nd ACM Technical Symposium on Computer Science Education*, 516–522. <https://doi.org/10.1145/3408877.3432536>
- BBC (Director). (2010, November 26). *Hans Rosling's 200 Countries, 200 Years, 4 Minutes—The Joy of Stats—BBC Four*. <https://www.youtube.com/watch?v=jbkSRLYSojo>
- Bhargava, R. (2017). *Data Storytelling Studio: Climate Change / Media Arts and Sciences*. MIT OpenCourseWare. <https://ocw.mit.edu/courses/cms-631-data-storytelling-studio-climate-change-spring-2017/>

- Bietti, L. M., Tilston, O., & Bangerter, A. (2019). Storytelling as Adaptive Collective Sensemaking. *Topics in Cognitive Science*, 11(4), 710–732.
<https://doi.org/10.1111/tops.12358>
- Brace, A., Finkelsten, B., & Diadrey-Anne, S. (2015). Evaluating the effectiveness of creating digital stories in a college classroom to promote a healthy food system. *Food Studies*, 6(1), 15–26.
- Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology*, 3(2), 77–101.
<https://doi.org/10.1191/1478088706qp063oa>
- Casola, L. (Ed.). (2020). *Roundtable on Data Science Postsecondary Education: A Compilation of Meeting Highlights*. National Academies Press.
<https://doi.org/10.17226/25804>
- Chan, B. S. K., Churchill, D., & Chiu, T. K. F. (2017). Digital Literacy Learning In Higher Education Through Digital Storytelling Approach. *Journal of International Education Research (JIER)*, 13(1), Article 1.
<https://doi.org/10.19030/jier.v13i1.9907>
- Chase, C. C., Chin, D. B., Oppezzo, M. A., & Schwartz, D. L. (2009). Teachable Agents and the Protégé Effect: Increasing the Effort Towards Learning. *Journal of Science Education and Technology*, 18(4), 334–352.
<https://doi.org/10.1007/s10956-009-9180-4>
- Cohn, N. (2013). Visual Narrative Structure. *Cognitive Science*, 37(3), 413–452.
<https://doi.org/10.1111/cogs.12016>
- Donoghue, T., Voytek, B., & Ellis, S. E. (2021). Teaching Creative and Practical Data Science at Scale. *Journal of Statistics and Data Science Education*, 29(sup1), S27–S39. <https://doi.org/10.1080/10691898.2020.1860725>
- Duran, D. (2017). Learning-by-teaching. Evidence and implications as a pedagogical mechanism. *Innovations in Education and Teaching International*, 54(5), 476–484. <https://doi.org/10.1080/14703297.2016.1156011>
- Dykes, B. (2019). *Effective Data Storytelling: How to Drive Change with Data, Narrative and Visuals*. John Wiley & Sons.
- Farahi, A., & Stroud, J. C. (2018). The Michigan Data Science Team: A Data Science Education Program with Significant Social Impact. *2018 IEEE Data Science Workshop (DSW)*, 120–124. <https://doi.org/10.1109/DSW.2018.8439915>
- Fisler, K. (2021). *Leveraging Data Science and Social-Impact Analysis to Broaden Participation in Introductory Computer Science Courses*. Gesellschaft für Informatik e.V. <http://dl.gi.de/handle/20.500.12116/37014>
- Ghani, R. (2018). Data Science for Social Good and Public Policy: Examples, Opportunities, and Challenges. *The 41st International ACM SIGIR Conference on Research & Development in Information Retrieval*, 3.

<https://doi.org/10.1145/3209978.3210231>

- Halwani, M. A., Amirkiaee, S. Y., Evangelopoulos, N., & Prybutok, V. (2021). Job qualifications study for data science and big data professions. *Information Technology & People*, 35(2), 510–525. <https://doi.org/10.1108/ITP-04-2020-0201>
- Henke, N., Bughin, J., Manyika, J., Chui, M., Saleh, T., Wiseman, B. & Sethupathy, G. (2016). *The Age of Analytics: Competing in a Data-Driven World*. McKinsey Global Institute. <https://www.cosmeticinnovation.com.br/wp-content/uploads/2017/01/MGI-The-Age-of-Analytics-Full-report.pdf>
- Herranen, J., & Aksela, M. (2019). Student-question-based inquiry in science education. *Studies in Science Education*, 55(1), 1–36. <https://doi.org/10.1080/03057267.2019.1658059>
- Hildreth, L. A., Miley, M., Strickland, E., & Swisher, J. (2022). Writing Workshops to Foster Written Communication Skills in Statistics Graduate Students. *Journal of Statistics and Data Science Education*, 0(0), 1–10. <https://doi.org/10.1080/26939169.2022.2138800>
- Kearns, M., & Roth, A. (2019). *The Ethical Algorithm: The Science of Socially Aware Algorithm Design*. Oxford University Press. https://www.google.com/books/edition/The_Ethical_Algorithm/z5OzDwAAQBAJ?hl=en&gbpv=1&dq=ethical+algorithm&pg=PP1&printsec=frontcover
- Knaflic, C. N. (2015). *Storytelling with Data: A Data Visualization Guide for Business Professionals*. John Wiley & Sons.
- Kosara, R., & Mackinlay, J. (2013). Storytelling: The Next Step for Visualization. *Computer*, 46(5), 44–50. <https://doi.org/10.1109/MC.2013.36>
- Lambert, J. (2018). *Digital Storytelling: Capturing Lives, Creating Community*. Routledge.
- Lewis, A., & Stoyanovich, J. (2021). Teaching Responsible Data Science: Charting New Pedagogical Territory. *International Journal of Artificial Intelligence in Education*. <https://doi.org/10.1007/s40593-021-00241-7>
- Liguori, A. (2020). Unlocking Contested Stories and Grassroots Knowledge. In P. P. Trifonas (Ed.), *Handbook of Theory and Research in Cultural Studies and Education* (pp. 465–479). Springer International Publishing. https://doi.org/10.1007/978-3-319-56988-8_35
- McDonald, N., Akinsiku, A., Hunter-Cevera, J., Sanchez, M., Kephart, K., Berczynski, M., & Mentis, H. M. (2022). Responsible Computing: A Longitudinal Study of a Peer-led Ethics Learning Framework. *ACM Transactions on Computing Education*, 22(4), 47:1-47:21. <https://doi.org/10.1145/3469130>
- Monroe-White, T., Wright, B., Hulsey, W., Kushins, E., & Hord, A. (2023). Establishing a Data Science for Good Ecosystem: The Case of ATLytiCS. *The Journal of the Southern Association for Information Systems*, 10(1), 1–19.

<https://doi.org/doi:10.17705/3JSIS.00029>

- National Science Foundation. (2021). *The STEM Labor Force of Today: Scientists, Engineers, and Skilled Technical Workers*. <https://nces.nsf.gov/pubs/nsb20212>
- Nolan, D., & Stoudt, S. (2021). *Communicating with Data: The Art of Writing for Data Science*. Oxford University Press.
- Norman, D. A. (2023). *Design for a Better World: Meaningful, Sustainable, Humanity Centered*. MIT Press.
- O’Neil, C. (2017). *Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy*. Crown.
- P. M. Ribeiro, S. (2016). Developing intercultural awareness using digital storytelling. *Language and Intercultural Communication*, 16(1), 69–82. <https://doi.org/10.1080/14708477.2015.1113752>
- Rambe, P., & Mlambo, S. (2014). Using digital storytelling to externalise personal knowledge of research processes: The case of a Knowledge Audio repository. *The Internet and Higher Education*, 22, 11–23. <https://doi.org/10.1016/j.iheduc.2014.04.002>
- Risam, R. (2018). *New Digital Worlds: Postcolonial Digital Humanities in Theory, Praxis, and Pedagogy*. Northwestern University Press.
- Ryoo, J. J., Morris, A., & Margolis, J. (2021). “What Happens to the Raspado man in a Cash-free Society?”: Teaching and Learning Socially Responsible Computing. *ACM Transactions on Computing Education*, 21(4), 31:1-31:28. <https://doi.org/10.1145/3453653>
- Tomašev, N., Cornebise, J., Hutter, F., Mohamed, S., Picciariello, A., Connelly, B., Belgrave, D. C. M., Ezer, D., Haert, F. C. van der, Mugisha, F., Abila, G., Arai, H., Almiraat, H., Proskurnia, J., Snyder, K., Otake-Matsuura, M., Othman, M., Glasmachers, T., Wever, W. de, ... Clopath, C. (2020). AI for social good: Unlocking the opportunity for positive impact. *Nature Communications*, 11(1), Article 1. <https://doi.org/10.1038/s41467-020-15871-z>
- Topping, K. J. (2005). Trends in Peer Learning. *Educational Psychology*, 25(6), 631–645. <https://doi.org/10.1080/01443410500345172>
- United Nations. (2015). *Transforming our World: The 2030 Agenda for Sustainable Development*. <https://sdgs.un.org/publications/transforming-our-world-2030-agenda-sustainable-development-17981>
- VanderWeele, T. J. (2017). On the promotion of human flourishing. *Proceedings of the National Academy of Sciences*, 114(31), 8148–8156. <https://doi.org/10.1073/pnas.1702996114>
- Williams, J. D., Lopez, D., Shafto, P., & Lee, K. (2019). Technological Workforce and Its Impact on Algorithmic Justice in Politics. *Customer Needs and Solutions*, 6(3),

84–91. <https://doi.org/10.1007/s40547-019-00103-3>

Willis, A., Charlton, P., & Hirst, T. (2020). Developing Students' Written Communication Skills with Jupyter Notebooks. *Proceedings of the 51st ACM Technical Symposium on Computer Science Education*, 1089–1095. <https://doi.org/10.1145/3328778.3366927>

Wu, J., & Chen, D.-T. V. (2020). A systematic review of educational digital storytelling. *Computers & Education*, 147, 103786. <https://doi.org/10.1016/j.compedu.2019.103786>

APPENDIX

Students' Responses to Interest Survey

We analyzed the 23 responses to the open-ended question in the interest survey “In your own words, what does living a good life look like?” Several themes emerged: The most popular concept is related to working, being productive, self-fulfillment and self-improvement, and realizing potential, coming from 48% of the respondents (n=11); followed by a related concept oriented toward meaning, purpose, and goal (39% of respondents, n=9); about 35% (n=8) of the students consider the access to basic resources and a certain level of quality of life as the hallmark for a good life. About 22% of the students believe helping and connecting with others are important factors. The other relatively infrequently mentioned concepts include mental health, subject well-being, and resilience (i.e., handling adversities)

Figure 2 summarizes students' responses to the general area of interest concerning the domains and enabling factors of human flourishing (VanderWeele, 2017).

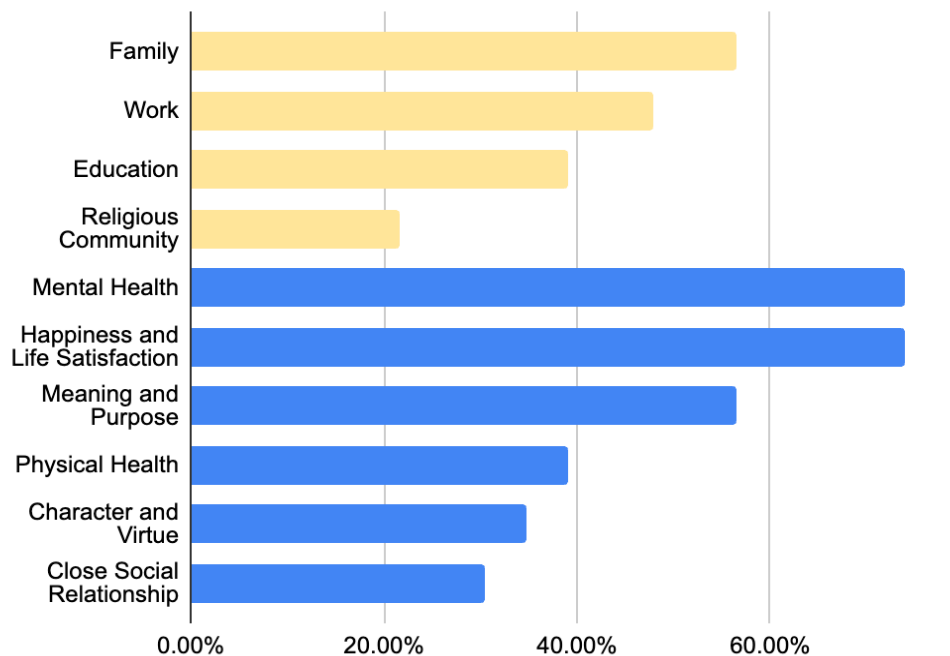


Figure 2. Summary of students' responses to one of the Interest Survey questions: Relating to the empirical evidence of human flourishing per VanderWeele 2017 paper, which of the following areas are you especially interested in exploring? Percentage of students choose a given selection item. The top four selection items (in yellow) are the four enabling factors of human flourishing, and the bottom six are the domain of human flourishing (VanderWeele, 2017).

Students' Topic Selection

Seven students (6 teams) selected the World Happiness project; 8 students (4 teams) selected the Climate Change topic; 8 students (6 teams) opted to work on healthcare topics which included a project investigating gamer's mental health and cardiovascular disease,

as well one students worked on Quantified self project which analyzed the data collected from himself and his friends on well-being related topics using Experience-Sampling Tool Expiwell.

Students' Teaming Choices

We permit students to form teams comprising up to three members. Nine students opted to form teams of three, while the remaining students chose to work independently. Regardless of their team size, all students collaborated with DST interns via cross-training and by participating in consultation or feedback processes.

Students' Curious Questions

After students finalized the topics, they were asked to do some preliminary research on the topic and dataset based on background reading or initial exploration of the data. At this stage, the questions can be large and vague, and the motivation is to empower the students to find an angle of analysis they find interesting. For example, teams working on the World Happiness project are interested in exploring the true definition of happiness and the factors contributing to happiness (e.g., material wealth, law and regulation), the property of happiness (e.g., whether it lasts, and whether everyone has the same definition of happiness, etc.). For teams working on the climate change project, the curious questions were centered around the state of carbon emission and how industrialization might be related and how it can be explained by policy, how natural disasters and severe weather might be linked to climate change, whether the sea-level rise is linked to the current global warming trend.

Example Digital Data Stories Produced by Students

The Story of World Happiness

Link to video: <https://youtu.be/-4XQqu-BXsU>

In this story, the student walked through his analysis step by step toward the revelation of the major contributing factors to the country-level happiness score. His story unfolds by introducing the candidate factors one by one. Here, his use of stock images and videos appears quite effective. He then attracts the audience's attention by posing an intriguing question: Why do some countries like Finland and Denmark scored the highest while countries like Afghanistan slipped to the bottom? Finally, he shared his aha moments of discovery. As he reflected at the end of the video on the meaning of happiness and potential pathways to improve world happiness:

"Happiness isn't just a commodity, it's a goal, something to strive for ... from my findings, the proper steps needed will involve looking at the happiness driving factors for countries, like GDP, Social Support, Healthy Life Expectancy, and seeing to it that countries found on the lower end of the ladder positively achieve these. I feel that with this data, we can

conclude, such as I have, that will point us to the kind of solutions and decisions needed to achieve true world happiness.”

Those narratives shed light on his deepened understanding of well-being related sustainability issues through his journey of data-driven inquiry.

The Story of Industrialization, Carbon Emissions, and Climate Change

Link to video: <https://youtu.be/g8qd-iQljBY>

This story starts with a definition of carbon emission and plausible causes linked to industrialization, population growth, and resource overconsumption, which were substantiated through data analysis. The story presents insights by contrasting the amount of carbon emission from one developed and one developing country from each continent. It follows with a historical account of industrialization and its impact on carbon emissions. The story further explores data showing the disproportional burden of air pollution-related health issues from developing countries compared to developed countries. The story concludes with a statement of the consequences of global warming, accompanied by a call to action by making small individual changes:

“Global warming threatens the survival of all living things and is directly linked to the health of the human species, especially in densely populated areas. Doing things like using reusable bags, using public transport instead of individual vehicles, and turning off electronic devices while not in use are some examples of small changes that many people can accomplish. If we all make small individual changes, we can accomplish great things together to ensure the health of future generations and our one and only planet.” Those narratives illustrate students’ emerging understandings of sustainability issues related to global warming, along with actions they could personally undertake to address them.

Glossary

Story Circle

In a story circle, participants gather (either physically or virtually) to share, listen to, and give feedback on each other's stories. The primary purpose is to create a collaborative, supportive environment for developing narratives.

Curious Questions

Students pose questions out of curiosity, given some basic understanding of the domain and data set. This pedagogy move is inspired by question-driven learning (Herranen & Aksela, 2019) or problem-based learning (D. E. Allen et al., 2011), where we use those student-initiated questions to motivate students in data-driven inquiry.

Minority Serving Institution

A Minority-Serving Institution (MSI) is a term used in the United States to describe colleges and universities that enroll a high percentage of minority students. There are

several types of MSIs, each serving a particular minority group. Some examples include Historically Black Colleges and Universities (HBCUs), Hispanic-Serving Institutions (HSIs), Tribal Colleges and Universities (TCUs), Asian American and Native American Pacific Islander-Serving Institutions (AANAPISIs), or Predominantly Black Institutions (PBIs).

¹ *ESD for 2030 toolbox: Priority action areas*. UNESCO. (2020, November 12). <https://en.unesco.org/themes/education-sustainable-development/toolbox/priorities#paa>

² SDG icons are from <https://www.un.org/sustainabledevelopment/>

³ [Innovations in Human Flourishing | Stories of Impact](#)