Research Article

Developing Data Analytics Skills in 21st Century Based on Open Data

Vipakorn Vadhanasin[1] and Dit Suthiwong[1*]

[1] *Information Studies Program, Faculty of Humanities, Srinakharinwirot University, 114 Sukhumvit 23 Road, Wattana, Bangkok 10110, Thailand*
*E-mail: dit@g.swu.ac.th

Abstract

Teaching data analysis with open data allows learners to practice the 21st century skills of data filtering and preparation, systematic analysis, creating output and communication, and problem-solving. The purposes of this research were to 1) apply open data for teaching and learning in data analysis course 2) compare achievements after learning process, and 3) study the satisfaction of learners received from open data exercises. Data was collected through interviews and classroom trails for open data lessons. The target group used in group discussion were students who registered for data analysis in academic year 2021 in 2 groups. Each group has 5 students with a total of 10 students. Researchers conducted in-dept interviews with 2 experts and developed class materials for teaching and learning. Quantitative data were collected from pre-test and post-test scores of 33 learners enrolled in the academic year 2022. The findings reveal an opportunity to develop and an open data application to be used in teaching and learning. The comparison of learning achievement by paired-sample t test showed that the mean score after learning was higher than the mean score before learning at a statistically significant level of 0.01. The detailed comparison results revealed that the average scores the average score after learning was higher than the average score before learning at the statistical significance level of 0.05 in the 21st century skills of 1) data accessing, filtering and preparation 2) creativity and presentations, and 3) problem solving. On the contrary, there was no difference in analytical thinking skills. The results showed that learners were satisfied with open data application used in teaching and learning at a high level ($\overline{x}$ = 4.21). The results of this study will be useful to apply open data in data analysis courses, which will develop 21st century skills of the learners.

Keywords: Open Data, 21st Century Skill, systematic thinking, Data Visualization

1. Introduction

Digital government transactions have led to increased collaboration in data sharing for public policy purposes (1). Collaboration in creating open data by government agencies under government oversight demonstrates transparency in operations, which creates confidence among the public, allowing them to access and verify data (2). The quality of open data can be evaluated based on its completeness, timeliness, accessibility, and compliance with legal regulations. Government open data sources in Asia, such as data.go.jp of Japan, data.gov.in of India, and data.go.id of Indonesia, are examples of such sources.

In recent years, data about COVID-19 has become popular for analysis and data visualization. Open data is an important resource for managing education because it can be used freely, without limitations, and can be reused. An example of this is the COVID-19 pandemic, which has made people around the world

interested in data about the number of people infected with the virus each day, which has been increasing rapidly.

However, having open data from the government that is freely available to the public does not necessarily reflect the value of the data. Users must have an understanding of the context of the data and have analytical skills to use the data effectively (3). Research on how to use government open data to create value is still limited (4).

Obtaining high-quality data helps students understand the benefits of data analysis and presentation. Developing deep analytical skills requires accurate, up-to-date data sources. Data should be relevant to the situations that students are interested in, in order to create understanding of the data structure and design the desired data output. Students can develop skills in creating data models to use in analysis. The results can be used to tell stories from the data that has occurred. Teaching and learning are designed to train students to have the skills necessary for the 21st century, which are essential for employment after graduation. The skills necessary for digital knowledge include: Understanding how to access and use digital data, The ability to analyze and interpret digital data, The ability to use digital tools to present and communicate data, The ability to use critical thinking to evaluate the credibility and accuracy of digital data.

Lie et al. (5) defines technology-based learning as an approach to education that uses technology as a tool to support and enhance the learning process. This can include the use of software, hardware, internet and network resources, and the development of teaching skills to create a more engaging and interactive learning experience. The goal of technology-based learning is to achieve desired learning outcomes for students, and for this, it is important that educational institutions prepare the necessary resources and infrastructure to support technology-based learning. This may include providing students with access to the appropriate technology, providing professional development opportunities for teachers to develop their technology-related skills, and creating a supportive environment for technology-based learning.

The ADDIE model (6-7) is a commonly used framework for the development of instructional materials and training programs, which stands for Analysis, Design, Development, Implementation, and Evaluation. It was first developed by the U.S. military in 1975, but it has since been adapted and applied to various fields and industries. The ADDIE model provides a systematic approach to instructional design and development, with the goal of increasing the effectiveness and efficiency of the learning process.

From reviewing literature that uses open data in the classroom, it can be seen that the benefits of open data can be seen in terms of training data skills and providing knowledge from real data, which is demonstrated by various research studies from other countries, such as (8) found that students are able to use tools beyond those taught in class, allowing non-science majors to analyze data in their field of expertise in a more profound way. Asamoah (9) found that students are able to connect data from various sources and have a good overall understanding of the data. While Nestorov et al. (10) found that using real-world data from outside organizations helped students understand the metrics used by the organization, the problems it faces in its context, and ways to help address these problems. Additionally, Saddiqa, Larsen, et al. (11) found that the use of open data made students more interested, fostered collaboration, and helped them develop critical thinking skills.

Consequently, in this study are employed open data analytics to motivate learners to develop 21st century skill. The research finds that the use of open data analysis in Business Intelligence can enhance 21st century skills for students, as it helps them to practice analytical thinking, problem solving, data filtering and preparation, and creative work presentation. Furthermore, this study applied Technology-based learning and the ADDIE model to the development of learning media, and tested the effectiveness of the learning method on a focus group of students through pre and post-course knowledge assessments and in-depth interviews with experts in open data services. Additionally, the research also found that the use of data visualization tools such as Tableau or Power BI can make data analysis more accessible for students and can focus on the process of thinking and presenting creatively.

2. Materials and Experiment

This research aims to study the use of open data sources and develop tools for teaching in the course. The research has conducted focus group discussions with students and gather feedback on the development of open data repositories in the classroom. Additionally, experts in open data policy had been interviewed to gather information on current open data services provided by the government, both domestically and internationally, and how they can be applied in the classroom. In the final phase of the research, the developed open data repository had been tested with all students registered in the class, starting with a pre-test at the beginning of the first week, and ending with a post-test at the end of the tenth week. Because this research was funded to study and develop learners in Information Studies. Therefore, 33 students with at least two years of experience in this field were selected. The test scores and satisfaction with the media used in the teaching had been collected from a sample as follows in Table 1.

The target group for the focus group discussion is students who have completed the course, and their experiences and suggestions for developing open data repositories in the classroom will be collected. The goal is to gather data from students about their experiences using open data sources in the course, including any challenges they encountered and how they overcame them, as well as their suggestions for improving the open data repository for future classes. This information will be used to improve the use of open data sources in the course and enhance students' learning experiences.

The target group for the in-depth interview is government officials who are currently working on open data repositories. The study aims to understand the current state of open data repositories provided by the government, both domestically and internationally, as well as their applications in education. The data will be collected through in-depth interviews with 2 experts who have completed a graduate degree and have played a significant role in promoting open data policy and implementation in the government. These experts have been involved in the policies, and are well-respected in their field.

3. Results and Discussion

3.1 The results from the student focus group interview.

The students surveyed data from kaggle.com and data.go.th as the main source for data analysis and presentation. The data from kaggle.com is more popular because it has a more comprehensive and diverse range of data, while students use data from data.go.th when presenting a topic that is official. Some students also survey image, video, and audio data for additional use in creating visualizations.

Because the open data that students choose to analyze is what they want to know themselves. The dataset from kaggle.com is diverse such as usage data of Sportify, YouTube, Netflix, Statistic of Japanese anime cartoon data, Salary data and Career Path data of Scientist career. These encourage students to concentrate on analyzing data and creating good work. But using data from only one data set can only provide basic analysis. It's not as useful as linking multiple data sets together and analyzing them, for example analyzing the Netflix dataset, comparing statistical numbers with the Disney dataset, for example.

For data from data.go.th, the problem was found that some open data groups had too few lines of data. Sometimes the data is not updated, so it takes time to search further to

Table 1 Illustrates the research design**.**

Research method	Research Objective	Participants	Data Collection Method
Focus Group	Interview	N= 10 studies	Literature review
Expert	In-depth Interview	N= 2 studies	Literature review
Focus Group	Compare students' learning achievement between pretest and posttest of the students learning through open data analytic activity	N= 33 studies	Systematic Review

obtain sufficient information for analyzing and creating Data Visualization.

By using open data, students learn from real-world situations, encounter various problems, and practice problem-solving skills. They also have the opportunity to communicate and collaborate with their group members and exchange ideas and feedback from their peers through the presentation of their dashboard projects in class.

3.2 The results from the expert interview.

Open data provided by the government is accurate and up to date, has clear standards for disclosure, and is in a format that is easily usable by the private sector and the public. It is also easier to access than in the past. However, users should be cautious and critically evaluate the data's accuracy and reliability. Data professionals should have 21st century skills in data filtering, analysis, prediction, diagnosis, and presentation. In addition to the skills related to data directly, legal and regulatory knowledge related to data is also important.

Using open data in the classroom allows students to practice 21st century skills by accessing various and selective data that interests them. In addition to training in the classroom, using open data also provides opportunities for students to create works and contribute to the community and network, and to present their ideas and insights based on credible data. This is done through the platform both at the national and international level.

3.3 The results from statistical data analysis using open data repositories.

Evaluations of students' learning results before and after learning using Open data repository were compared. Results were used for research purposes only, without affecting learners' learning outcomes. The test results of 33 students who took the pre-test and post-test were used to analyze the results. The analysis in Table 2 found that after using an open data to develop 21st century skills using an Open data repository, students had a statistically significant improvement in their learning performance at a level of 0.01.

Table 2 The results of learning performance before and after using an open data repository.

Test	N	$\bar{x}$	S.D.	t	Sig.
Pre-test	33	8.45	2.166	-4.796	0.000*
Post-test	33	10.55	2.785		

p-value < .01

When considering the learning performance based on 21st century skills, Table 3 shows that students had a statistically significant improvement in their performance compared to before learning at a level of 0.05 in the following skills: data access, screening, and preparation, creative work production and presentation, and problem-solving, while the skill of systematic analysis did not show a significant difference.

Table 3 The results of learning performance before and after, categorized by 21st century skills.

Score	Pre-test		Post-test		t	Sig.
	$\bar{x}$	S.D.	$\bar{x}$	S.D.		
1. Data access, screening, and preparation	2.52	1.176	3.12	0.960	-2.789	0.009*
2. Systematic Analysis	2.36	1.084	2.45	1.063	-.432	0.669
3. Creative work production and presentation	1.85	0.972	2.48	1.064	-2.592	0.014*
4. Problem-solving	1.73	1.069	2.48	0.972	-3.029	0.005*
TOTAL	8.45	2.166	10.55	2.785	-4.796	0.000*

p-value < .05

The results from the t-test dependent in Table 3 to compare the mean scores of students before and after using open data in the class showed that the mean score after the class was higher than the mean score before the class at the 0.01 level of significance (t = -4.796). The results of this research are a reflection of the introductory data analysis course. Demonstrates

an opportunity to develop students' critical thinking skills. Teachers should improve the assignment of students to practice working with datasets that require more critical thinking skills.

From Table 4, it was found that the students have a high level of satisfaction with the open-source media used in teaching and learning ($\bar{x}$ = 4.21). When considering each item, it was found that all items have an average score in the high level. The item that students have the highest satisfaction is the media being modern ($\bar{x}$ = 4.52), followed by students can benefit from the media ($\bar{x}$ = 4.39) and the convenience of using media in teaching and learning ($\bar{x}$ = 4.33) in order.

Table 4 Satisfaction with open-source media used in teaching and learning.

Test	$\bar{x}$	S.D.	Level of Satisfaction
1. The media has appropriate and clear content.	4.30	0.637	High
2. The media is well-organized and easy to understand.	4.09	0.843	High
3. The media is interesting.	4.06	0.827	High
4. The media is modern.	4.52	0.755	High
5. The media can attract the students' interest.	3.82	0.808	High
6. The media has appropriate and clear content.	4.00	0.901	High
7. The media stimulates students' interest in learning through technology.	4.18	0.846	High
8. Students can use the media independently.	4.30	0.883	High
9. Convenience of using media in teaching and learning.	4.33	0.736	High
10. Students can benefit from the media.	4.39	0.704	High
7. Open Data from real-world situations is beneficial for learning.	4.30	0.637	High
Total	**4.21**	**0.797**	**High**

4. Conclusions

The study provides clear evidence that using open data in the class can enhance 21st century skills. Demonstrating the benefits of open data in Thailand and abroad student learning process. The importance of the study results of students' learning achievement before and after the learning process by developing an open data warehouse for teaching and learning management as well as student satisfaction.

The students are highly satisfied with the open data media used in teaching and learning, with an average satisfaction level of 4.21. The aspect that received the highest satisfaction level is the media being up-to-date, with an average of 4.52. The other aspects that received high satisfaction levels include the students being able to use the media to their advantage (4.39), and the convenience of using the media in teaching and learning (4.33) in that order.

A suggestion for further research is to study the use of open data in class with students who have expertise. To allow students to conduct in-depth analysis in their areas of expertise and exchange knowledge discovered from data such as finance, education, entertainment, health sciences, agriculture, etc.

Declaration of conflicting interests

The authors declared that they have no conflicts of interest in the research, authorship, and this article's publication.

References

1. Ruijer E, Meijer A. Open government data as an innovation process: Lessons from a living lab experiment. Public Performance & Management Review. 2020;43(3):613-35.
2. Piotrowski SJ. The "Open Government Reform" movement: The case of the open government partnership and US transparency policies. The American Review of Public Administration. 2017;47(2):155-71.

3. Gebre EH, Morales E. How "accessible" is open data? Analysis of context-related information and users' comments in open datasets. Information and learning sciences. 2020;121(1/2):19-36.

4. Utamachant P, Anutariya C, editors. An analysis of high-value datasets: a case study of Thailand's open government data. 2018 15th international joint conference on computer science and software engineering (JCSSE); 2018.

5. Lie V, Kauutuk DA, Palilingan R, Tambingon HN. Technology-Based Learning Management at State Vocational Schools In Minahasa Selatan. Eduvest-Journal of Universal Studies. 2022;2(11): 2316-29.

6. Allen WC. Overview and evolution of the ADDIE training system. Advances in developing human resources. 2006;8(4): 430-41.

7. Welty G. The'design'phase of the ADDIE model. Journal of GXP Compliance. 2007; 11(4):40-53.

8. Lo LY-H, Ming Y, Qu H, editors. Learning vis tools: Teaching data visualization tutorials. 2019 IEEE Visualization Conference (VIS); 2019.

9. Asamoah D. Improving Data Visualization Skills: A Curriculum Design. Int. j. educ. dev. using inf. commun. technol. 2022; 18(1):213-35.

10. Nestorov S, Jukić N, Rossi S. Design and Implementation of a Data Visualization Course with a Real-World Project Component in an Undergraduate Information Systems Curriculum. J. Inf. Syst. Educ. 2019;30(3):202-211.

11. Saddiqa M, Rasmussen L, Magnussen R, Larsen B, Pedersen JM. Bringing open data into Danish schools and its potential impact on school pupils. Proceedings of the 15th international symposium on open collaboration. 2019;1-10