INFORMATION SYSTEMS EDUCATION JOURNAL

Volume 24, No. 1
January 2026
ISSN: 1545-679X

In this issue:

iscap
Information Systems & Computing Academic Professionals

The **Information Systems Education Journal** (ISEDJ) is a double-blind peer-reviewed academic journal published by **ISCAP** (Information Systems and Computing Academic Professionals). Publishing frequency is five times per year. The first year of publication was 2003.

ISEDJ is published online (https://isedj.org). Our sister publication, the Proceedings of the ISCAP Conference (https://iscap.us/proceedings) features all papers, abstracts, panels, workshops, and presentations from the conference.

The journal acceptance review process involves a minimum of three double-blind peer reviews, where both the reviewer is not aware of the identities of the authors and the authors are not aware of the identities of the reviewers. The initial reviews happen before the ISCAP conference. All papers, whether award-winners or not, are invited to resubmit for journal consideration after applying feedback from the Conference presentation. Award winning papers are assured of a publication slot; however, all re-submitted papers including award winners are subjected to a second round of three blind peer reviews to improve quality and make final accept/reject decisions. Those papers that are deemed of sufficient quality are accepted for publication in the ISEDJ journal. Currently the target acceptance rate for the journal is under 35%.

Information Systems Education Journal is pleased to be listed in the Cabell's Directory of Publishing Opportunities in Educational Technology and Library Science, in both the electronic and printed editions. Questions should be addressed to the editor at editor@isedj.org or the publisher at publisher@isedj.org. Special thanks to volunteer members of ISCAP who perform the editorial and review processes for ISEDJ.

INFORMATION SYSTEMS EDUCATION JOURNAL

Editors

The Rapid Rise of Generative AI Adoption among First-Year College Students

Mark Frydenberg
mfrydenberg@bentley.edu
Bentley University
Waltham, MA 02452

Kevin Mentzer
kevin.mentzer@nichols.edu
Nichols College
Dudley, MA 01571

Adam Patterson
adam.patterson@uconn.edu
University of Conneticut
Storrs, CT

Abstract

This paper explores the rapid adoption of Artificial Intelligence (AI) tools among 1,597 first-year college students at two New England college campuses during three semesters (Spring 2023, Fall 2023, and Spring 2024), with a specific focus on Generative AI (GAI) tools like ChatGPT. After a comprehensive literature review and empirical analysis, findings indicate a significant increase in GAI awareness and utilization, primarily for academic tasks such as homework assignments, with some use in quizzes and exams. Regression analysis reveals that strong data literacy skills, specifically those related to data discovery, collection, and analysis, are linked to the adoption of AI technologies, while general digital literacy skills such as ability to use productivity applications and databases were not found to have a similar correlation. These results amplify the importance of enhancing data literacy to facilitate effective AI tool integration in academic settings. The study highlights the need for targeted educational strategies to improve data literacy, thereby promoting equitable access to AI technologies and mitigating potential biases. A limitation of this study is that the scope is limited to incoming college students. This research contributes to the understanding of AI adoption dynamics in higher education, providing insights for educators and policymakers to support the ethical and effective use of GAI tools in academic settings.

Keywords: Artificial Intelligence, ChatGPT, Data Literacy, Digital Literacy, Technology Adoption.

Recommended Citation: Frydenberg, M., Mentzer, K., Patterson, A., (2026). The Rapid Rise of Generative AI Adoption among First-Year College Students *Information Systems Education Journal.* v24, n1 pp 4-18. DOI# https://doi.org/10.62273/HVNN2048

The Rapid Rise of Generative AI Adoption among First-Year College Students

Mark Frydenberg, Kevin Mentzer, and Adam Patterson

1. INTRODUCTION

The topic of Artificial Intelligence (AI) is sweeping across college campuses, highlighting the universally high interest in this emerging technology. Faculty are struggling with the ethical issues that arise from the use of AI while also looking for ways to capitalize on this technology for the educational opportunities it enables (Mew & Money, 2024; Zhong & Kim, 2024).

The adoption of AI technologies among students introduces new opportunities and challenges in educational environments (Fuchs & Aguilos, 2023; Murugesan & Cherukuri, 2023). Many students are leveraging AI for various purposes, from generating ideas and conducting research to completing assignments and preparing for exams (Fuchs & Aguilos, 2023; Jovanović & Campbell, 2022). This widespread use signifies a shift in how knowledge is acquired and processed, prompting a need to reassess traditional educational models (BaïDoo-Anu & Owusu Ansah, 2023). Understanding the patterns of AI usage among students can offer insights into their learning behaviors and preferences, which, in turn, can inform the development of more effective teaching strategies and policies.

The rapid rise of generative AI tools into the student experience necessitates understanding more about the students using them and for what purposes. This research aims to contribute to the increasing literature on uses of AI in education by considering first-year college student behavior and attitudes toward AI tools. The following research questions drive this analysis:

RQ1: How quickly are first-year college students adopting the use of AI?
RQ2: What demographic characteristics and technology competencies describe first-year college students who are likely to adopt AI tools?
RQ3: For what purposes are first-year college students using AI tools to support their academic work?
RQ4: How do first-year college students perceive the ethics of AI usage in the classroom, and what factors influence their perceptions?

2. LITERATURE REVIEW: TECHNOLOGY ADOPTION AND AI

The use of generative Artificial Intelligence tools have been a major topic of debate in academia since Chat Generative Pretrained Transformer (ChatGPT) was released in November 2022 by OpenAI (Murugesan & Cherukuri, 2023). Generative AI tools use large language models that generate human-like responses from inputs, or prompts, through natural language processing and statistics (Jovanovic & Campbell, 2022). In addition to generating text, some generative AI tools also create images and video, generate speech, and write and debug code. The technology has a wide range of expertise and domain knowledge as the models are trained on over 175 billion parameters, making it ideal for personalized learning in academia (Brown et al., 2020).

Generative AI tools have experienced swift adoption within students at higher education institutions (Wong, 2024). For educators, the benefits of novel, creative, and always available learning are often compared against the potential for undermining academic integrity and increased bias (Eke, 2023). For students, the adaptive and interactive learning is often weighed against ethical and proper use of the technology (BaïDoo-Anu & Owusu Ansah, 2023).

According to Baldassarre et al. (2023), a primary area of concern is the potential for adoption bias increasing social inequality. Social injustice may be enhanced as learning outcomes become unfair to students that do not use AI tools (Chan & Wu, 2023). Thus, it is of interest for educators to better understand mechanisms driving adoption in creating interventions for technology laggards to embrace GAI tools (Venkatesh, 2022).

Kasneci et al. (2023) report that a clear strategy is needed for educational institutions to optimize GAI in educational settings. To optimize learning with GAI, insights about student usage are imperative in closing the theoretical gap as potential influencing factors may be properly addressed (Fuchs & Aguilos, 2023). According to Hwang and Chen (2023), it is predicted that GAI

applications will increase rapidly in the coming years, thus it is imperative to understand student usage in creating efficient learning strategies.

Sharples (2023) details adoption of GAI as a social process of exploration dependent not only on AI systems themselves, but with interactions among other students. According to Ivanov et al. (2024), social impact has potential to be a major driver of behavioral intention within AI adoption at academic institutions. Jo and Bang (2023) corroborate these findings by reporting that social influence may be an essential driver of ChatGPT adoption at the University level through evaluation of theoretical technology acceptance models. Empirical research by Gupta (2024) denotes social influence, domain experience, technology familiarity, and training to be factors that impact GAI adoption. However, empirical research also reports that social influence does not play an important role in student usage of ChatGPT (Matalka et al., 2024).

A study by Saif et al. (2024) confirms the usefulness of Technology Acceptance Model (TAM) regarding GAI usage. The Unified Theory of Acceptance and Use of Technology (UTAUT) model is shown to align with ChatGPT adoption (Strzelecki, 2023). Literature surrounding the Value-based Adoption Model (VAM) is less prominent in relation to GAI. The VAM model compares perceived benefits and costs among users when considering adoption (Sohn & Kwon, 2020). Although not specifically utilizing the TAM framework, research by Tiwari et al. (2023), found that student adoption of GAI is driven primarily by usefulness, social presence, and enjoyment; however, perceived ease of use was not found to significantly determine usage.

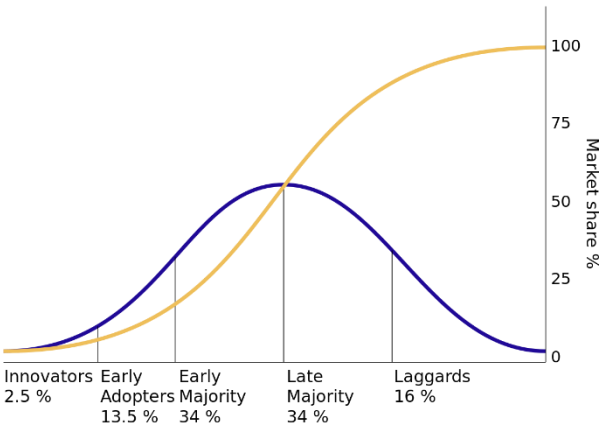


Figure 1. Rogers Diffusion of Innovation Curve

Rogers (1962) developed the Diffusion of Innovations Theory (DOI) to explain how new ideas or technologies spread among a social group (See Figure 1). According to Rogers, innovators are the first adopters of technology and are characterized by being willing to take risks. Early adopters follow and are influential in the technological adoption process. Early majority users adopt innovations after they have been proven by early adopters. The late majority are more skeptical and join later after the majority have adopted a technology. Laggards resist change and prefer more traditional approaches.

Raman et al. (2023) state that understanding adopter categories utilizing Rogers' DOI assists in strategic and successful diffusion of technological innovations by examination of innovators and early adopters. Although extensive literature exists detailing the usefulness of TAM and DOI for new technologies, a literature review specifically for large language models and GAI showed that an acceptance of established models in education exists, but with doubts (Baytak, 2023).

3. METHODOLOGY

This descriptive research study aims to gain insights into the adoption and usage patterns of generative AI applications by first-year college students at two different business-focused New England institutions. This study extends the work of McCarron & Frydenberg (2023) which focused on digital skills of first-year students, and Mentzer, et al. (2024), which focused on their data literacy skills.

Limiting this study to first-year students provides insights into the AI skills they have upon entering college. This research analyzes students' AI adoption during a three-semester period beginning with Spring 2023, just weeks after ChatGPT was introduced.

Sample

Students enrolled in either the "Solving Business Problems with Information Technology" (Bentley University) or "Introduction to Data Literacy" (Nichols College) were offered this survey. These courses primarily serve first-year students in each university and include students across all majors. Most students in these courses have not selected a major this early in their college careers. Both courses are mandatory introductory technology courses for all first-year students at their respective institutions. A total of 1597 participating students completed this survey

during the first weeks of the Spring 2023 (n=522), Fall 2023 (n=610), and Spring 2024 (n=465) semesters.

We surveyed the enrolled students regarding their own familiarity with and usage of AI tools. To understand characteristics of first-year students who were likely to adopt ChatGPT usage, we adopted the series of data literacy and digital literacy questions as developed by Mentzer, et al. (2024). In addition, to ascertain their self-assessed familiarity with technology, we asked the respondents whether they felt they were "tech savvy," which we define as "the extent to which one is informed or proficient about the use of digital technologies and devices (Mentzer et al., 2024)."

To understand whether students were using AI and/or familiar with the technology we asked survey questions shown in Appendix I.

Independent Variables

Demographics
- Age
- Gender
- Ethnicity
- High School Location
- First Generation Student

Data Literacy
- Reading: Data Discovery
- Reading: Quality – Trustworthiness
- Reading: Quality – Errors
- Writing: Collection
- Writing: Management
- Comprehension: Analysis
- Comprehension: Interpretation
- Comprehension: Visualization
- Comprehension: Presentation
- Comprehension: Decision Making

Digital Literacy
- Application Usage
- Word Processing
- Spreadsheets
- Presentations
- Database Tasks
- Operating System Tasks
- Cloud Tasks
- Web Tasks
- Media Tasks

Control Variables:
- Tech Savviness
- Others' use of ChatGPT
- Semester

Table 1. Independent and Control Variables

A logistic regression model was used to determine characteristics driving adoption of ChatGPT (Cox, 1959). The dependent variable was ChatGPT usage with responses coded "non-use" if the respondent replied they had never heard of it or had never used it and coded "use" for all other responses.

$$\ln\left(\frac{p}{1-p}\right) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_n X_n$$

Independent variables included demographics, data literacy skills, digital literacy skills, and tech savviness. We controlled for the semester to understand likelihood of adoption over time (Lever et al., 2016). Table 1 shows the independent variables and control variables used in our analysis.

The logistic regression model makes predictions based on the log-odds of the positive outcome (i.e. ChatGPT Use). The parameters are exponentialized to deliver a more interpretable result, odds ratio, rather than log odds ratio (Nick & Campbell, 2007). Parameter probabilities are calculated by dividing the odds ratio by 1 plus the odds ratio, and then subtracting the baseline model probability (Huang & Moon, 2013; Liberman, 2005).

4. RESULTS

This section discusses the initial adoption patterns for GAI tools, adoption characteristics, and student use during the three-semester timeframe since GAI became widely available.

Initial Adoption Patterns

As can be seen in Figure 3, at the start of the Spring 2023 semester shortly after the release of ChatGPT, many first-year students (66.67%) had neither used ChatGPT nor known other students using it. One year later (Spring 2024) this had dropped to just 4.21% not using nor knowing others using the tool. Personal use went from 15.63% in Spring 2023 to 80.93% usage in Spring 2024.

Initially, we would expect that students do not use technology, nor do they know people who do (C). As the word of the technology spreads, people may know a few early adopters, though most still will not have used or tried the technology themselves (A), or alternatively, some users might adopt the technology independently (D). Later, we would expect users to actively use the technology themselves. As mass adoption is obtained, personal usage, along with knowing others who use the technology would occur (B).

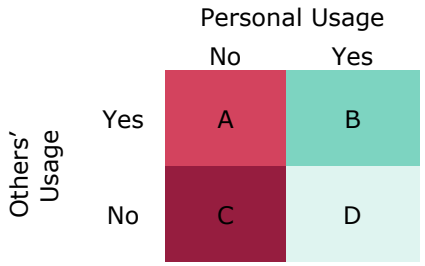


Figure 2. Personal vs Others' Usage

We looked across the immediate three semesters (since the release of ChatGPT in November 2022) to determine trends for how students and their peers used AI tools, as shown in Figure 3.

As can be seen in Figure 3, at the start of the Spring 2023 semester shortly after the release of ChatGPT, many students (66.67%) had neither used ChatGPT nor known other students using it. One year later (Spring 2024) this had dropped to just 4.21% not using nor knowing others using the tool. Personal use went from 15.63% in Spring 2023 to 80.93% usage in Spring 2024.

These numbers suggest a rapid progression through Rogers' Diffusion of Innovation curve. Within months of release of ChatGPT, with personal usage at 15.63%, we were nearing the end of the Early Adopters phase and approaching the Early Majority. By Fall 2023 with personal usage at 53.63% we had moved through the Early Adopters and moved into Late Majority.

By Spring 2024 we were approaching the end of Late Majority and nearing the Laggards phase. If this trend continues then it suggests use of ChatGPT will be in the Laggard phase in fewer than 18 months from its initial release.

Adoption Characteristics

Using the Personal Usage variable as our target variable and the independent variables in Table 1, we ran our logistic regression model. Appendix II shows the results of this model. In logistic regression, confidence intervals that contain the value 1 do not reject the null hypothesis for each parameter. In contrast, traditional ordinary least squares techniques reject the null hypothesis if the parameter contains the value 0.

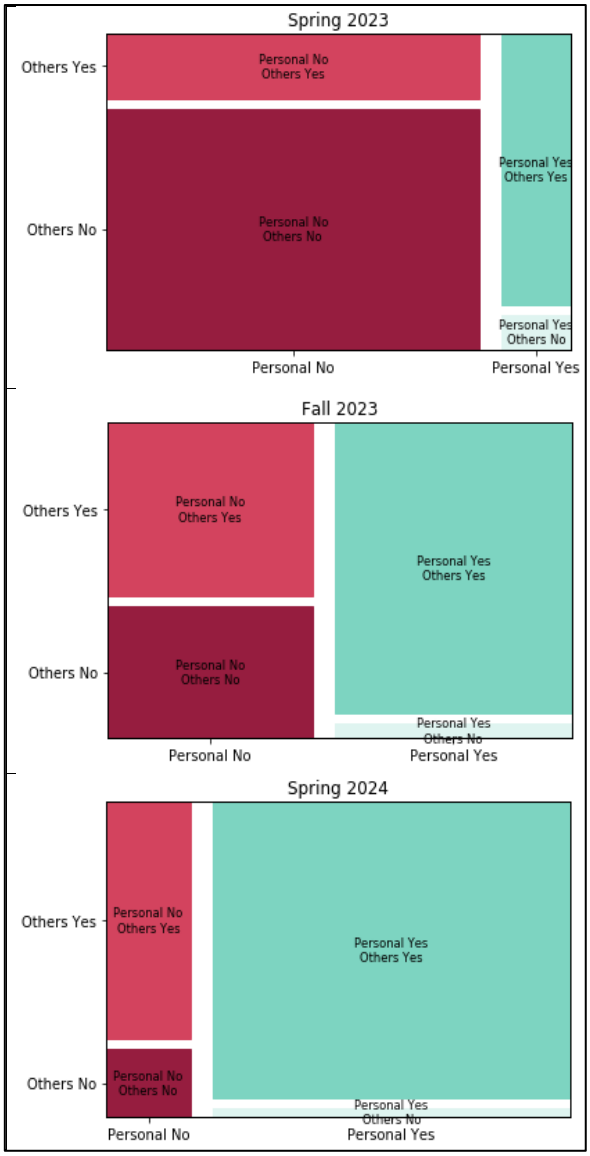


Figure 3. Generative AI Usage Trends of First-Year College Students and their Peers

Turning first to the demographic variables related to the first-year students surveyed, in Appendix II, gender and high school location were significant. Males had a 70% increase in probability of using ChatGPT over females holding all other variables constant, while international students had a 68% increase over U.S. students, holding all other variables constant. Age, ethnicity, and first-generation classification were non-significant. Students may come from high schools where they had (or did not have) exposure to or access to new technologies, and that could impact their own willingness to adopt AI technologies.

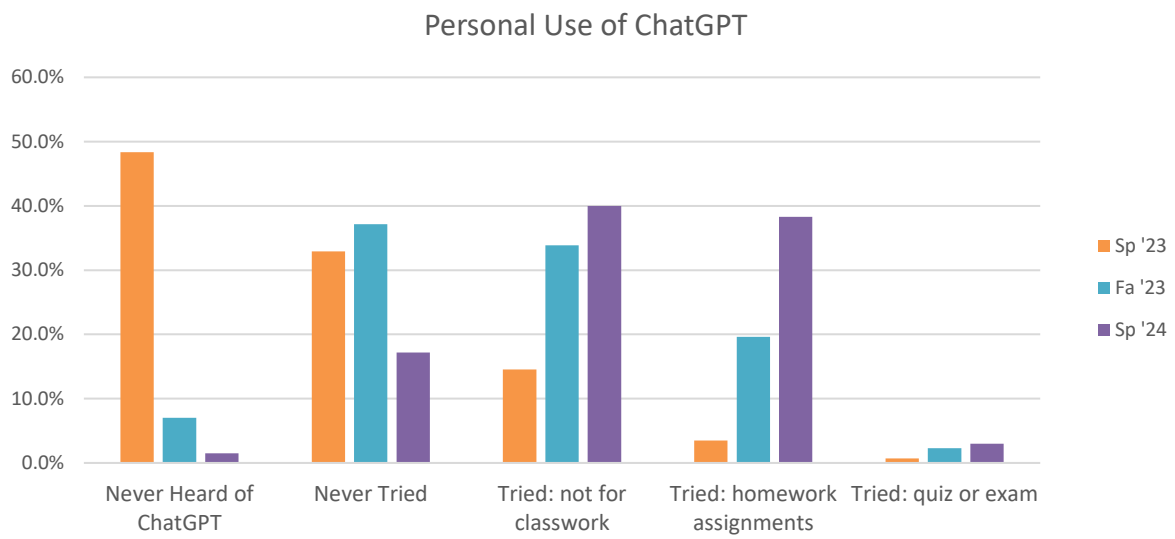


Figure 4. Personal Use of ChatGPT

In Appendix III, the 95% confidence interval does not intersect the value of 1 (shown by the red dashed line) for parameters such as gender and tech savviness, indicating a significant positive effect of these factors. The interval intersects the value 1 for web tasks, visualization, and several other factors, suggesting that these do not have a statistically significant effect on ChatGPT usage. Data and AI skills are hypothesized to be related as according to Schüller (2022), data literacy and AI literacy skills intersect within educational standards. While research by Ng et al. (2022) indicate that AI literacy should emerge as a new literacy skill set in response to digital literacy, our results corroborate that AI literacy is more related to data literacy skills and has less in common with digital literacy. Upon examining data literacy and digital literacy skills, no digital literacy skills are significant while the data literacy skills of Discovery, Collection, and Analysis are significant.

The significant variables within data literacy have differing influences on AI usage. Students with higher self-reported skills in Discovery and Collection have greater likelihood of using ChatGPT. For each one unit increase in Discovery and Collection skills, the likelihood of students using ChatGPT increased by 16% and 25%, respectively. Conversely, respondents with higher skills in Analysis have decreased odds of using ChatGPT. For every one unit increase in Analysis skills, the likelihood of using ChatGPT decreased by 21%. Students with increased analysis skills may not use ChatGPT if they are confident in their own abilities to analyze information.

These skills directly relate to understanding, managing, and analyzing information which are important when using ChatGPT and other AI tools. Students need to be able to validate that AI-generated results are accurate.

The lack of significance of digital literacy skills suggests that basic competencies in using digital tools do not significantly influence AI adoption. This might be because of the perception that as digital natives, most students have basic digital literacy skills already, even though prior research has shown otherwise (McCarron & Frydenberg, 2023). While many students are proficient in launching a web browser and navigating to the ChatGPT website to interact with it, they lack the ability to critically evaluate the information it generates. (Mentzer et al., 2024). A study by Paris (2002) showed that a majority of students lack critical thinking processes when asked questions about websites that they were visiting. The ease of use of AI tools lessens the influence of a student's digital literacy skills as a differentiating factor between those who have adopted the use of AI tools, and those who have not.

The significance of some of the data literacy skills and the insignificance of all digital literacy skills in predicting AI adoption suggests that proficiency

in handling data is indicative of students' likelihood to use AI tools. This underscores the need for providing instruction to enhance students' data literacy to foster greater AI adoption, as these skills are evidently more relevant to effectively engaging with AI technologies.

Finally considering the control variables, all three were significant. Students who consider themselves more Tech Savvy had a 58% increase in probability of using ChatGPT holding all other variables constant, students who knew others who used ChatGPT had a 74% increase in probability of use, and as each semester passed there was a 74% increase in the probability of using ChatGPT compared to the prior semester.

These results suggest that gender, high school location, certain data literacy skills, tech savviness, and contextual factors like the use of ChatGPT by others and the academic semester significantly impact students' adoption of ChatGPT.

Student Use

Figure 4 shows the results when first-year college students were asked how they personally use ChatGPT. While 48.3% of students had never heard of ChatGPT in the beginning of Spring 2023 semester, this number dropped to just 7.0% in the Fall 2023 semester and 1.5% by spring 2024. This highlights how rapidly knowledge of AI spread among college age students.

Combining the "Never Heard of" and "Never Tried" we see that accounted for a total of 80.2% of students surveyed in the spring 2023, this number dropped to 44.2% in the Fall 2023 and dropped further to 18.7% by the Spring 2024. This suggests a rapid familiarity with the tool spreading between Spring 2023 and Spring 2024.

Survey questions were structured to increase in severity (See Appendix I for questions). We see a rapid increase from Spring 2023 to Fall 2023 for those Trying the tool but not for classwork. From Fall 2023 to Spring 2024 we see a rapid increase of those using it for homework assignments (but not quizzes or exams). As expected, this suggests that students were getting comfortable with the tools before using them for classwork and then subsequently used them for classwork. Across all three semesters there is low usage for ChatGPT to help with quizzes and exams. This could be a result of waiting one more semester before applying it to quizzes and exams (if it follows the same trend of moving from non-classwork to homework usage) or could be due to self-reporting bias. We will explore this more next.

Others' Use

We then asked students how other first-year college students were using ChatGPT (Figure 5). As expected, those who weren't familiar with ChatGPT remained consistent. By the Spring 2024 semester, 43.4% of respondents indicated they knew others who were using it for classwork but not quizzes or exams. 29.5% knew of others who were using it to help with quizzes or exams. This latter finding suggests that a low value for self-reported use for quizzes or exams could be a sign of self-reporting bias. Alternatively, first-year students could be familiar with a few well-known students who have used it for quizzes or exams. Overall, 72.9% of these students know of other students using ChatGPT for either homework assignments or quizzes or exams.

Generative AI Helpfulness

Our next set of questions focused on the perceived helpfulness of Generative AI tools (see Figure 6). While participants responded based on a 5-point Likert scale (Strongly Disagree, Disagree, Neutral, Agree, Strongly Agree), to simplify the presentation, Figure 6 combines the Strongly Disagree/Disagree and Agree/Strongly Agree categories. Figure 6 has ordered the questions from those applications least likely to be perceived useful to most likely be perceived as useful, as indicated by the increasing values of the green bars. This shows us that students think usefulness in their job is least applicable while usefulness for personal tasks is most applicable.

These results suggest that students find Generative AI tools helpful for a variety of tasks, most notably personal and homework or academic related tasks. The lower response for "in my job" may be because most first-year students who completed this survey are not employed or have not had the opportunity to use Generative AI tools in a professional setting.

It is interesting to note that students find Generative AI more useful in writing essays or papers than for writing job or college applications. One reason for this could be that they are more likely to have exposure to Generative AI recently, and therefore, have written essays and papers more recently than job or college applications. This result could also reflect that students do not have as much confidence in using Generative AI tools for the personal reflection that is needed for job or college applications.

ChatGPT Use

Our final set of questions is meant to summarize

feelings about the ethics of ChatGPT usage (see Figure 7). Usage for college work correlates strongly with a student's perception of whether it is acceptable to use it for coursework with approximately 60% of students indicating that they disagree that they will use it for coursework. 43.1% of first-year students are unsure or disagree that it is acceptable to use Generative AI tools even for personal use.
This suggests that these students in general are unsure when it is appropriate to use Generative AI tools and when it is not appropriate.

Finally considering the control variables, all three were significant. Students who consider themselves more Tech Savvy had a 58% increase in probability of using ChatGPT holding all other variables constant, students who knew others who used ChatGPT had a 74% increase in probability of use, and as each semester passed there was a 74% increase in the probability of using ChatGPT compared to the prior semester.

These results suggest that gender, high school location, certain data literacy skills, tech savviness, and contextual factors like the use of ChatGPT by others and the academic semester significantly impact students' adoption of ChatGPT.

Student Use
Figure 4 shows the results when students were asked how they personally use ChatGPT. While 48.3% of students had never heard of ChatGPT in the beginning of Spring 2023 semester, this number dropped to just 7.0% in the Fall 2023 semester and 1.5% by spring 2024. This highlights how rapid knowledge of AI spread among college age students.

Combining the "Never Heard of" and "Never Tried" we see that accounted for a total of 80.2% of students surveyed in the spring 2023, this number dropped to 44.2% in the Fall 2023 and dropped further to 18.7% by the Spring 2024. This suggests a rapid familiarity with the tool spreading between Spring 2023 and Spring 2024.

Survey questions were structured to increase in severity (See Appendix I for questions). We see a rapid increase from Spring 2023 to Fall 2023 for those Trying the tool but not for classwork. From Fall 2023 to Spring 2024 we see a rapid increase of those using it for homework assignments (but not quizzes or exams). As expected, this suggests that students were getting comfortable with the tools before using them for classwork and then subsequently used them for classwork. Across all three semesters there is low usage for ChatGPT to help with quizzes and exams. This could be a result of waiting one more semester before applying it to quizzes and exams (if it follows same trend of moving from non-classwork to homework usage) or could be due to self-reporting bias. We will explore this more next.

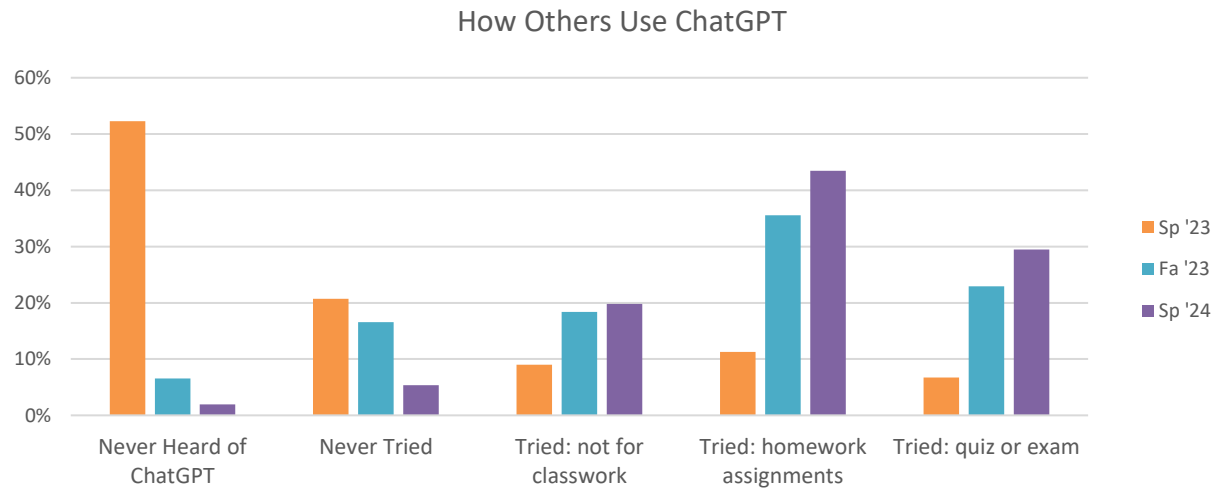


Figure 5. How Others Use ChatGPT

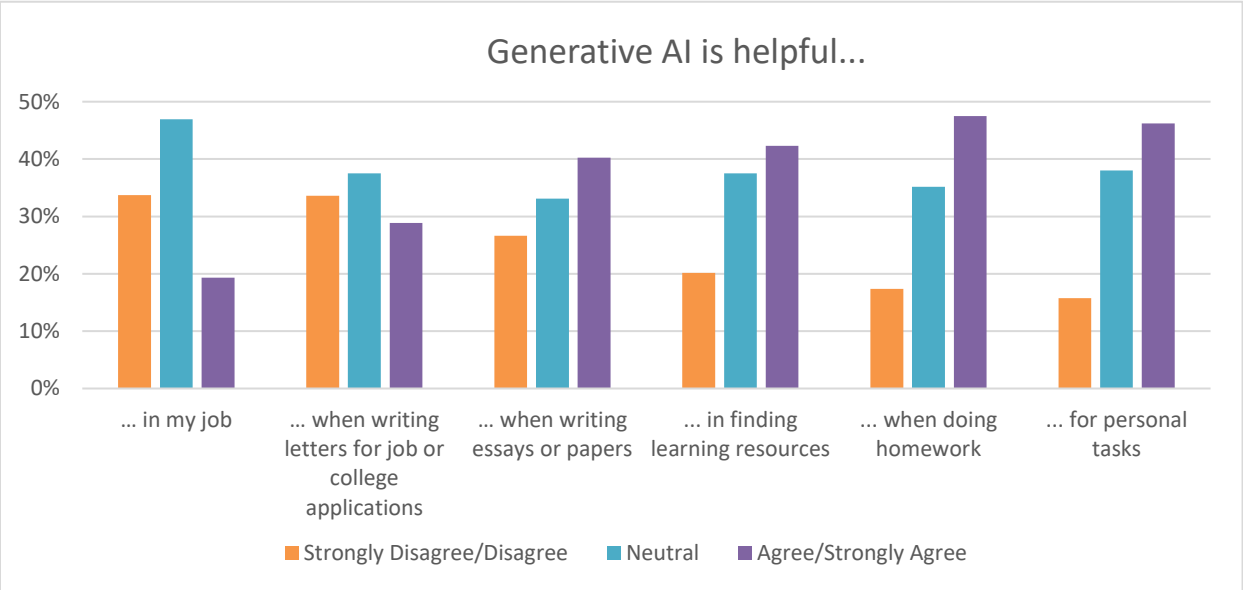


Figure 6. Perceived Helpfulness of Generative AI Tools

Others' Use

We then asked students how other students were using ChatGPT (Figure 5). As expected, those who weren't familiar with ChatGPT remained consistent. By the Spring 2024 semester, 43.4% of respondents indicated they knew others who were using it for classwork but not quizzes or exams. 29.5% knew of others who were using it to help with quizzes or exams. This latter finding suggests that low self-reported use for quizzes or exams could be a sign of self-reporting bias. Alternatively, students could be familiar with a few well-known students who have used it for quizzes or exams. Overall, 72.9% of students know of other students using ChatGPT for either homework assignments or quizzes or exams.

Generative AI Helpfulness

Our next set of questions focused on the perceived helpfulness of Generative AI tools (see Figure 6). While participants responded based on a 5-point Likert scale (Strongly Disagree, Disagree, Neutral, Agree, Strongly Agree), to simplify the presentation, Figure 6 combines the Strongly Disagree/Disagree and Agree/Strongly Agree categories. Figure 6 has ordered the questions from those applications least likely to be perceived useful to most likely be perceived as useful, as indicated by the increasing values of the green bars. This shows us that students think usefulness in their job is least applicable while usefulness for personal tasks is most applicable.

These results suggest that students find Generative AI tools helpful for a variety of tasks, most notably personal and academic related tasks. The lower response for "in my job" may be because most first-year students who completed this survey are not employed or have not had the opportunity to use Generative AI tools in a professional setting.

It is interesting to note that students find Generative AI more useful in writing essays or papers than for writing job or college applications. One reason for this could be that they are more likely to have exposure to Generative AI recently, and therefore, have written essays and papers more recently than job or college applications. This result could also reflect that students do not have as much confidence in using Generative AI tools for the personal reflection that is needed for job or college applications.

ChatGPT Use

Our final set of questions is meant to summarize the students' feelings about the ethics of ChatGPT usage (see Figure 7). Usage for college work correlates strongly with a student's perception of whether it is acceptable to use it for coursework with approximately 60% of students indicating that they disagree that they will use it for coursework. 43.1% of students are unsure or disagree that it is acceptable to use Generative AI tools even for personal use.

This suggests that students in general are unsure when it is appropriate to use Generative AI tools and when it is not appropriate.

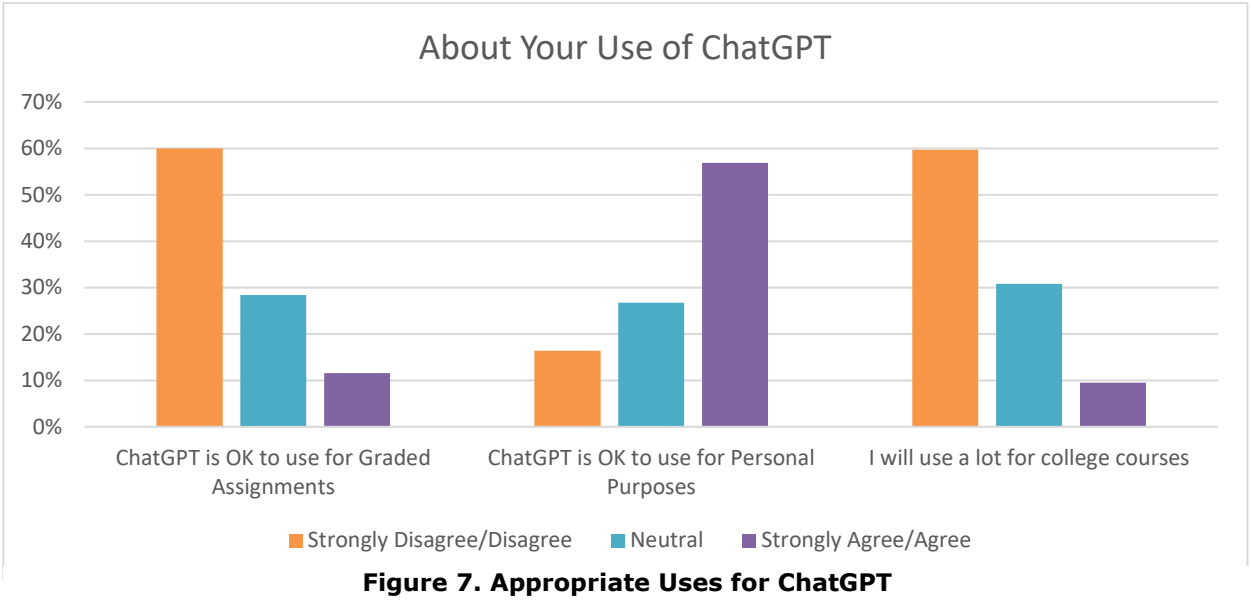


Figure 7. Appropriate Uses for ChatGPT

5. DISCUSSION AND CONCLUSIONS

This discussion summarizes the findings as they relate to the research questions that guided this study.

RQ1: How quickly are first-year college students adopting the use of AI?

Our results demonstrate a clear progression of increased awareness, experimentation, and adoption of ChatGPT. The rapid adoption of ChatGPT among first-year college students during a 12-month period demonstrates the importance of including ample instruction for using it effectively and responsibly in personal, academic, and professional contexts. Initially, most students had not used nor known peers who used ChatGPT, but as the tool became more widely used, this trend shifted drastically. By Spring 2024, nearly 81% of all students used ChatGPT and 72.9% knew others who were using ChatGPT for classwork (homework or exams). These results demonstrated a rapid movement through Rogers' Diffusion of Innovation Curve from early adopters to late majority in 18 months.

RQ2: What demographic characteristics and technology competencies describe first-year college students who are likely to adopt AI tools?

Students who identified as male and international students had higher likelihood of using ChatGPT. Those with data literacy skills are more likely to use ChatGPT for more complex tasks. Our study concludes that data literacy, gender, tech savviness and social influence all contribute to students' adoption of AI tools. College curricula need to emphasize data literacy to encourage students to use Generative AI tools and emerging technologies effectively and efficiently.

It is possible that students with increased analysis skills do not use ChatGPT as they are confident in their own abilities to analyze information.

RQ3: For what purposes are first-year college students using AI tools to support their academic work?

As students become more familiar with the capabilities and benefits of Generative AI tools like ChatGPT, they are more likely to integrate them into their academic work, particularly for homework and, to a lesser extent, for assessments such as quizzes and exams.

RQ4: How do first-year college students perceive the ethics of AI usage in the classroom, and what factors influence their perceptions?

Our data indicates that students are still uncertain about how appropriate it is to use ChatGPT and that students are looking to see how their peers are using it. This suggests that colleges and universities need to offer stronger guidance to students as to the appropriateness of using Generative AI tools such as ChatGPT.

Limitations and Future Research

This study emphasizes the critical importance of

having data literacy skills to fully utilize AI tools. AI literacy goes beyond traditional data analysis and interpretation and requires an understanding of the limitations and capabilities of AI tools, the ability to recognize bias in and reliability of its results, and ethical considerations regarding its use. While this research contributes to our understanding of AI adoption among first-year college students, future research will focus on identifying specific skills that contribute to AI literacy and measuring the extent to which students have these skills, so they can engage with AI tools as informed and responsible users.

Analysis of survey data from the Fall 2024 semester regarding generative AI usage trends of students and their peers as shown in Figure 3 did not reveal any significant differences from the Spring 2024 semester. This suggests that the adoption of GAI tools such as ChatGPT has plateaued. This suggests that students who are likely to adopt the tool have already done so, and the few who have not used them are not likely to do so. This behavior aligns with Rogers' Diffusion of Innovation Curve, as laggards are least likely to adopt new technologies.

This study emphasizes the critical importance of having data literacy skills to fully utilize AI tools. AI literacy goes beyond traditional data analysis and interpretation and requires an understanding of the limitations and capabilities of AI tools, the ability to recognize bias in and reliability of its results, and ethical considerations regarding its use. While this research contributes to our understanding of AI adoption among college students, future research will focus on identifying specific skills that contribute to AI literacy and measuring the extent to which students have these skills, so they can engage with AI tools as informed and responsible users.

6. REFERENCES

BaïDoo-Anu, D., & Owusu Ansah, L. (2023). Education in the Era of Generative Artificial Intelligence (AI): Understanding the Potential Benefits of ChatGPT in Promoting Teaching and Learning. *Journal of AI*, *7*(1), 52–62. https://doi.org/10.61969/jai.1337500

Baldassarre, M. T., Caivano, D., Fernandez Nieto, B., Gigante, D., & Ragone, A. (2023). The Social Impact of Generative AI: An Analysis on ChatGPT. *Proceedings of the 2023 ACM Conference on Information Technology for Social Good*, 363–373. https://doi.org/10.1145/3582515.3609555

Baytak, A. (2023). The Acceptance and Diffusion of Generative Artificial Intelligence in Education: A Literature Review. *Current Perspectives in Educational Research*, *6*(1), Article 1. https://doi.org/10.46303/cuper.2023.2

Brown, T., Mann, B., Ryder, N., Subbiah, M., Kaplan, J. D., Dhariwal, P., Neelakantan, A., Shyam, P., Sastry, G., Askell, A., Agarwal, S., Herbert-Voss, A., Krueger, G., Henighan, T., Child, R., Ramesh, A., Ziegler, D., Wu, J., Winter, C., … Amodei, D. (2020). Language Models are Few-Shot Learners. *Advances in Neural Information Processing Systems*, *33*, 1877–1901. https://proceedings.neurips.cc/paper/2020/hash/1457c0d6bfcb4967418bfb8ac142f64a-Abstract.html

Cox, D. R. (1959). The Regression Analysis of Binary Sequences. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, *21*(1), 238–238. https://doi.org/10.1111/j.2517-6161.1959.tb00334.x

Fuchs, K., & Aguilos, V. (2023). Integrating Artificial Intelligence in Higher Education: Empirical Insights from Students about Using ChatGPT. *International Journal of Information and Education Technology*, *13*(9), 1365–1371. https://doi.org/10.18178/ijiet.2023.13.9.1939

Gupta, V. (2024). An Empirical Evaluation of a Generative Artificial Intelligence Technology Adoption Model from Entrepreneurs' Perspectives. *Systems*, *12*(3), Article 3. https://doi.org/10.3390/systems12030103

Huang, F. L., & Moon, T. R. (2013). What Are the Odds of That? A Primer on Understanding Logistic Regression. *Gifted Child Quarterly*, *57*(3), 197–204. https://doi.org/10.1177/0016986213490022

Hwang, G.-J., & Chen, N.-S. (2023). Editorial Position Paper: Exploring the Potential of Generative Artificial Intelligence in Education: Applications, Challenges, and Future Research Directions. *Educational Technology & Society*, *26*(2), Article 2. https://www.jstor.org/stable/48720991

Jo, H., & Bang, Y. (2023). Analyzing ChatGPT adoption drivers with the TOEK framework. *Scientific Reports*, *13*(1), Article 1. https://doi.org/10.1038/s41598-023-49710-0

Jovanović, M., & Campbell, M. (2022). Generative Artificial Intelligence: Trends and Prospects. *Computer*, *55*(10), Article 10. https://doi.org/10.1109/MC.2022.3192720

Kasneci, E., Sessler, K., Küchemann, S., Bannert, M., Dementieva, D., Fischer, F., Gasser, U., Groh, G., Günnemann, S., Hüllermeier, E., Krusche, S., Kutyniok, G., Michaeli, T., Nerdel, C., Pfeffer, J., Poquet, O., Sailer, M., Schmidt, A., Seidel, T., … Kasneci, G. (2023). ChatGPT for good? On opportunities and challenges of large language models for education. *Learning and Individual Differences*, *103*, 102274. https://doi.org/10.1016/j.lindif.2023.102274

Lever, J., Krzywinski, M., & Altman, N. (2016). Logistic regression: Regression can be used on categorical responses to estimate probabilities and to classify. *Nature Methods*, *13*(7), 541–543.

Liberman, A. M. (2005). How Much More Likely? The Implications of Odds Ratios for Probabilities. *American Journal of Evaluation*, *26*(2), 253–266. https://doi.org/10.1177/1098214005275825

Matalka, M., Badir, R., Ahmad, A., Al-Said, K., Nassar, H., Alzoubi, S., & Alzoubi, M. (2024). The adoption of ChatGPT marks the beginning of a new era in educational platforms. *International Journal of* Data and Network Science, 8(3), Article 3.

McCarron, E., & Frydenberg, M. (2023). Digitally Prepared for Success? Technology Skills of Incoming First-Year College Students. 21(3), 70–90.

Mentzer, K., Frydenberg, M., & Patterson, A. (2024). Are Tech Savvy Students Tech Literate? Digital and Data Literacy Skills of First-Year College Students. Information Systems Education Journal, 22(3), 4–24. https://doi.org/10.62273/MFSD3135

Mew, L., & Money, W. (2024). Leveraging ChatGPT for Higher Education Course Offerings. ISEDJ, 22(5), 72.

Murugesan, S., & Cherukuri, A. K. (2023). The Rise of Generative Artificial Intelligence and Its Impact on Education: The Promises and Perils. Computer, 56(5), 116–121. https://doi.org/10.1109/MC.2023.3253292

Ng, D. T. K., Luo, W., Chan, H. M. Y., & Chu, S. K. W. (2022). Using digital story writing as a pedagogy to develop AI literacy among primary students. Computers and Education: Artificial Intelligence, 3, 100054. https://doi.org/10.1016/j.caeai.2022.100054

Nick, T. G., & Campbell, K. M. (2007). Logistic Regression. In W. T. Ambrosius (Ed.), Topics in Biostatistics (pp. 273–301). Humana Press. https://doi.org/10.1007/978-1-59745-530-5_14

Paris, P. G. (2002). Critical Thinking and the Use of the Internet as a Resource. International Education Journal, 4(1), 30–41.

Raman, R., Mandal, S., Das, P., Kaur, T., JP, S., & Nedungadi, P. (2023). University students as early adopters of ChatGPT: Innovation Diffusion Study. https://doi.org/10.21203/rs.3.rs-2734142/v1

Rogers, E. M. (1962). Diffusion of Innovations. Free Press of Glencoe.

Schüller, K. (2022). Data and AI literacy for everyone. Statistical Journal of the IAOS, 38(2), 477–490. https://doi.org/10.3233/SJI-220941

Sharples, M. (2023). Towards social generative AI for education: Theory, practices and ethics. Learning: Research and Practice, 9(2), Article 2. https://doi.org/10.1080/23735082.2023.2261131

Venkatesh, V. (2022). Adoption and use of AI tools: A research agenda grounded in UTAUT. *Annals of Operations Research*, *308*(1), Article 1. https://doi.org/10.1007/s10479-020-03918-9

Wong, W. K. O. (2024). The sudden disruptive rise of generative artificial intelligence? An evaluation of their impact on higher education and the global workplace. *Journal of Open Innovation: Technology, Market, and Complexity*, *10*(2), Article 2. https://doi.org/10.1016/j.joitmc.2024.100278

Zhong, C., & Kim, J. B. (Joo B. (2024). Teaching Case: Teaching Business Students Logistic Regression in R With the Aid of ChatGPT. *Journal of Information Systems Education*, *35*(2), 138–143.

Appendix I. Survey Questions about AI Usage

Which of these statements best describes how <u>you</u> have used ChatGPT: (Select one)

- I've never heard of ChatGPT
- I've heard of ChatGPT but have never tried it myself
- I've tried ChatGPT but have never used it for classwork
- I've used ChatGPT to help answer homework assignments but not quizzes or exams
- I've used ChatGPT to help answer quizzes or exam questions

Which of these statements best describes how others you know have used ChatGPT: (Select one)

- I've never heard of ChatGPT
- I've heard of ChatGPT but don't know anyone else who has tried it
- I know others who have tried ChatGPT but not for classwork
- I know others who have used ChatGPT to help answer homework assignments but not quizzes or exams
- I know others who have used ChatGPT to help answer quizzes or exam questions

Before coming to college, how often have you used ChatGPT or other AI tools to:
(Responses: Never, Little, Somewhat, Frequently, Very Frequently)

1. help you with homework
2. help you with writing essays or papers
3. write a cover letter for a job or college application
4. help you in your job
5. help with personal (not-school or work related) tasks
6. find learning resources

To what extent do you agree with these statements about ChatGPT or other generative AI tools?
(Responses: Strongly Disagree, Somewhat Disagree, Neutral, Agree, Strongly Agree)

1. They are helpful when doing homework
2. They are helpful when writing essays or papers
3. They are helpful when writing letters for job or college applications
4. They are helpful in my job
5. They are helpful for personal tasks
6. They are helpful in finding learning resources

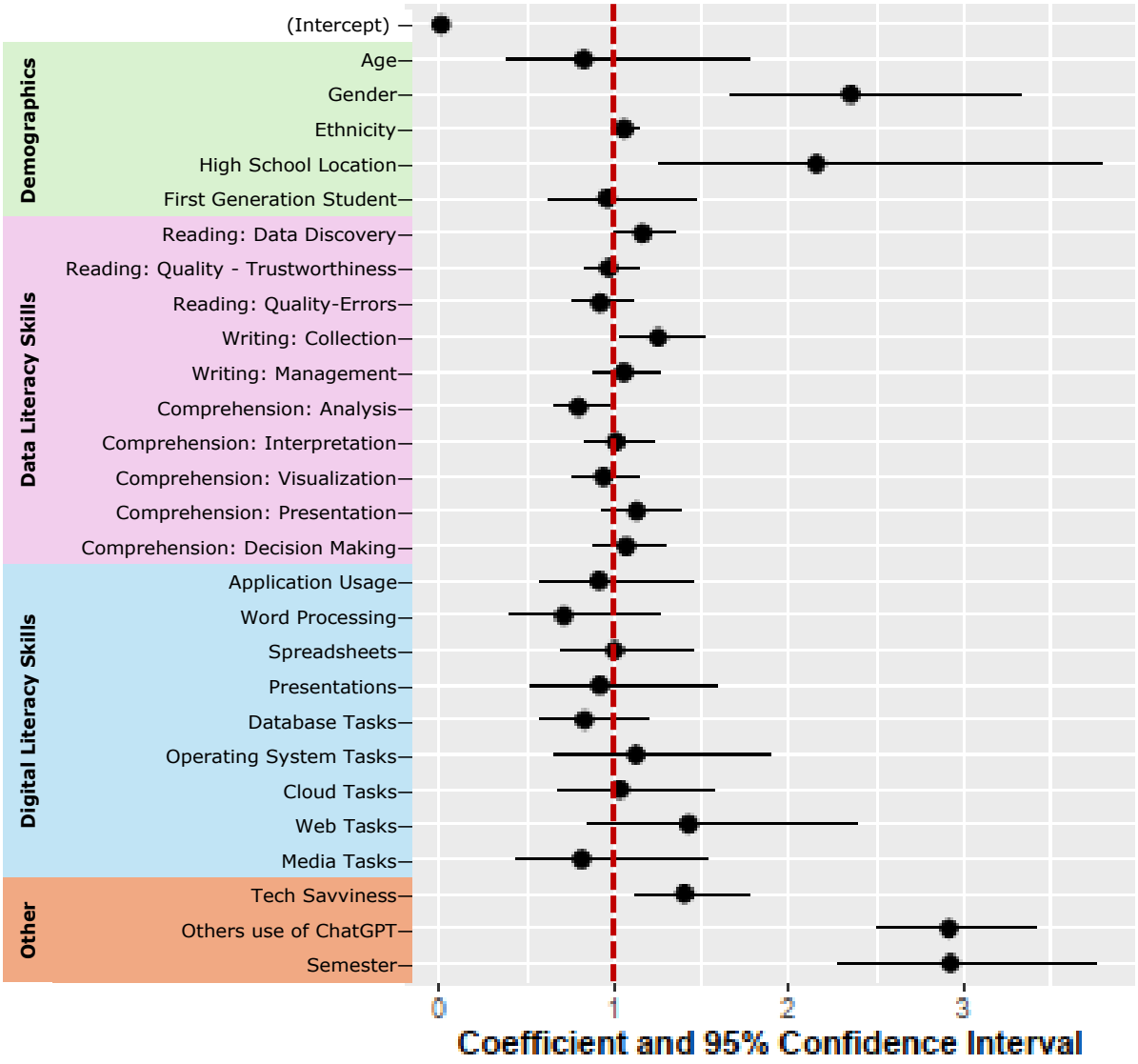
To what extent do you agree with these statements:
(Responses: Strongly Disagree, Somewhat Disagree, Neutral, Agree, Strongly Agree)

1. ChatGPT helps me reach my learning goals
2. It is okay to use ChatGPT to help me with graded class assignments
3. It is okay to use ChatGPT for personal purposes
4. I think I will use ChatGPT a lot to help me with my college courses

Appendix II. Regression Analysis

	Characteristic	Estimate Exp	Std. Error	z value	Pr(>|z|)	Sig
	(Intercept)	0.00602392	1.01129428	-5.0549262	4.31E-07	
Demographics	Age	0.82447206	0.38675851	-0.4990505	0.61774383	
	Gender	2.3541209	0.17715419	4.83289379	1.35E-06	***
	Ethnicity	1.05666723	0.04109728	1.34120385	0.17985428	
	High School Location	2.15767549	0.28155654	2.73135723	0.00630741	***
	First Generation Student	0.95888348	0.22125888	-0.1897583	0.84949851	
Data Literacy Skills	Reading: Data Discovery	1.16063837	0.07840564	1.89999296	0.05743404	**
	Reading: Quality - Trustworthiness	0.9679529	0.08352707	-0.3899557	0.69656933	
	Reading: Quality-Errors	0.91823002	0.09452987	-0.902438	0.36682424	
	Writing: Collection	1.25119137	0.09897494	2.26417094	0.0235636	***
	Writing: Management	1.05443753	0.09076459	0.58401049	0.55921326	
	Comprehension: Analysis	0.79318478	0.10359064	-2.2366795	0.02530729	***
	Comprehension: Interpretation	1.00846432	0.0978591	0.08613095	0.93136232	
	Comprehension: Visualization	0.93876078	0.10477302	-0.6031571	0.54640416	
	Comprehension: Presentation	1.12839831	0.10543814	1.14568792	0.25192431	
	Comprehension: Decision Making	1.06859367	0.10080572	0.65813181	0.51045344	
Digital Literacy Skills	Application Usage	0.91165035	0.23646576	-0.3911719	0.6956702	
	Word Processing	0.71059835	0.29474392	-1.1591347	0.24640128	
	Spreadsheets	1.00099828	0.18608623	0.00536195	0.99572181	
	Presentations	0.91679522	0.2831927	-0.3067563	0.75902887	
	Database Tasks	0.82847411	0.19116859	-0.9843128	0.32496175	
	Operating System Tasks	1.12379166	0.26484177	0.44067209	0.65945041	
	Cloud Tasks	1.03217864	0.21697188	0.14597168	0.88394375	
	Web Tasks	1.4252037	0.26310288	1.34667759	0.17808409	
	Media Tasks	0.81324336	0.3248304	-0.6364087	0.5245101	
Other	Tech Savviness	1.4018558	0.11913182	2.83548857	0.00457557	***
	Others use of ChatGPT	2.91666884	0.07901982	13.5465017	8.31E-42	***
	Semester	2.92557721	0.12661898	8.47812718	2.29E-17	***

Appendix III. Coefficient and 95% Confidence Intervals



Teaching Case

An Introductory IS Course Culminating Project – Weather Easy as Pi

Anthony Serapiglia
Anthony.Serapiglia@stvincent.edu
CIS Department, Saint Vincent College
Latrobe, PA 15650

Abstract

Many introductory courses in CS/IS are modular by nature. This allows for multiple topics to be introduced to beginning students and to serve as a springboard to higher-level material in later courses. Unfortunately, in attempting to fit many topics in, connections between topics or a cohesive theme can be difficult to establish or maintain. The introduction of a final, multifaceted, project that incorporates the individual topics and concepts of the course can serve as the catalyst enabling the whole to gel within the nascent CS/IS student. The purpose of this paper is to present such a culminating project that is flexible in the amount of topics included, and difficulty level - a mini weather station Raspberry Pi project.

Keywords: Information Systems, Internet of Things, IT Project Management, EATPUT

Recommended Citation: Serapiglia, A., (2025). An Introductory IS Course Culminating Project – Weather Easy as Pi. *Information Systems Education Journal* v24, n1 pp 19-29. DOI# https://doi.org/10.62273/GBOX4060

An Introductory IS Course Culminating Project – Weather Easy as Pi

Anthony Serapiglia

1. INTRODUCTION

There have been many approaches to structuring introductory Computer Science (CS) and Information Systems (IS) courses. It can quickly become a "chicken or the egg" question when determining whether it is better to go deep and narrow or wide and shallow. How best to pack all of it in?

Many incoming freshmen or beginning students have only experienced a very narrow set of technologies or specialties in the discipline. As consumers of technology, they have been very proficient and accomplished. Whether for entertainment or within a classroom, students have been able to do what they have been asked to do, and at very high levels. They are Digital Natives after all. However, many have not been asked to create complex systems or go beyond single task/single technology exercises. Getting under the hood or straying from the guided path is not something that many high school aged students have been asked to do.

CS/IS are popular and sought-after degrees. According to the National Center for Education Statistics (NCES), in 2021-22 CS/IS degrees were the fifth highest category with 97,047 degrees issued (NCES, 2023). Many students are motivated by continued job growth and salary growth as projected by the Bureau of Labor Statistics (USBLS). Both job openings and salaries have increased much faster than the average for all occupations and the sustained growth is expected to continue (USBLS, 2023).

The level of preparedness for a CS/IS major varies widely. According to a 2023 study by Code.org, only 8 states have legislation requiring public school students to pass a CS course for graduation, two of which were added in the 2022/23 academic year. The same study reported 14 states where less than half of high schools offer CS instruction. Nationwide, 57.5% of US public schools offer a foundational CS class. This leaves more than 10,000 high schools across the country that do not offer a single CS course. This number is also unevenly spread. Within the lowest 35 states, only 5.8% of students are enrolled in a foundation CS course (Code.org, 2023).

Without a head start in high school, many students arrive at a college already having declared for a CS/IS major with little to no formal introduction to the disciplines. This makes the role of an introductory CS/IS course even more important and valuable in filling in blanks, leveling the playing field, and establishing a solid foundation.

This paper describes the introduction of a capstone project into an introductory CS/IS course. The project incorporates subject areas of System Design, Hardware Integration, Networking, Python Scripting, Operating Systems, IoT, and System Lifecycle. Other topics of Cybersecurity, Privacy, and Data Management can easily be added.

Included in this paper is a background of the course and the evolution towards a capstone project, a description of the materials and environment necessary, an outline and instructions for each stage of project development, commentary and observations from implementation, and options for scaling the project up or down from the framework.

2. BACKGROUND

This project is part of a Computing and Information Systems department at a small private religious liberal arts college. The department supports three degree programs, Bachelor of Science degrees in Computer Science, Cybersecurity, and Information Systems.

Prior to 2012, the first course for all three programs was introductory programming. To more align the degree programs with model curriculum examples, and also in response to student surveys, a new course was added as a "CS-Zero" orientation class. This course was designed to be complimentary to the introductory programming course and to include a multitude of subjects as a stepping stone to more advanced courses. As such, a new introductory course was designed in a very modular fashion to allow for flexibility and evolution of topics as the years progressed.

Feedback from student course evaluations remained positive for the next decade. For many students the broadness of the introductory material was comforting. Some were thankful that a stated goal of the course was to create a level start for everyone, knowing that there was a wide disparity of experience within each incoming freshman class.

Beginning in 2021, feedback from students began to evolve to include several comments from students that the course was starting to become too fragmented. Students noted that the connections between topics were not always visible or emphasized. In a course that emphasizes systems, the greater picture as a whole was being missed.

Reflecting upon the original intent of the introductory course, the primary goal was for the content of the course to be a steppingstone to other, deeper, experiences through the degree program. This was accomplished with introductory material for topics such as Networking and Databases, but a primary area not addressed was project management and preparation for classes such as Software Engineering, Systems Analysis and Design, or the Senior Project capstone.

Introducing a mini-capstone project within the introductory course would serve to provide a platform for cohesively bringing together the topics of the course into one as well as preparing students for the complexity of larger projects later in the degree program.

The EATPUT Information Systems Model

Dr. Anthony Debons was an experimental psychologist and early pioneer in Information Science. Debons worked closely with the Army Air Corps and US Air Force in the years after World War II developing command and control systems. These were heady days of advancements in Information Systems, Decision Support Systems, and ADIK (Advanced Data Information Knowledge) systems. While these specific labels may have gone out of favor, their core simplicity in structure and framework is worth revisiting as models for modern "complex" systems. (Serapiglia, 2022)

Beginning in 1960, Debons led a project to establish a conceptual framework for the design of an information system to support command and control for the Strategic Air Command. This project was a contemporary of the group led by J.C.R. Licklider at DARPA, with Debons and Licklider both having backgrounds in psychology and wide interdisciplinary views of information systems. According to Debons, they conferred on a number of occasions, including consultations on funding devoted to projects focused on the development of better software to train computer programs that would benefit both of them (Asprey, 1999). These efforts in developing command and control systems for the military had great influence on the development of early management information systems and decision support systems that have evolved into the systems of today (Asprey, 1999).

Debons' work culminated in an abstract model of an Information System – EATPUT (Debons, Horne, Cronenweth 1988). Consisting of six basic components, the first letters of which produce the acronym. The six components of EATPUT are:

Event World – The occurrences that are relevant to the objective and functioning of the information system. It includes the classifying and categorizing of events and the representation of them in symbolic form.

Acquisition – The initial physical component of the system, used to capture matter and energy describing an event from the external environment (data).

Transmission – The actual movement of signals (data) within and between components of the system.

Processing – The ordering, storage, and retrieval of data for the ultimate purpose of applying it to problem solving, decision making, or general development (knowledge formulation).

Utilization – The component that represents the evaluative, interpretive requirement of information systems

Transfer – the action component of the system; the implementation of the decider function through the system's transfer medium. The Transfer function in this model can be seen as communication or information transfer.

Internet of Things Six Layer Model

In 1999, Kevin Ashton coined the term Internet of Things (IoT) in describing a system of uniquely identifiable sensors connected to the Internet (Ashton, 2009). Since that time, there have been

many variations of models of Internet of Things (IoT) structure. Some may be as few as three layers (Khajenasiri et al., 2017), and others as large as seven (Padmanabhan, 2022). As an introduction, this exercise builds from a six-layer IoT model that is derived from Dr. John Barrett (Barrett, 2012). Dr. Barrett's original model included five layers. A sixth wrapper layer of security has been added to emphasize the need to incorporate Cybersecurity consciousness in the system lifecycle. The basic layer of Dr. Barrett's original model are:

Identity Layer – The need to uniquely identify each system/sensor. This can be accomplished through existing network identifiers, such as MAC or IPV6 addresses.

Sensing Layer – This may also be termed the perception layer. This layer includes any device that enables the gathering of data from outside of the system.

Communications Layer – This may also be termed the network layer. This layer allows for the movement of data throughout the system and beyond. This layer may utilize multiple technologies, Including Bluetooth, Ethernet, Wi-Fi, and more.

Data Storage Layer – This layer determines how data is handled both in use and at rest. Technologies that may be utilized in this layer include MicroSD cards and High Bandwidth Memory.

Application Layer – This layer can also be termed as a middleware, business, or management layer. As an application layer, this is a multifaceted catch-all for working with collected data and allowing user access.

Security Layer – This layer is not often included in IoT models. It is important to stress to students the need to ingrain cybersecurity concepts into all systems. This layer can include the acknowledgment of encryption for data, access control models, backup/recovery methods, and much more.

3. WEATHER STATION PROJECT

Project goal: This project was designed to engage students in the process of developing an information system to solve a need or problem. Students build a working information system that will gather measurements of the environment and display them on an accessible website within the department network. A successful project results in the ability of the instructor to access a webpage served by the student's system that displays the temperature, humidity, and time.

This project has been delivered across a two-week window, delivered in six class sections on a Monday/Wednesday/Friday fifty-minute format.

Project Stages

1. Project overview, environment analysis, solution identification, project planning
2. Proposal
3. Base OS installation
4. Apache installation and verification
5. Sensor installation, script configuration, and output verification
6. Web page creation and accessibility verification
7. Documentation

The timing of the project was designed such that the initiation and introduction (Stage 1) took place on a Friday. This allowed for students to have the weekend to develop and submit a project proposal that was due the following Monday. This arrangement also allowed for the final project day presentations to take place on the Wednesday of the last week of classes. Friday, the last class of the semester, was able to be reserved for review.

While students were given the opportunity to work outside of the classroom, most students worked within the class time and computer lab environment.

Project Kit

The following seven items were bundled together in resealable gallon-sized bags and provided to students:

Raspberry Pi Model 4B computer board
MicroSD card
DHT11 Temperature Humidity Sensor Module
GPIO pin jumper wire
Micro HDMI to Display Port cable
USB cable for power
USB power block

Working within the department computer lab, students were able to utilize the existing monitor, keyboard, mouse, and network cable at each station.

Based on spring 2025 prices, each kit costs approximately $80 with most components available through Amazon.

Project Overview

Stage1:

At the beginning of the project, students are reminded of the course topics that have preceded this capstone exercise. At the conclusion of the recap, a scenario is developed with class discussion.

For this project, a problem scenario was initiated related to the specific room in which the class is held. Discussion is started based on the Internet of Things sections that precede the final exercise. With a focus on visibility into the environment, students are asked what variables within the room would be most valuable to monitor. The most common responses have been related to occupancy. Students immediately identify the need to know if the room is occupied and if any computers are available for use. Beyond those variables, with some prompting other environmental aspects are identified. This classroom has a set of windows facing west, and at various times during the year can become relatively warm as the sun sets. Several years previously, there was also an emergency situation within this room where a pipe had burst in the ceiling, resulting in some flooding.

As an introductory course, the layout of the system is provided to the students in a general outline: The project will consist of students creating a temperature and humidity reporting system. The system will be comprised of a sensor and a Raspberry Pi computer. The sensor will collect data through a Python script on the Raspberry PI and save the readings to a text file. The Raspberry Pi will also support an Apache web server which will display a web page configured to display the contents of the text file of collected readings and refresh on a determined interval. The Raspberry Pi/Apache server will be connected to the school network such that the page displaying the collected data will be accessible to other computers on the school network.

Given the outline of the system, the first assignment/stage artifact for students is two-fold. First, students are asked to re-state the problem and solution in their own words. This paper should include a description of the environment, resources available, restrictions, and other factors of note (such as traffic of rotating classes throughout the day). The second piece is an inventory of the devices, software, and other resources required. This inventory list should also contain a preliminary budget with links to possible purchase locations and pricing. In re-stating the solution, students are directed to project the proposed system to the EATPUT Information System model, as well as a six-layer IoT system model.

Stage 2:

Stage two is a gateway assignment. Students are required to develop and submit a one-page executive summary proposal to request permission to proceed with the project. The proposal must include a summary of the situation, a description of the system, and a bottom-line estimated cost. Upon submission and review of the proposal, students receive the project kit containing the equipment needed to build the system.

Stage 3:

Once students have a project kit, the first action stage is to establish the Raspberry Pi as a working computer. Initially, as a way of reducing possible failure points, students were provided with a microSD card pre-flashed with the latest version of the Raspberry Pi OS. The OS had not been set up, requiring students to still work through the initial installation info screens. Students are directed to the official Raspberry Pi website for documentation and initial Getting Started instructions. Once at a desktop, students are required to verify connectivity and to update the OS through a terminal command.

The artifact for Stage 3 is a dialog/journal of installation steps that includes a final screenshot of the Raspberry Pi displaying a terminal window showing a successful update of the OS.

Stage 4:

Once the Raspberry Pi is updated, the next stage is the installation of Apache, a web server. Again, students are directed to install Apache through a terminal window. While the process for this stage is short compared to others, it is stressed to students that it is important to separate and isolate different tasks.

The artifact for Stage 4 is a dialog/journal of the installation steps required to install Apache on their Raspberry Pi. Screenshots are required to show the installation success in the terminal window. A screenshot is also required to show the default Apache landing page accessible through the network IP address and the local loopback address.

Stage 5:

Stages 5 and 6 combined may be the most challenging for introductory students. Each requires some programming. Depending on the level of the students, some initial code may be provided and then modified by the students.

The focus of stage 5 is to attach the DHT11 Temperature/Humidity sensor. There are several ways to interact with the DHT11 sensor. The DHT11 sensor is a popular device and has many libraries associated with it for free use. The library utilized by this project is adafruit-circuitpython-dht available through the Adafruit Industries GitHub repository. The DHT11 is also a digital device, so there is no need for an extra analog-digital convertor. The sensor can attach directly to the General Purpose Input Output pins (GPIO) of the Raspberry Pi.

The first part of Stage 5 is connecting the sensor to the Raspberry Pi. The DHT11 sensor has three connectors. These are labeled with "+", "out", and "–" markings.

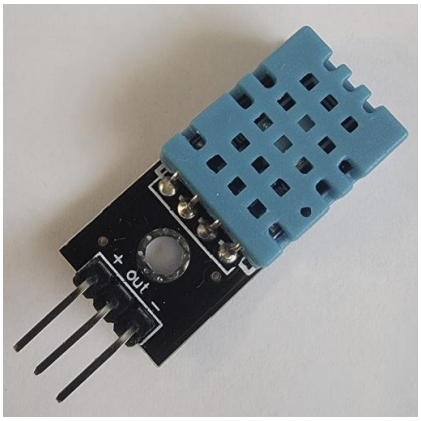


Figure 1: DHT11 Temperature/Humidity Sensor

Connecting to the Raspberry Pi GPIO pins is simple when utilizing female to female jumper cables. The "+" connector will be connected to pin 2 on the Pi. The "out" connector will be connected to pin 11, GPIO17 within the Pi. The "-" pin will be connected to pin 6 on the Pi.

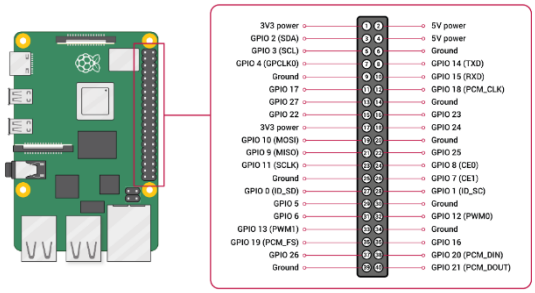


Figure 2: Raspberry Pi GPIO Pin Assignments (learn.sparkfun.com)

Take special note of the orientation and labeling of the pins on the DHT11 and the corresponding pins on the Raspberry Pi. Improper connection, the reversal of the plus and minus, can lead to the destruction of the sensor.

Once connected, Python and the necessary libraries can be installed through the following commands:

```
sudo apt install python3 python3-pip python3-venv
mkdir ~/dht11
cd ~/dht11
python3 -m venv env
source env/bin/activate
python3 -m pip install adafruit-circuitpython-dht
```

A simple way of saving the data is to place the text file in a subfolder of Apache. In a terminal window, browse to the default html folder for Apache (/var/www/html). Create a subfolder for the output file name DHT11. Assign RWX permissions appropriately (… I.E. NOT 777). Open a second terminal window for the following commands:

```
cd /var/www/html
mkdir DHT11
cd DHT11
sudo chmod -R 327 ./
```

Configuring the Python script

The Raspberry Pi OS has a few preinstalled scripting tools to help create and edit code. From the top Pi menu, go to the Programming section and open Thonny. Thonny is an editing environment that will allow you to create or edit Python scripts.

Once Thonny opens, first save the opened blank document as dht11.py in the "dht11" folder created in the home folder during the previous step. Sample code has been provided in Appendix A. Copy the Python code into the editor. Save the file.

Returning to the original terminal window. Verify that it is in the "dht11" folder and the environment is still active.

Run the script:

```
python3 dht11.py
```

The output of the script shows the temp in Celsius, Fahrenheit, and Humidity:

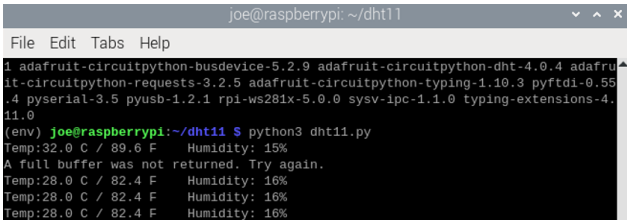


Figure 3: Script Output

Verify that the "output" file exists in the html folder. Move to the second terminal window. Show the directory contents (ls) of the DHT11 folder and verify the output file exists.

The Stage 5 artifact is a dialog/journal of the installation steps required to install Python, configure directories, create/modify the DHT11 Python script, and to run/verify the DHT11 Python script. Screenshots are required to show the script running in a terminal window, validation of the output file creation, and contents written to the output file.

Stage 6:
With a working sensor successfully writing data to an output file, the next step is to configure a web page that will display the contents of the output file in a web browser.

With an introductory class, sample code has been provided. See Appendix B for the sample html code. There are multiple approaches to displaying the contents of a text file from an HTML document, this sample utilized fetch. Depending on the level of the course, students could be asked to develop their own code and pursue other methods.

To utilize the provided sample HTML code, open a terminal window on the Raspberry Pi and browse to the /var/www/html directory. Rename the existing index.html page with the following command:

```
mv index.html index.ori
```

Open the Thonny editor that was utilized in Stage 5. In Thonny, create a new file. Save this file to the default location for Apache by browsing to other locations – computer – var – www – html. Save the new file as index.html. Copy the contents of the index file sample code into Thonny. For personalization, have students add their name as the top displayed line to the index page. Save the file.

Open a web browser and enter 127.0.0.1 as the loopback address. Confirm that the page now appearing is the new index.html page. If the original Apache page is still showing, try a "hard refresh" by hitting the Ctrl-F5 buttons together.

Open a terminal window and find the IP number assigned to the Raspberry Pi utilizing the "ip a" command. Enter the IP number into a web browser to confirm that the temperature and humidity data are displayed. Share the IP number with another student and the instructor to confirm that the index page can be viewed by others on the network.

The stage 6 artifact is a dialog/journal of the steps required to rename the existing index.html page, creation of a new index.html page, the modification/configuration of the index.html page, and the confirmation that the index.html page displaying the temperature/humidity data is accessible on the network. A screenshot of the index.html page accessed through the LAN address with the student's name and current data displayed must be included.

Stage 7:
The final piece of the project is documentation. With a limited time window to complete the project at the end of a semester, the emphasis is placed on "document as you go". Each of the previous stages culminates with artifacts documenting the actions of those stages. The final stage is an exercise in combining those artifacts into a cohesive whole. Introducing organization, note-taking, and documentation concepts in the early stages of a student's journey is just as important as any technical skill. The idea of a twenty-page report on a project will seem very intimidating to first-year students. However, once they add together the number of pages from the combined assignments leading up to the final report, they quickly realize that they are most

likely past that number already.

At the low end of the rubric are reports/documentation that simply package the previous sections together with a table of contents. More developed reports will revisit and revise sections with broader observations, details, and commentary. Fully developed documentation will include all sections, a cohesive narrative of the project, and appropriate reflection with tips for improvement and further development paths.

4. PROJECT REFLECTION

This mini weather station Raspberry Pi project was designed to serve as a culminating capstone exercise in an introductory CS-Zero or Introductory Information Systems course. There are many goals for the project, and they can be managed in a flexible fashion to emphasize sections as needed or desired. The level of difficulty can also be adjusted by managing the amount of prepared code provided to students in different stages.

This project has been delivered in two semesters with twenty-five students during the fall semester and twelve students in the spring. Overall, student response to the project was positive. A short student survey of five questions was conducted at the end of the project each semester. Combined, thirty-one students responded. Each question was ranked on a five-point scale with 1 as low and 5 as high. For each question, the lowest response was a 3.

Question 1 – This project helped me understand the different topics of the course and how they interrelate to each other. Responses to question one averaged 4.52.

Question 2 – This project helped me understand Information Systems in general. Responses to question two averaged 4.55.

Question 3 – This project helped me understand the stages of IT project management. Responses to question two averaged 4.42.

Question 4 – After completing this project, I feel more confident about how different technologies interact with each other. Responses to question two averaged 4.52.

Question 5 – Overall, this project increased my interest in Information Systems. Responses to question two averaged 4.58.

During the project, students expressed that they enjoyed being able to physically work with the Raspberry Pi and the sensor. While the hardware interactions are very simple, many students have not had the experience of piecing a system together. Several students also expressed their gratitude for how the project was broken into stages. They had originally felt overwhelmed by the description of the project, but that anxiety diminished as they saw how the work was broken out into manageable pieces. At the end of the project, several students began discussing ideas on how to expand their system, expressing interest in incorporating authentication methods and cloud hosting.

4. REFERENCES

Ashton, K. (22 June 2009). "That 'Internet of Things' Thing". rfidjournal.com. Retrieved 9 May 2024. https://www.rfidjournal.com/expert-views/that-internet-of-things-thing/73881/

Aspray, W. (1999). Command and control, documentation, and library science: The origins of information science at the University of Pittsburgh. *IEEE Annals of the History of Computing*, *21*(4), 4–20. https://doi.org/10.1109/85.801528

Barrett, J. The Internet of Things [Video]. Ted Conferences, TEDxCIT. September 6, 2012. Cork, Ireland. Retrieved fro: https://www.youtube.com/watch?v=QaTIt1C5R-M

*Code.org Advocacy Coalition. (2023). Code.org. https://advocacy.code.org/stateofcs/*Retrieved June 14, 2024, from https://advocacy.code.org/stateofcs.

Debons, A., Horne, E., & Cronenweth, S. (1988). *Information science an integrated view*. G.K. Hall.

Introduction to the Raspberry Pi GPIO and Physical Computing - learn.sparkfun.com. (n.d.). Learn.sparkfun.com. https://learn.sparkfun.com/tutorials/introduction-to-the-raspberry-pi-gpio-and-physical-computing/gpio-pins-overview

Khajenasiri, I., Estebsari, A., Verhelst, M., & Gielen, G. (2017). A review on internet of things solutions for intelligent energy control in buildings for Smart City Applications. *Energy Procedia*, *111*, 770–779. https://doi.org/10.1016/j.egypro.2017.03.239

National Center for Education Statistics. (2023). Undergraduate Degree Fields. *Condition of Education.* U.S. Department of Education, Institute of Education Sciences. Retrieved May 30, 2024, from https://nces.ed.gov/programs/coe/indicator/cta.

Padmanabhan, A. (2022, February 15). *IOT security model*. Devopedia. Retrieved March 2, 2024, from https://devopedia.org/iot-security-model

Serapiglia, A., (2023). Command and Control – Revisiting EATPUT as an IS Model for Understanding SIEM Complexity. *Cybersecurity Pedagogy and Practice Journal*2(1) pp 77-84. http://CPPJ.org/2023-1/ ISSN : 2832-1006. A preliminary version appears in The Proceedings of EDSIGCON 2022

The NCES Fast Facts Tool provides quick answers to many education questions (National Center for Education Statistics). National Center for Education Statistics (NCES). Retrieved June 14, 2024, from https://nces.ed.gov/fastfacts/display.asp?id=37

U.S. Bureau of Labor Statistics. (2023, September 6). *Field of degree: Computer and Information Technology*. U.S. Bureau of Labor Statistics. Retrieved June 14, 2024, from https://www.bls.gov/ooh/field-of-degree/computer-and-information/computer-and-information-technology-field-of-degree.htm

APPENDIX A
Python Code

The following is Python code that will collect data from a DHT11 temperature/humidity sensor. It will display data to the screen with the temperature in Celsius and Fahrenheit, with humidity. This script will also write the data to an output file "output" located in /var/www/html/DHT11/. The data written to the output file also contains a time stamp for each line written.

```
import time
import adafruit_dht
import board

dht_device = adafruit_dht.DHT11(board.D17)

while True:
    try:
        temperature_c = dht_device.temperature
        temperature_f = temperature_c * (9 / 5) + 32

        humidity = dht_device.humidity
        currentTime = time.ctime()

        print("Temp:{:.1f} C / {:.1f} F    Humidity: {}%".format(temperature_c, temperature_f, humidity))


        if humidity is not None:
            with open('/var/www/html/DHT11/output','a') as w:
                w.write('Temp={0:0.1f}*  Humidity={1:0.1f}%  '.format(temperature_c, temperature_f, humidity))
                w.write(f"{currentTime}")
                w.write('\n')
        else:
            print('Failed to get reading. Try again!')

    except RuntimeError as err:
        print(err.args[0])


    time.sleep(10.0)
```

APPENDIX B
HTML code

The following code for a basic HTML page that will display the contents of an output file containing data. It assumes an output file exists in a subdirectory named "DHT11".

```
<html>
<head>
   <title>Temp in W214</title>
   <meta http-equiv="refresh" content="60">
</head>
<body>
   <h1>Contents of "output" file:</h1>
   <pre id="fileContents"></pre>
   <script>
      // Function to fetch file contents using Fetch API
      function fetchFileContents() {
         fetch('./DHT11/output')
            .then(response => {
               if (!response.ok) {
                  throw new Error('Network response was not ok');
               }
               return response.text();
            })
            .then(data => {
               document.getElementById('fileContents').textContent = data;
            })
            .catch(error => {
               console.error('Error fetching file:', error);
               document.getElementById('fileContents').textContent = 'Error fetching file. Check console
for details.';
            });
      }

      // Call the function when the page loads
      window.onload = fetchFileContents;
   </script>
</body>
</html>
```

The Importance of Understanding Global Data Governance in an Interdisciplinary Research Methods Course

Loreen M. Powell
lpowell@maryu.marywood.edu
Marywood University
Scranton, PA USA

Michalina Hendon
michalina.hendon@ucumberlands.edu
University of Cumberlands
Williamsburg, KY USA

ABSTRACT

This study investigates the extent of global data governance content in academic "Research Methods" textbooks, crucial for interdisciplinary courses involving national and international students. A content analysis was conducted using the O'Reilly Media, Inc. website, known for its comprehensive IT resources. Over the duration of one month, research methods textbooks published in the last two years were analyzed for terms related to international or global research and data governance. The initial search yielded 3,848 results, filtered to 544 books relevant to the "data" sub-topic. Among these, only 47 were valid research methods textbooks. Of these, just three included data governance content. None provided detailed country-specific legislation. This research highlights the significant gap in global data governance coverage in research methods textbooks, indicating a need for more comprehensive resources to support interdisciplinary and international research training.

Keywords: data governance, global data legislation, research methods, interdisciplinary approach

Recommended Citation: Powell, L.M., Hendon, M., (2025). The Importance of Understanding Global Data Governance in an Interdisciplinary Research Methods Course. *Information Systems Education Journal,* v24,n1, pp 30-41. DOI# https://doi.org/10.62273/BPN6886

The Importance of Understanding Global Data Governance in an Interdisciplinary Research Methods Course

Loreen M. Powell and Michalina Hendon

1. INTRODUCTION

Due to the declining enrollment numbers in higher education, many academic institutions are choosing to consolidate or merge sections of courses with similar content. Brint and Clotfelter (2016) note that interdisciplinary courses are being introduced to improve undergraduate retention and graduation rates while addressing declining enrollment. The main goal of this strategy is to cut costs. For example, a Master of Business Administration (MBA) research method course is combined with a Master of Management Information Technology (MSIT) or even a Master of Science in Healthcare. Although this interdisciplinary approach to teaching is cost-effective, challenges may arise when addressing trending topics such as data governance within the program to provide the understanding in balancing data utility and security while considering both technical and ethical implications for student researchers.

As business, healthcare, and technology environments continue to evolve globally, the use of electronic data and data sharing are becoming standard practice (Frieden, 2021; Romo, 2022). Additionally, the growing reliance on cloud computing for data storage and sharing are further emphasizing the importance of global data governance among researchers (Brous et al., 2026; Ngesimani et al., 2022). Yet, there is still a limited number of research studies on global data governance (Bližnák et al., 2024; Brous et al., 2016; Marucci et al, 2023). Currently, many data governance standards and documents are in the implementation phase and are sector-specific (Marcucci et al, 2023; Bližnák et al., 2024) or lack global framework principles (Brous et al, 2016).

Due to the massive amount of data available (Benfeldt Nielsen, 2017) along with the normality and customary procedures of data sharing in global business and healthcare (Frieden, 2021; Romo, 2022), educational institutions should emphasize the importance of global data governance in an interdisciplinary research method course encompassing international students. However, the problem is that awareness data governance has typically been covered in the CITI exam, which, depending on the module that is selected within the sponsoring Institutional plan or subscriptions, may or may not address global data governance. Regardless, the focus on data governance typically focuses on the United States (U.S.) legislation regarding data governance. As a result, many international students return to their home country and may not be aware of their country's data governance legislation within their discipline.

Furthermore, while many papers and documents have been published in ethical research practices, many lack knowledge regarding global governance and the differences between multiple global sectors. This research aims to better understand the degree of information centered around global data governance, which can be used for an interdisciplinary graduate research methods course with national and international students. Specifically, the contents of graduate academic "Research Methods" textbooks were examined via a content analysis for global data governance content. This research has practical implications for graduate programs, faculty, and future publishers and authors. The remainder of this paper is structured as follows: review of literature, goal, methodology, results, recommendations, and conclusion.

2.REVIEW OF LITERATURE

The definition of data governance varies across organizations and sectors. Data governance, within the biomedical field, is characterized by stringent regulations and ethical considerations due to the sensitive nature of patient information. Thus, data governance in the biomedical field extends beyond privacy and security to encompass ethical considerations and regulatory compliance in data usage and sharing (Bortoli, 2024). As a result, the importance of informed consent and maintaining patient confidentiality when accessing and utilizing health data for research purposes is essential due to data governance standards such as the Health Insurance Portability and Accountability Act (HIPAA) in the U.S. and/or General Data Protection Regulation (GDPR) in the European Union. The ethical data standards aid in ensuring patient data is used responsibly via anonymization or de-identification when collected, stored, processed, shared, and secured (Hassani et al. 2023; Zmed Solutions, 2024).

Moreover, data sharing involving consortium or collaborative efforts among healthcare institutions, research organizations, and regulatory bodies is commonplace within the biomedical field (Bernier et al., 2022). As a result, it is also important to note that while sharing anonymized data can accelerate medical research and improve patient outcomes, stringent protocols, and agreements are in place to safeguard against data breaches and unauthorized access. Therefore, data governance in the biomedical field emphasizes data quality and accuracy to support clinical decision-making and research outcomes. This involves protocols for data validation, integrity checks, and ensuring that data used in research studies are reliable and reproducible.

A research study by Brous et al. (2016) conducted a systematic review of data governance publications between 2000 and 2015. They found a limited amount of research published within this area. Specifically, they identified 17 publications within databases and 1710 within google scholar. Among their findings, they explain that there is not a single standard or approach that works for all organizations and governments.

Additionally, the World Development Report 2021 by the World Bank (2021) agrees that there is a heightened need for global data governance standards as there is a lack of safeguards, security, and standards for most of the world (Marcucci et al., 2023; Mungai et al., 2022). The report supports the argument for global data governance across sectors and types of organizations.

A more recent study by Marcucci et al. (2023) identified 100 data governance documents across 37 organizations, eight national or local governments, and four regional entities. They examined 58 of the identified documents in detail and found that only 63% of documents were at the beginning of their development and served more as starting points rather than compliance standards. Additionally, they found that only 39% of the documents examined included global data governance or international data and/or human rights protection standards.

Similarly, Bližnák et al. (2024) conducted an extension review of published data governance literature from 2017 to 2023. Among the many findings reported was the overwhelming non empirical research methodologies within the published works. Thus, the majority of recent data governance publications are literature reviews and/or conceptual based. Most importantly, they also concluded by suggesting future research can be explored in sector specific characteristics of data governance such as higher education and healthcare.

There is a need for a more holistic understanding of the difference in data governance and data governance frameworks amongst various sectors, such as the global, public, and private business, technology, and healthcare, when conducting research (Bližnák et al., 2024). Oftentimes, within a research method course, data governance is addressed when a student researcher completes the ethical research practices provided by the Collaborative Institutional Training Initiative (CITI) program. The CITI program provides training in various aspects of research, such as ethics, compliance, and regulations. The content varies depending on the specific module in which the researcher enrolled. Some but not all modules cover topics related to data governance, privacy, and security, which may include international standards and legislation (Collaborative Institutional Training Initiative, 2020). For example, CITI's "Information Privacy and Security" courses often discuss key regulations like the General Data Protection Regulation (GDPR) in the European Union and the Health Insurance Portability and Accountability Act (HIPAA) in the U.S. The privacy and security modules are aimed at educating researchers on how to handle personal data responsibly and in compliance with relevant laws. However, the coverage of global data governance standards and laws is limited in depth and scope.

The adoption of specialized modules which are supported by institutions are requested or should be considered to ensure that researchers are specialized in their research. The modules serve a dual purpose to provide educational guidance on ethical practices and procedures as well as equip researchers with the critical faculties to scrutinize their study’s scope and assess the applicability of pertinent regulations. The learning module approach can facilitate a thorough preparation for the human subject ethical review processes (Collaborative Institutional Training Initiative, 2020). However, it is noteworthy that some institutions do not mandate ethical reviews for archival data (The National University, 2025). In such instances, the responsibility of explaining data usage falls upon the researcher. This responsibility necessitates a thorough explication of the data utilization methodology, ensuring adherence to ethical standards and regulatory compliance despite the absence of a formal institutional review process.

Global Data Governances for Usage and Sharing

Global data governance refers to the systematic and transparent management of data worldwide, aiming to ensure responsible data use, protection, and coordination across various stakeholders, organizations, and governments (Kuzio et al., 2022). The concept has evolved from private organizational governance to include governmental and institutional bodies.

Various data governance standards and regulations apply to specific sectors, but these are often overlooked when working across multiple disciplines. For example, the application of data governance for a technology organization may differ from that of a bio-medical organization.

Additionally, there are also differences within policies in data governance for biomedical organizations and private organizations, which differ significantly due to their distinct objectives, regulatory environments, and data types. For example, within the IT business field, as well as the IT department, data governance focuses on maximizing the value of data assets to drive organizational performance and innovation. As a result, data governance frameworks place emphasis on ensuring data transparency, accessibility, and usability across different business units (Abraham et al., 2020). They emphasize leveraging data as a strategic asset to enhance decision-making processes, improve operational efficiency, and gain a competitive advantage in the marketplace. Therefore, data governance in the IT business field often revolves around establishing clear policies and procedures for data management, including data stewardship roles, data lifecycle management, and compliance with industry regulations such as GDPR or California Consumer Privacy Act (CCPA). For instance, while GDPR and CCPA regulations provide legal frameworks for data compliance, organizations often struggle to balance the demands of these regulations with their need to innovate and maintain competitiveness in the marketplace. The trade-offs between maximizing data utility and ensuring data security highlight the need for a more holistic approach to data governance encompassing technical and ethical considerations (Giordani, 2021).

Furthermore, many organizations may share aggregated or anonymized data with global third-party vendors, business partners, or customers to enhance product development, customer insights, and marketing strategies (Brewis et al., 2023). Thus, data governance in the IT business field emphasizes data global integration, interoperability, and the use of advanced analytics to derive insights that drive business innovation and growth. As a result, global data governance measures, including data masking and encryption, are implemented to protect sensitive information, including PHI information, and mitigate risks associated with data breaches (Winter & Davidson, 2019).

Bernier et al. (2022) asserted, "Rapid legal change, unsettled ambiguities in the law, or conflicting obligations arising in different jurisdictions can exhaust the limited legal compliance resources of biomedical consortia or deter the secondary use of research data for fear of legal non-compliance." (p.2) To evaluate the difference in applicability or the paradigm of the differences and similarities in overarching regulations as assessed by the literature begins with the establishment of data privacy laws and the application of country and local regulations.

Although the legislative privacy laws have been enhanced and have been vastly added to consider the changing data landscape. The first approach to sector-specific data privacy laws includes laws such as the HIPAA, which governs covered entities in the health sector, and the Federal Trade Commission (FTC) Privacy of Consumer Financial Information Rule (Bernier et al., 2022). Following the US, the Council of Europe, OECD, and the GDPR were enacted in other European countries to ensure the regulation and data governance to protect personal and proprietary data. The GDPR follows the Data Governance Act (DGA), the act established in 2022; however, it was enacted and applicable in 2023. The privacy and security legislation of data is part of the European Union's efforts to create a more integrated data economy and enhance trust in data sharing across the EU.

The privacy and security of data legislation establishes rules and guidelines for reusing protected data held by public sector bodies, including personal and commercially confidential information. The use of data requires established consent; confidential or proprietary information can only be disclosed for reuse with appropriate consent or permission. Additionally, exclusive data reuse agreements must be put in place, accommodating reasonable fees (European Commission, 2024). The United States has comparable agreements for data sharing and utilization. Furthermore, the European Data Innovation Board (EDIB) has been created to oversee regulation and ensure compliance.

It is important to note that there are many different data privacy legislations around the world, and these may vary by country. However, as Kuzio et al. (2022) noted, there is a lack of a

comprehensive global structure for safeguarding consumer data, which leads to a situation that is both incomplete and fragmented. The insufficiency of data protection mechanisms for cross-border data transfers is evident, with many regulations currently in development of a common framework is lacking due to the need for global consensus and fragmented policies, inability to resolve overlapping jurisdictional claims, and normative differences among actors by articulating distinct visions of data governance. Furthermore, the emergence of machine learning (ML) and large learning model (LLM) advancement has emphasized the necessity for global language data governance, proposing a multi-party international governance structure focused on language data management as well as the data lifecycle, encompassing data handled by information resellers. Additionally, there is a notable shortage of safeguards for cross-border and international data transfers, potentially resulting in the adoption of the least stringent protection standards as the norm. Lastly, emerging data collection domains possess a worldwide reach but lack adequate regulatory measures at either the domestic or international level (Marcucci et al. 2023).

In addition to these challenges, the global nature of data governance brings forward complex considerations surrounding ethical data use and equitable access. Different regions prioritize various aspects of data protection based on cultural, economic, and political factors, resulting in an uneven global data governance landscape. This disparity raises concerns about the potential for "data colonialism," where more powerful nations or corporations exploit data from less regulated regions for their gain, without adequately addressing local privacy or ethical concerns (Couldry & Mejias, 2019). Therefore, the importance of making students aware of international data governance, policy, and guidelines becomes increasingly important as future employers may be borderless, international corporations.

3. GOAL & RESEARCH QUESTION

The goal of this research is to better understand the degree of information centered around global data governance, which can be used for interdisciplinary research methods courses with national and international students. Specifically, the contents of academic "Research Methods" textbooks were examined via a content analysis for global data governance content. The following research question was explored in this paper.

1. To what extent do graduate-level research methods textbooks incorporate content on global or international data governance, as measured by the percentage of inclusion texts?

4. METHOD

The O'Reilly Media, Inc. website (https://www.oreilly.com) was used to evaluate academic text focusing on research methods. Since data governance is currently a trending IT topic, the O'Reilly website was selected due to its status as a renowned knowledge-sharing platform for IT professionals and academic professors. This free resource offers a comprehensive range of online training materials, encompassing books, videos, interactive events, and structured learning paths. (O'Reilly Media Inc., 2024; Powell et al., 2021). Additionally, the O'Reilly Media, Inc. website was utilized as a valid tool for content analysis research focused on IT textbooks containing grid computing and volunteer computing content (Powell, et al., 2021).

Over a one-month period, from June 8, 2024, to July 8, 2024, a content analysis was conducted on research methods texts found on the O'Reilly website. A content analysis represents a rigorous methodological approach for systematically evaluating textual, visual, and multimedia content. The content analysis research design has been used in many research studies for over 50 years (Elo & Kynqas, 2008).

In this study, we utilized the O'Reilly website to perform a content analysis of research methods textbooks. The website's search functionality allowed us to refine our data collection by applying specific filters to ensure the accuracy and relevance of our results. The researchers utilized the search page (https://www.oreilly.com/search/?q=*&rows=100) and restricted our query. Specifically, under the "publication date" category, we selected publications "within the last two years". This was selected as we wanted to find current research method texts with relevant data privacy content. Under the "format" category, we selected "books". The researchers selected "books" since we were not interested in videos, or any other content. Under the "topic" category, we selected "data". Once these filters were selected, we type in "research methods" into the search box and pressed enter.

As the results appeared, the data was copied and pasted in Microsoft Excel. Once in Microsoft Excel the find function was used to perform an initial content analysis on the title of the text was conducted to ensure that valid research methods

texts appeared. We excluded texts if they did not have "Research Methods" in their title. For instance, works like "Ensemble Methods for Machine Learning," despite appearing in the search results, were deemed irrelevant and excluded based on this criterion. A PRISMA flow diagram is presented in Figure 1 within Appendix A.

Once a list of valid research methods texts was identified, we manually reviewed the electronic versions of the table of contents for each text. Specifically, we used the find function within the Firefox browser to examine the table of contents for any of the following three terms: international research, global research, and data governance. Electronic texts which contained these terms were further manually examined for the extent of the specific content upon which they contained. We manually examined the content for country-specific legislation or, at the very least, the GDPR Data Protection Act. Based upon our findings, we reported the raw data. No statistical analysis was conducted on the data.

5. RESULTS

The initial search initially yielded 3,848 search results for "research methods" books. The search was further narrowed down to 544 books via the following filters: Research methods published within the last two years and the sub-topic "data" Upon further examination among the results, 544 books that were found, only 47 were identified as a valid research method textbook.

Among the evaluation of the 47 identified books, only three of the texts contained data governance language or legislation information. Figure 2 provides a visual representation of the visual percentage of textbooks with some data governance coverage.

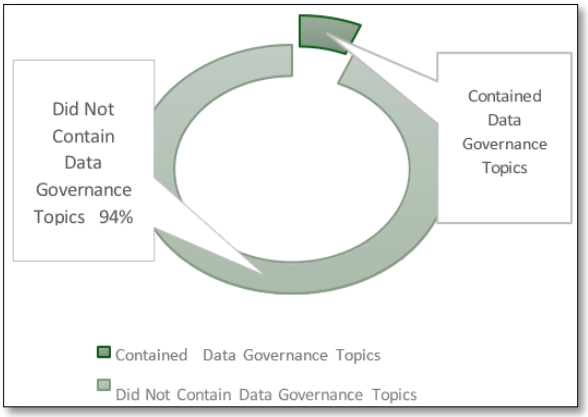


Figure 2. Percentage of texts with data governance topics

It is important to note that upon examination of the specific text chapter among each of the three identified textbooks, none of the texts provided a list of country-specific legislations regarding data governance.

6. DISCUSSION AND RECCOMENDATIONS

Based upon the results showing that research methods textbooks do not cover global or international data governance information, we argue that this may be because this is a fairly new topic. Many laws and legislations are in progress or have not been implemented. Furthermore, many research methods textbooks address the topic by stating the importance of the CITI program, which does not thoroughly address data governance from global data governance perspectives, regulations, and standards. Thus, students are inadequately prepared or undereducated regarding global data usage, privacy, and sharing practices across different countries. This lack of preparation can hinder the student's ability to collect data or conduct research in international contexts due to an insufficient understanding of diverse data governance laws. Consequently, the gap in knowledge can lead to significant legal and financial repercussions for their employers and organizations, including substantial fines and compliance issues when data protection is not internally governed. The absence of comprehensive global data governance training undermines students' capability to navigate complex international regulations, thus impacting their effectiveness in global research initiatives and business operations. Therefore, it is imperative to integrate international data governance education into academic curricula to better equip students for the challenges of global data management and ensure compliance with international standards.

Moreover, since there are limited resources for educators on global data governance, we recommend supplementing the texts with information explaining established data privacy legislation enacted by countries. We adapted Marcucci et al.'s (2023) work, which identified 100 different legislation documents and provided a simplified list of a representative sample of significant data privacy legislation from various countries, providing educators and student researchers a starting point for understanding the global landscape of data governance. Each of these laws reflects similarities as well as different approaches to data protection, influenced by cultural, political, and economic factors specific to their respective jurisdictions. Focusing on data

privacy laws, educators can provide student researchers with a comprehensive overview of global data privacy trends, regulatory frameworks, and the varying approaches to data protection across different countries. Understanding, as well as applicable global knowledge, is crucial for developing a complicated understanding of the intricacies involved in global data governance and the challenges faced by organizations operating in multiple jurisdictions.

Table 1, listed in Appendix B, presents a selection of data privacy laws currently in place across different countries. Although not comprehensive, it provides a foundation for understanding international research policy requirements, regulations, and guidance. This resource highlights gaps in existing textbook coverage of data governance. Consequently, we recommend incorporating additional scholarly resources that focus on country-specific data governance legislation.

7. CONCLUSION

As higher education institutions continue to face declining enrollments, they will continue to look for creative ways to cut costs. As a result, academic institutions may merge various graduate subject areas, such as business administration, healthcare, and IT. One logical area to place an interdisciplinary graduate course is within the research methods subject area. However, these challenges in preparing students, particularly international students, for the complexities of teaching global data governance. While the Collaborative Institutional Training Initiative (CITI) offers foundational guidance, it remains focused primarily on U.S. regulations, leaving students underprepared to handle international data governance issues.

This research study examined possible textbooks that addressed global data governance in interdisciplinary graduate research methods courses. The study revealed a critical gap in global data governance knowledge sharing as well as importance in international data governance provided within research methods texts. We argue that education about global data governance is particularly relevant and needed given the rapidly evolving global business, healthcare, and technology landscapes. As demonstrated through this research, the academic content of research methods textbooks, as currently available, inadequately covers global data governance topics. This learning gap can leave students, especially international ones, unprepared to navigate data privacy laws and governance standards in their home countries or in a global context. Our findings align with Marcucci et al. (2023) and Kuzio et al. (2022) research, which also emphasizes the fragmented and inconsistent nature of data governance frameworks, as well as the challenges posed by emerging technologies and international data exchanges.

Furthermore, this study underscores the need for more comprehensive and globalized approaches to data governance education, particularly for students in disciplines that increasingly require interdisciplinary research and international collaboration. Integrating global data governance content into research methods courses, academic institutions can better prepare students for the complexities of data governance in a globalized world. Furthermore, educators are encouraged to supplement existing materials with up-to-date legislative documents from various countries, thereby providing a broader, more informed perspective on the topic. Addressing these gaps is essential for equipping future researchers with the necessary tools to navigate data privacy, security, and governance challenges in an increasingly interconnected world. This paper provides a foundation of country-specific legislation and emerging trends in data governance. By doing so, educational faculty, programs, and institutions can better prepare students, future researchers, and professionals for the complexities of data governance in an increasingly interconnected world.

This research is not without limitations as it is limited to 100+ publishers found on the O'Reilly site. There may be texts which were not updated or included on the site or have not yet been released. Second, the recommendations to integrate global data governance information stem from the current enrollment crisis push toward interdisciplinary courses like research methods. Third, this research focuses on the fact that international students may be within the course or students wishing to conduct research overseas. Fourth, this research assumes that professors use texts in a research methods course. Firth, no statistical analysis was performed. Finally, this research only focuses on using global data governance. Thus, this research does not imply that understanding global data governance is the key to understanding international research methods. Future research should address the limitations described and reevaluate the text content as there is a need for global data governance materials in the future.

8. REFERENCES

Abraham, R., Schneider, J., & vom Brocke, J. (2019). Data governance: A conceptual

framework, structured review, and research agenda. International *Journal of Information Management,* 49, 424–438. https://doi.org/ 10.1016/j.ijinfomgt.2019.07.008

Bernier, A., Molnár-Gábor, F., & Knoppers, B. M. (2022). The international data governance landscape. *Journal of Law and the Biosciences,* 9(1), lsac005. https://doi.org/10.1093/ jlb/lsac 005

Benfeldt Nielsen, O. (2017). A comprehensive review of data governance literature. https://aisel.aisnet.org/iris2017/3/

Bližnák, K. Munk, M. &. Pilková, A. (2024). A systematic review of recent literature on data governance (2017–2023), in *IEEE Access*, 12 (1), 149875-149888, https://doi.org/ 10.1109 /ACCESS. 2024.3476373.

Bortoli E (2024). Navigating the terrain of health data governance: Ensuring privacy, security, and ethical use. *EJBI*., 20(2):236-237. https://doi.org/10.24105/ejbi.2024.20.2.236-237

Brewis, C., Dibb, S., & Meadows, M. (2023). Leveraging big data for strategic marketing: A dynamic capabilities model for incumbent firms. *Technological Forecasting and Social Change*, 190, 122402. https://doi.org/ 10.1016/j.techfore.2023.122402

Brint, S. & Clotfelter, C. T. (2016). U.S. higher education effectiveness. *RSF: The Russell Sage Foundation Journal of the Social Sciences*, *2*(1), 2–37. https://doi.org/10.7758/rsf.2016.2.1.01

Collaborative Institutional TrainingInitiative (CITI). (2020). About CITI Program courses. https://www.citiprogram.org/about-citi-program/citi-program-courses/

Couldry, N., & Mejias, U. A. (2019). Data Colonialism: Rethinking Big Data's Relation to the Contemporary Subject. *Television & New Media.*, *20*(4), 336–349. https://doi.org/10.1177/1527476418796632

Elo, S. & Kynqas, H. (2008). The qualitative content analysis process. *Leading Global Nursing Research,* 62(1). https://doi.org/10. 1111/j.1365-2648.2007. 04569.x

European Commission. (2024). Data governance act explained, shaping Europe's digital future. data governance act explained. https://digital-strategy.ec.europa.eu/en/ policies/data-governance-act-explained

Frieden, J. (2021, April 23). Health data sharing improved during pandemic, but barriers remain. *Medpage Today*. https://www.med pagetoday.com/practicemanagement/informationtechnology/92263

Giordani, J. (2021). How to balance trade-offs between security and privacy in IT, *Forbes. https://www.forbes.com/councils/forbestechcouncil/2021/06/03/how-to-balance-trade-offs-between-security-and-privacy-in-it/*

Hassani, H., & MacFeely, S. (2023). Driving excellence in official statistics: unleashing the potential of comprehensive digital data governance. *Big Data and Cognitive Computing*, *7*(3), 134. https://doi.org/10.3390/bdcc7030134

Kuzio, J., Ahmadi, M., Kim, K.-C., Migaud, M. R., Wang, Y.-F., & Bullock, J. (2022a). Building better global data governance. *Data & Policy*, 4(25). https://doi.org/10.1017/dap.2022.17

Marcucci, S., Alarcón, N.G., Verhulst, S. G., Wüllhorst, E. (2023). Informing the global data future: Benchmarking data governance frameworks. *Data & Policy,* 5(30). https://doi.org/10.1017/dap.2023.24

Nadal, S., Jovanovic, P., Bilalli, B., & Romero, O. (2022). Operationalizing and automating data governance. *Journal of Big Data*, 9(1), 117. https://doi.org/10.1186/s40537-022-00673-5

Ngesimani, N.L., Ruhode, E. & Harpur, P-A., (2022). Data governance in healthcare information systems: A systematic literature review. *South African Journal of Information Management.* 24(1), 1475. https://doi.org/ 10.4102/sajim. v24i1.1475

O'Reilly Media Inc. (2024). https:// www.oreilly. com/

Powell, L. & Karpinski, A. (2021). Awareness of volunteer computing on the grid: Content within IT texts. *Issues in Information Systems*. 22(1), 246-254 https://doi.org/ 10.48009/1_iis_2021_246-254

Romo, J. (2021, March). Data Sharing during the COVID-19 Pandemic. https://blogs.ischool. berkeley.edu/w231/2022/03/16/data-sharing-during-the-covid-19- pandemic/

The National University. (2025). Institutional Review Board (IRB): Archival/Secondary Data. https://resources.nu.edu/irb/ archival-data

The World Bank (2021). World Development Report 2021: Data for Better Lives—Research Directions and Measuring Country Capacity. https://www.worldbank.org /en/news/video/2021/05/25/data-for-better-lives-research-directions-and-measuring-country-capacity

Winter, J. S., & Davidson, E. (2019). Big data governance of personal health information and challenges to contextual integrity. *The Information Society*, *35*(1), 36–51. https://doi.org/10.1080/01972243.2018.1542648

Zmed Solutions (2024). Why is It Important to Maintain Patient Confidentiality in Healthcare? https://www.zmedsolutions.net/why-is-it-important-to-maintain-patient-confidentiality-in-healthcare.

Appendix A

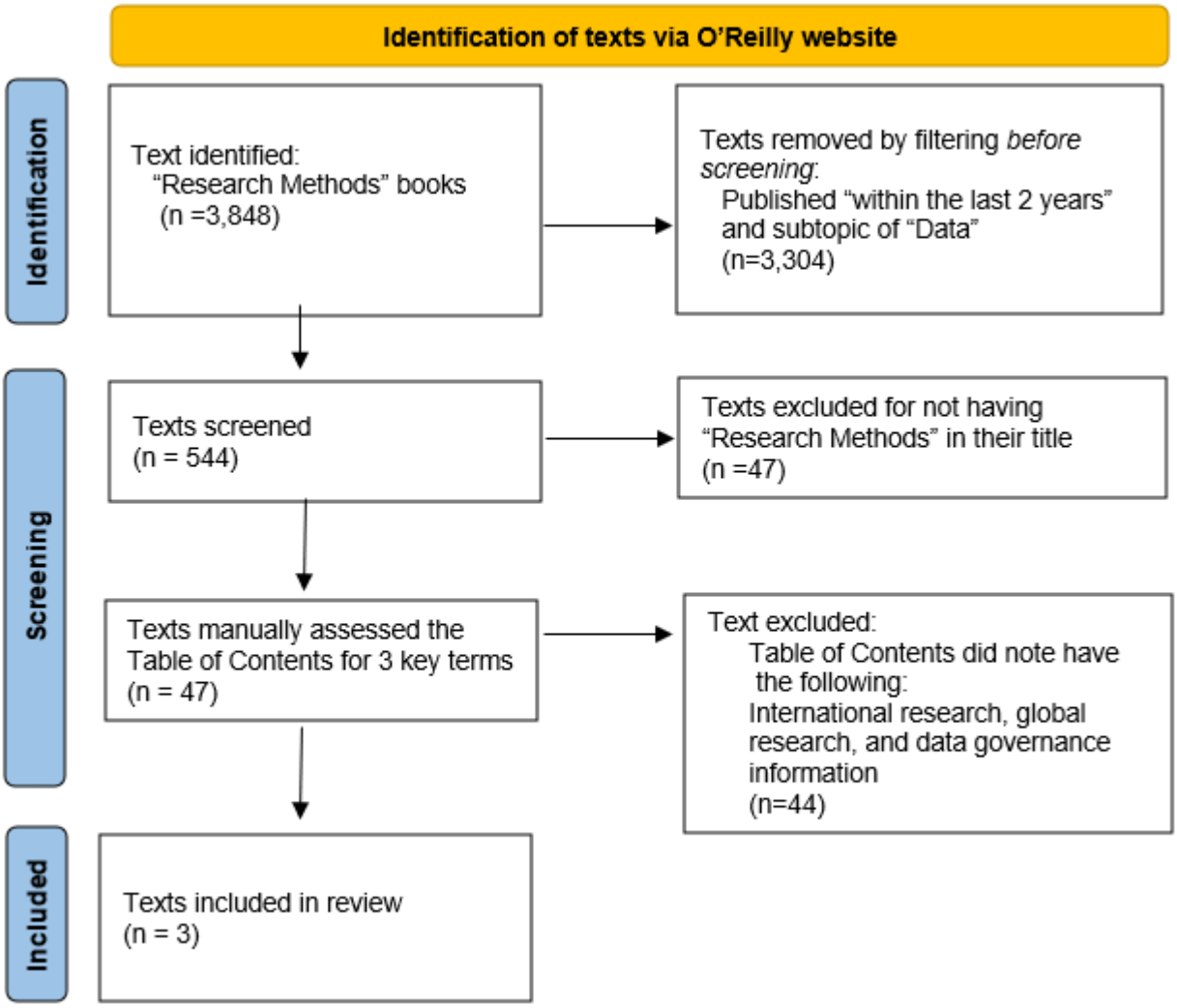


Figure 1. PRISMA flow diagram for textbook selection process

Appendix B

Country	Privacy Legislation	Website
Australia	Privacy Regulation 2013	https://www.legislation.gov.au/F2013L02126/latest/versions
Bangladesh	Data Protection Act (DPA) of 2023 (proposed)	https://bangladesh.gov.bd/
Brazil	Lei Geral de Proteção de Dados (LGPD) Law No. 13,709/2018	https://www.gov.br/esporte/pt-br/acesso-a-informacao/lgpd
Canada	Digital Charter Implementation Act (proposed)	https://ised-isde.canada.ca/site/innovation-better-canada/en/canadas-digital-charter-trust-digital-world
China	Personal Information Protection Law (PIPL)	http://www.npc.gov.cn/npc/index.html. (not translated to English)
Columbia	Statutory Law 1266 of 2008 Statutory Law 1581 of 2012.	https://www.sic.gov.co/en/deputy-superintendence-for-the-protection-of-personal-data
Denmark	Danish Data Protection Act	https://denmark.dk/privacy-notice
Egypt	Personal Data Protection Law	https://mcit.gov.eg/en/TeleCommunications/Regulations/Laws_and_Legislations
European Union England	GDPR Data Protection Act 2018	https://gdpr.eu/ https://www.gov.uk/data-protection
France	GDPR	https://www.diplomatie.gouv.fr/en/navigation/about/article/privacy-policy
Germany	Nationale de l'Informatique et des Libertés (CNIL)	https://www.diplomatie.gouv.fr/en/navigation/about/article/privacy-policy https://www.bfdi.bund.de/EN/Home/home_node.html
	GDPR	
Hong Kong	Federal Commissioner for Data Protection and Freedom of Information (BfDI)	https://www.bfdi.bund.de/EN/Home/home_node.html https://www.pcpd.org.hk/english/data_privacy_law/ordinance_at_a_Glance/ordinance.html
	Personal Data (Privacy) Ordinance (PDPO)	
India	Personal Data Protection Bill, 2023	https://prsindia.org/files/bills_acts/bills_parliament/2023/Digital_Personal_Data_Protection_Act,_2023.pdf
Israel	Protection of Privacy Law	https://www.wipo.int/wipolex/en/text/347462
Ireland	Data Protection Act 2018	https://www.dataprotection.ie/en/who-we-are/data-protection-legislation
Japan	Act on Protection of Personal Information	https://www.japaneselawtranslation.go.jp/en/laws/view/4241/en
Malaysia	Personal Data Protection Act 2010 (PDPA)	https://www.malaysia.gov.my/portal/content/654

Mexico	Federal Law on the Protection of Personal Data Held by Private Parties	https://s3.amazonaws.com/creel.mx.public/creelprod/2021/07/Data-Protection-in-Mexico-Overview.pdf
New Zealand	Privacy Act 2020	https://www.justice.govt.nz/justice-sector-policy/key-initiatives/key-initiatives-archive/privacy/
Philippines	Data Privacy Act of 2012 (Republic Act No. 10173)	https://privacy.gov.ph/data-privacy-act/
Qatar	Qatar Law No. 13 of 2016 concerning Personal Data Privacy	https://www.qfc.qa/-/media/project/qfc/qfcwebsite/documentfiles/resource-center/data-protection/regulations-and-rules/qfc-data-protection-regulations-and-rules-2021-guidance.pdf
Russia	Federal Law on Personal Data (No. 152-FZ)	https://pd.rkn.gov.ru/docs/Federal_Law_On_personal_data.doc
South Africa	Protection of Personal Information Act	https://popia.co.za/
Singapore, Thailand, Taiwan	Personal Data Protection Act (PDPA)	https://www.trade.gov/market-intelligence/thailand-personal-data-protection-act
Switzerland	Federal Act on Data Protection (FADP)	https://www.kmu.admin.ch/kmu/en/home/facts-and-trends/digitization/data-protection/new-federal-act-on-data-protection-nfadp.html
Turkey	Law on Protection of Personal Data	https://www.dlapiperdataprotection.com/index.html?t=law&c=TR
United Arab Emirates	The Dubai Data Law	https://www.digitaldubai.ae/docs/default-source/default-document-library/dde-module-1---overview-of-the-dubai-data-manual_v5.pdf?sfvrsn=aaae240_0
United States of America	No comprehensive federal law, but notable state laws include	
	California Consumer Privacy Act (CCPA)	https://oag.ca.gov/privacy/ccpa
	Virginia Consumer Data Protection Act (CDPA)	https://law.lis.virginia.gov/vacodefull/title59.1/chapter53/
Vietnam	Law on Cybersecurity and Law on Protection of Personal Information	https://baovedlcn.gov.vn/

Table 1. Privacy legislation by country

Teaching Case

A Business Intelligence Class Project Combining Business Knowledge, Critical Thinking, and Data Exploration for an Online Retail Store

Jason Triche
jason.triche@umontana.edu
University of Montana
Missoula, MT 59812

Mark McDiffett
markmcdiffett@yahoo.com
Real Chemistry

David Firth
david.firth@umontana.edu

Shawn Clouse
shawn.clouse@umontana.edu

University of Montana
Missoula, MT 59812

Abstract

Teaching Business Intelligence (BI) courses comes with challenges. Unlike traditional IS courses, where the entire course focuses on a specific topic, BI courses utilize skills from several domains, including business knowledge, data analytics, problem-solving, effective communications, and critical thinking. Finding relevant assignments and projects is difficult for instructors. In this paper, we introduce a three-phased class project from a fictitious online retail store that contains different customer segmentations, product offerings, shipping options, and sales regions. Each phase has different deliverables and is geared for different audiences within the business. The project can be modified for difficulty levels, student grade levels, and different course timelines. We introduce the project, lessons learned from assigning the project for several years, and teaching tips to adapt the project for different BI courses. Teaching tips discuss the use of different technologies, BI principles like star schemas and transforming data, and updating the dataset to work in subsequent semesters. After completing the project, the feedback from the students has been positive.

Keywords: Business Intelligence, IS Curriculum, Teaching Case, Class Project

Recommended Citation: Triche, J., McDiffett, M., Firth, D., Clouse, S.F., McDiffett, M., (2025). Teaching Case: A Business Intelligence Class Project Combining Business Knowledge, Critical Thinking, and Story-Telling for an Online Retail Store. *Information Systems Education Journal*. v24, n1, pp 42-51. DOI# https://doi.org/10.62273/BBEK9897

A Business Intelligence Class Project Combining Business Knowledge, Critical Thinking, and Data Exploration for an Online Retail Store

Jason Triche, Mark McDiffett, David Firth and Shawn Clouset

1. INTRODUCTION

Given today's data-driven economy, the demand for skilled Business Intelligence (BI) professionals continues to rise and BI professionals are in high demand (Olavsrud, 2023). Higher education is responding to the growing market demand by offering new courses, programs, and degrees in the discipline (Mills et al., 2016; Romanow et al., 2020). However, this growth in course offerings comes with significant challenges for educators. For example, unlike traditional IS courses, where the entire course focuses on a specific topic (e.g., relational databases), BI utilizes skills from several domains, including business knowledge, data analytics, problem-solving, effective communications, and critical thinking. Specifically, students learning BI must learn the appropriate tools and techniques but must also learn to become data-driven decision-makers (Jeyaraj, 2019). In addition, it is difficult for instructors to cover both breadth and depth in the timeframe of a typical college course (Negash, 2004; Shi et al., 2024). Students are expected to build the skill sets of storytelling and critical thinking based on data and outputs from BI applications (Knaflic, 2015).

To address some of these challenges, we created a BI class project that incorporates a variety of BI skills, including business knowledge, data modeling, problem-solving, data visualization, effective communication, storytelling, and critical thinking. In this project, students are provided data from a fictitious online retail store that has different customer segmentations, product offerings, shipping options, and sales regions. We also ask students to find and combine external data into their data model to enhance business insights. The project is broken into three separate phases with different audiences and deliverables for each phase. Our goal in this paper is to share the project details and our experience in delivering this project over the past several years. Our hope is that other BI instructors can use some or all of this project to enhance student learning in BI.

The project has several built-in features that are beneficial for building skills in BI. The first feature is related to the type of data. Although the data is simulated, it was created and modified based on a real online retail company. Students receive seven interrelated data files and are challenged to create a cohesive data model. Another feature is that the instructor can hide insights throughout the datasets to see if the students can find them when working on the project. We outline how instructors can do this below. The last feature is that the project's complexity level can be modified based on the class schedule, topics covered in the course, or student level. For example, instructors can assign this project in an undergraduate class by using only a few of the datasets, making the hidden insights easier to find, or only assigning the first phase of the project.

In this paper, we outline the prerequisite knowledge needed for the students and the instructors to complete the project successfully. We then introduce the entire project, the datasets, and the deliverables for all three project phases. Next, we provide lessons learned from offering this project for the past several years, including a technology discussion. Lastly, we explain how to hide insights in the datasets. The grading rubrics are provided in the appendix.

2. PREREQUISITE KNOWLEDGE

The audience for this project is primarily graduate students. We assign this project in a graduate-level business intelligence class. The class is a semester-long course consisting of 3 graduate-level credits. The class meets weekly for three hours, and each class consists of lectures, readings, discussions, and technology presentations. Although we implement this project in a graduate-level course, there is enough flexibility in the project that it could be retooled for an undergraduate business intelligence class. To retool for an undergraduate class, the instructor can choose to assign fewer datasets or limit the number or scope of the deliverables.

There are several prerequisites the students need for the project to be successful. Some of this knowledge must be covered prior to the student taking the course, some knowledge can be

covered in the course but before the project is assigned, and some knowledge can be covered after the project is assigned while students are working on the first project phase. Prior to taking the course, students must be familiar with public speaking, presentation skills, basic business knowledge (e.g., marketing, management, supply chain, operations, revenue, profit), and a basic understanding of data schemas (e.g., primary/surrogate key, foreign key). Before the project is assigned, the students must learn the basic functionality of Microsoft's Power BI (or equivalent software). Microsoft's Power BI is a collection of software services, apps, and connectors that work together to turn unrelated sources of data into coherent, visually immersive, and interactive insights (Microsoft, 2024). Other prerequisite knowledge includes basic principles of extracting, transforming, and loading data, star schema principles, and visualization principles. Power BI will allow students to transform and load the data, create a star schema, create custom calculations and fields, and create visualizations using the data. Once the project is assigned, but before the second and third deliverables are due, students must understand effective dashboards and the principles behind key performance indicators (KPIs). We introduce the full project below, and the prerequisites are listed again under each deliverable.

3. PROJECT DETAILS

The project involves a fictional online retail company called Big Sky Prints. The company asks a business intelligence consultant (i.e., a graduate student) for help analyzing sales data for the last six months to determine where the company can improve its business. The expert is asked to make sense of the data, find solutions to issues, and communicate the results to different audiences in the company over three different time horizons.

Big Sky Prints is a fictional Montana-based company that sells office supplies, furniture, and tech products entirely online. They started as a small regional operation but have grown steadily over the past few years. While they have kept things simple on the surface, the company has expanded into a wide variety of products, customer types, and shipping logistics. As they continue to expand, they are trying to get a better handle on their operations through data—figuring out what is working, where they can improve, and how to make smarter decisions moving forward. Although the company is fictitious, the product data is adapted from the SuperStore Sales Dataset (Starr, 2020). The remaining data (orders, shipping, sales reps, returns) are simulated in order to provide students with different avenues of exploration.

Table 1: Dataset Details

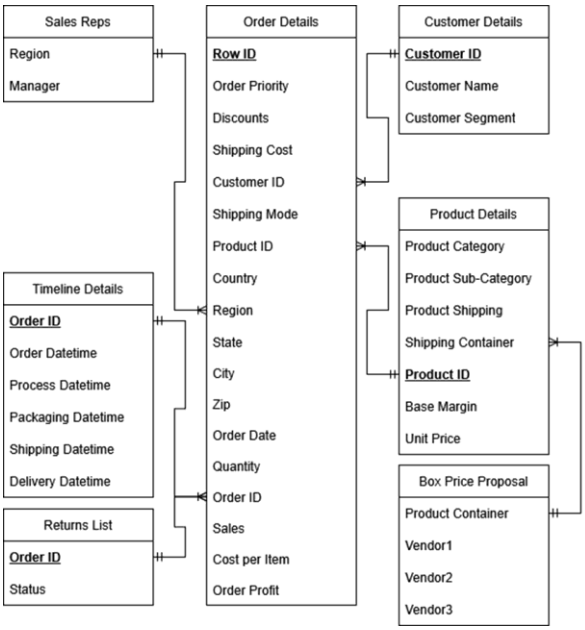
File	Explanation	Fields *primary key
Order Details	Individual records for each order.	• Row ID* • Order Priority • Discounts • Shipping Cost • Customer ID • Shipping Mode • Product ID • Country • Region • State • City • Zip • Order Date • Quantity • Order ID • Sales • Cost per Item • Order Profit
Product Details	Product category Info and product prices	• Product Category • Product Sub-Category • Product Shipping Container • Product Name • Product ID* • Base Margin • Unit Price
Customer Details	Customer Information	• Customer ID* • Customer Name • Customer Segment
Timeline Details	Detailed timestamps for processing and shipping orders	• Order ID* • Order Datetime • Process Datetime • Packaging Datetime • Shipping Datetime • Delivery Datetime
Returns List	List of returned orders	• Order ID* • Status
Box Price Proposal	Prices of seven different box types by three different vendors	• Product Container* • Vendor1 • Vendor2 • Vendor3
Sales Reps	Sales Manager names by region	• Region* • Manager

The course project is divided into three major deliverables – Past, Present, and Future. The first deliverable, Past, is a written report addressing the question, "What does all the data tell us about the company's performance?" The second deliverable, Present, is a dashboard addressing

the question, "How can we best use this data for information on a daily basis?" The third deliverable, Future, is a presentation addressing the question, "How can we use this data to improve our business?"

The students are provided with seven data files, all in the format of .csv. The datasets are explained in Table 1.

Figure 1: Entity Relationship Diagram (ERD)



Phase 1 - Past

The first phase, Past, asks for a written report to the audience of a large group of managers, accountants, and stakeholders answering the question, "What does all the data tell us about the company's performance". Defining the company's performance is up to the student, depending on their focus on the report. For example, students have used sales, revenue, profit, shipping efficiency, and customer loyalty in the past. The audience is broad on purpose, so the report can be interpreted by non-business intelligence professionals. This report requires three major sections. The first section is a summary report of five tables or charts of their choice and the importance of these tables or charts. These can all address one aspect of the business (products, processing efficiency, etc.) or be a high-level summary of the business.

The second section asks the students to find additional data and add it to the data model. The students must create at least two tables or graphs utilizing outside data in addition to the company's data. This data can be merged on timestamps (e.g., stock market data, weather, interstate closures), geography (e.g., demographic data, locations of other office supply stores, list of top ten vacation destinations for office managers), or anything else (e.g., competitor pricing).

The third section is more complex and asks the students to address three questions from management. The prompt tells the students that management would like help in organizing the data to make an informed decision. The prompts ask the students to create at least one table and one graph to summarize the data for each question below.

a. Management structure: Management is restructuring their account management team. They previously had four managers (one for each region) but have determined they need to increase the number of managers to seven. Create a chart/graph showing the number of customers and total value of accounts for each subdivision based on region (South, East, West, Central) and customer segmentation (consumer, corporate, consumer, small business) and propose the ideal way to split responsibilities between the seven managers. Alternatively, a case could be made to realign regions by comparing numbers by individual U.S. states.

b. The company has received three proposals for box suppliers for the next year. Management wants to know how much they would have paid each supplier in the last year. The company has decided to use one vendor only and cannot decide. Additionally, Vendor 3 has a promotion where small packages and wrap bags are free.

c. Management is reevaluating its complicated order priority structure and trying to simplify to only two levels - high priority and standard. Before doing so, management wants to know how each

priority level (critical, high, medium, low, not specified) is currently being processed. Determine the average process time (order to packaging) and total time (order to delivery) for each level by region. Then determine what the average times would be if these five levels were binned into two priority levels. Critical and high would be combined to high priority, and medium, low, and not specified would be combined to standard.

This phase forces students to create a story around how the business is performing. The students must weave a narrative on business performance for a wide audience.

The prerequisite knowledge to complete phase 1 of the project includes a basic understanding of basic business principles (e.g., marketing, management, supply chain, operations), Microsoft's Power BI or equivalent BI software, basic loading and transforming data, star schemas, visualization principles, and professional, business, and technical writing skills.

Phase 2 - Present

The second deliverable, Present, asks for a dashboard for the audience of a department manager or managers. The dashboard should be for a department manager to help them with a snapshot of their performance. Examples could include marketing managers, operation managers, sales managers, or supply chain managers.

The prerequisite knowledge to complete phase 2 of the project includes building an effective dashboard with effective data visualization principles.

Phase 3 – Future

The third phase, Future, asks for students to create a KPI and deliver the KPI in a presentation. The audience for this presentation is executive leadership. The prompt informs the students that the future portion will consist of a presentation to senior management proposing a KPI to use to help track and improve some aspects of the business. We ask the students to create a KPI, explain how the KPI is calculated, explain how the KPI drives results, and create a visualization used to utilize this KPI. The presentation should include the business goal, the KPI(s), how the KPI(s) is calculated, how the KPI(s) impacts the business goal, and a graph displaying the KPI(s) over the 6-month period using the data.

The prerequisite knowledge to complete phase 3 of the project includes building and tracking effective KPIs, building a presentation, and delivering a presentation to the appropriate audience.

4. LESSONS LEARNED

Teaching Tips

Before the introduction of the first phase, the students are provided with a prompt that directs them to import the data into a technology platform, create a data schema, and analyze the data. These directions are purposefully vague for several reasons. First, the project was created so students could explore different aspects of the business. For example, students could explore supply chain issues, vendor packaging optimization, delivery issues, processing delays, sales forecasts, revenue management, discounts, returns, and/or product margins. We wanted the students to choose a part of the organization that interests them. We did not want to lead the students into a specific area of the business.

Second, we did not want to limit the students to a specific technology. We encourage students to use Microsoft's Power BI. We cover several lessons in Power BI during the semester, including building star schemas, creating charts and graphs based on data visualization principles, and creating effective KPIs and dashboards. Therefore, students are familiar with Power BI. However, some students use other BI software like DataPine, Orange, KNIME, and Tableau.

The third reason for the vague instructions is for the student to explore making the star schema without any guidance from the instructor. As demonstrated in Table 1, some datasets do not have a primary key, nor does the primary key listed show the entire picture. For example, the order table contains a row ID as the primary key. The order table contains a row for each product ordered, but several rows could be tied to one order. Therefore, the order ID is repeated throughout the dataset. The proper way to tie an order to the timeline table is a combination of row ID and order ID. This happens throughout the datasets. The ERD in Figure 1 shows how the data can be connected. We would caution providing this to the students so they can try to create the data schema on their own.

Another teaching tip for instructors is spacing out the due dates for the phases. We found it useful to make the third phase due during the final

instruction week of the semester and back into the other two deliverable dates. We operate on a 16-week semester, where the 16th week is finals week. We made the first phase due during week 11 of the semester. Two weeks later, week 13, the second phase was due. Then two weeks later, week 15, the final phase and presentation was due. The disadvantage of spacing out the deliverables so early in the semester is that the instructor needs to cover all the concepts for phase 1 prior to week 11. The specific prerequisites are defined above in the project details. We found the shortened timeline was not a problem for us if we planned the course topic schedule correctly at the beginning of the semester.

An additional teaching tip is advice around the final presentation (i.e., phase 3). Each student had a strict 5-minute time limit to present their final deliverable with a 1-minute Q&A. At our university, the class was listed as a hybrid, which means students could be in-person, synchronous remote via Zoom, or completely asynchronous. We required all the asynchronous students to pre-record their presentations and upload their presentations to an unlisted YouTube link prior to class starting. During the class period, we randomly drew names to present. The students had to present either live in-person, live on Zoom, or through the prerecorded links. The 1-minute Q&A followed, and for those asynchronous, the students had to answer the question via the Teams class link the following day.

In Phase 3 – Future, students create KPIs for management. This is a difficult assignment for some students who may not have had strategic management courses previously. Some questions to help guide the students include, "What business goal are you trying to achieve," "What is being measured," "What is the timeframe," "Who is responsible for the KPI," "What are the industry benchmarks for this type of KPI." Sample KPIs could include "number of new customers" or "same-day shipping percentage."

Keeping students on schedule while delivering high quality is always a challenge for instructors. We set aside the last 15 minutes of each class period to check in with the students on their projects. We provided hints and tips to individual students who were struggling. We avoided answering individual questions to the entire class so each student could attempt to solve problems on their own. For the asynchronous students, we set up discussion boards within Microsoft Teams for students to ask questions. Again, we avoided posting possible solutions to individual problems to the entire channel to encourage individual problem-solving.

The final teaching tip involves updating the dataset so the project can be repeated every semester or year. The instructor only needs to update the dates in the Order Details and the Timeline Details files. We found the easiest way to do this is in either Power BI or Excel. In Excel, we searched for the string '2022' and changed it to '2023' or whatever year we needed. Moving the date up one year created more interesting insights since it appeared that Sunday produced a lot of sales, and shipping could fall on a Sunday accidentally. Some students caught this and suggested that the company reprioritize shipping on weekends.

Hidden Insights

One of the tenets of business intelligence is to discover hidden insights in the vast amount of business data. Teaching students to find hidden insights can be difficult. A dynamic aspect of the project is that the instructor can modify the datasets to hide business insights for the students to discover. The instructor can make these hidden insights easy to find or challenging to discover. An example of an easy insight is to change the returns dataset to only list returns for a few customers or by a specific regional sales rep. If the student discovers this insight, then the student can report suspicious activity of the sales rep if they have an extraordinary amount of returned products compared to their sales rep peers. In this example, the instructor would need to sort the order table by region, copy the order IDs in a certain region, and paste the order IDs over the order IDs in the returns table. The same procedure can be done for the customer returns but sorting the order table by customer IDs and then grabbing order IDs linked to a specific customer. Then, the instructor can paste those order IDs over the order IDs in the returns table.

A challenging insight would be if a student calculated the time difference between the order being placed, when the order is being processed, and when the order is being packaged. The students would discover that some products take an extraordinarily long time to process and package certain products. This could indicate supply chain issues for certain products or product categories/sub-categories. The instructor can modify these dates in the timeline dataset and change these dates by order ID.

Student Feedback

After completing the project, the feedback from

the students was positive. Quotes from students include, "Project Overall was a good test of knowledge and simulated real-work business expectations," "Good project, challenging," and "the final project were all great practice and helped me learn the most." There were no negative comments from the students about the project.

5. CONCLUSION

This paper introduces a graduate-level BI course project for instructors to use in their courses. The project covers a wide variety of BI skills, including business knowledge, data modeling, problem-solving, data visualization, effective communication, storytelling, and critical thinking. The project is introduced in three phases with different time horizons, deliverables, and audiences for each phase. Because of the flexibility of the project phase, datasets, and hidden insights, the project can be modified for an undergraduate class, or shortened in scope or duration to fit different BI courses.

6. REFERENCES

Jeyaraj, A. (2019). Pedagogy for business analytics courses. Journal of Information Systems Education, 30(2), 67.

Knaflic, C. (2015). *Storytelling with Data: A Data Visualization Guide for Business Professionals.* John Wiley & Sons, Ltd. https://doi.org/10.1002/9781119055259.

Microsoft. (2024, March 22). *What is Power BI?* https://learn.microsoft.com/en-us/power-bi/fundamentals/power-bi-overview.

Mills, R. J., Chudoba, K. M., & Olsen, D. H. (2016). IS Programs Responding to Industry Demands for Data Scientists: A Comparison Between 2011-2016. *Journal of Information Systems Education,* 27(2), 131.

Negash, S. (2004). Business Intelligence. *Communications of the Association for Information Systems,* 13(1). https://doi.org/10.17705/1CAIS.01315.

Olavsrud, T. (2023, December 1). 10 business intelligence certifications and certificates to advance your BI career*. CIO.Com.* https://www.cio.com/article/219932/11-business-intelligence-certifications-to-advance-your-bi-career.html.

Romanow, D., Napier, N. P., & Cline, M. K. (2020). Using Active Learning, Group Formation, and Discussion to Increase Student Learning: A Business Intelligence Skills Analysis. *Journal of Information Systems Education,* 31(3).

Shi, Y., Gebauer, J., Kline, D., & Gillenson, M. (2024). Teaching a Report-Oriented Business Intelligence Course: A Pedagogical Experience. *Journal of Information Systems Education,* 73–85. https://doi.org/10.62273/RTPL4395.

Starr, Joe (2020). US Superstore data, Mendeley Data, V1, doi: 10.17632/bfgzt5rh35.1

APPENDIX A

Grading Rubrics

Phase 1 – Past (written report)
Total: / 100

Criteria	Scale
Summary Report 5 tables or charts Explain their importance 15 pts	1 – Less than 3 reports. 2 - 3 or 4 reports or missing explanations of their importance. 3 - 5 reports with some explanations of their importance. 4 - 5 reports all with explanations of their importance. 5 – 5 reports all with clear explanations of their importance and how they are tied to the business.
Outside Data 20 pts	0 – no outside data used. 3 – outside data used, but not relevant or used correctly. 5 – outside data used correctly and relevant to address the business.
Management Structure Chart/graph with number of customers and total value of accounts per division, region, customer segment. Propose new responsibilities with 7 managers 15 points	1 – Missing graphs, evidence, and proposal is questionable. 2 - Graph is missing dimensions, and proposal is satisfactory and missing evidence and proposal is questionable. 3 - Graph is missing dimensions, or proposal is satisfactory and backed up with evidence or proposal is questionable. 4 - Contains all graphs, proposal is complete and backed up with evidence. Proposal is questionable. 5 – Contains all graphs, proposal is complete and backed up with evidence. Proposal is feasible.
Container Proposal How much was paid to each vendor Recommended vendor Contains assumptions about box per item or per order 15 points	1 – Incorrect table/chart with amount paid to vendor, recommendation is missing. Missing assumptions. 2 - Incorrect table/chart with amount paid to vendor, recommendation is unclear. Missing assumptions. 3 - Contains table/chart with amount paid to vendor, recommendation is unclear. Missing assumptions. 4 - Contains table/chart with amount paid to vendor, recommendation is clear and feasible. Missing assumptions. 5 – Contains table/chart with amount paid to vendor and recommendation is clear and feasible. All assumptions are clear.
Priority structure How each priority level is being processed Average process time (order to packaging) per region Total time (order to delivery) per region Average time with new bins 15 points	1 – Missing calculations. 2 - Contains mostly incorrect calculations. 3 - Contains half-correct calculations. 4 - Contains mostly correct calculations. 5 – Contains all correct calculations.
Professionalism/Structure/Grammar 20 points	1 - Poorly written 2 - Several errors, minimum professional 3 - Several errors, professionalism is average 4 - Few errors, written well, professional 5 – No errors, written well, professional

Phase 2 – Present (performance dashboard)
Total: / 100

Criteria	Scale
Dashboard 50 pts	1 – Dashboard does not meet data visualization principles (all axis are labeled, use of color is correct, graphs are easy to read, uncluttered) 2 - Dashboard does not meet most data visualization principles (all axis are labeled, use of color is correct, graphs are easy to read, uncluttered) 3 - Dashboard meets some data visualization principles (all axis are labeled, use of color is correct, graphs are easy to read, uncluttered) 4 - Dashboard meets almost all data visualization principles (all axis are labeled, use of color is correct, graphs are easy to read, uncluttered) 5 – Dashboard meets all data visualization principles (all axis are labeled, use of color is correct, graphs are easy to read, uncluttered)
Audience 50 pts	1 – The audience for the dashboard is not clear. The graphs do not directly help the audience. Audience is not focused. 2 – The audience for the dashboard is not clear. The graphs typically do not directly help the audience. Audience is not focused. 3 – The audience for the dashboard is clear. The graphs may directly help the audience. Audience is not focused. 4 – The audience for the dashboard is clear. The graphs will directly help the audience. Audience is not focused. 5 – The audience for the dashboard is clear. The graphs will directly help the audience. Focused audience.

Phase 3 – Future
Total: / 100

Criteria	Scale
Dashboard 20 points	1 – Dashboard does not meet data visualization principles (all axis are labeled, use of color is correct, graphs are easy to read, uncluttered) 2 - Dashboard does not meet most data visualization principles (all axis are labeled, use of color is correct, graphs are easy to read, uncluttered) 3 - Dashboard meets some data visualization principles (all axis are labeled, use of color is correct, graphs are easy to read, uncluttered) 4 - Dashboard meets almost all data visualization principles (all axis are labeled, use of color is correct, graphs are easy to read, uncluttered) 5 – Dashboard meets all data visualization principles (all axis are labeled, use of color is correct, graphs are easy to read, uncluttered)
Business Goal 15 points	1 – No mention of business goal. 2 – The business goal is implied by not stated 3 – The business goal is defined but not clear. 4 – The business goal is defined and is somewhat clear. 5 – The business goal is defined and clear.
KPI overview & calculation What is being measured? Timeframe? 15 points	1 – The measurement is not stated (ratio, %, ranking, #), and timeframe is not stated, and benchmark is not stated. 2 – The measurement is not stated (ratio, %, ranking, #), or timeframe is not stated, or benchmark is not stated. 3 – The measurement is not clear (ratio, %, ranking, #), or timeframe is not clear, or benchmark is not clear. 4 – The measurement is stated (ratio, %, ranking, #), timeframe is stated, benchmark is stated. 5 – The measurement is clear (ratio, %, ranking, #), timeframe is clear, benchmark is clearly stated.
KPI impact on business goal Does success of this measure lead to success of the business goal? 15 points	1 – Not stated. 2 – Implied, but not stated 3 – Stated, but unclear 4 – Stated 5 – Stated and clear
KPI Graph for 6 months 15 points	1 - Not present 5 - Present
Presentation 20 points	1 - The presenter is unclear and lacks confidence, making it impossible to understand the presentation. 2 - The presenter is unclear and lacks confidence, making it difficult to understand the presentation. 3 - The presenter is hesitant and lacks confidence, making it difficult to follow the presentation. 4 - The presenter is clear and mostly confident, but could be more engaging. 5 - The presenter is engaging, confident, and delivers the presentation with enthusiasm.