Comparative analysis between theoretical and simulatory learning methods by data science methodology approach

Patani Rucha Sharadkumar¹, S Shankar Narayan²

¹Final year student, Department of Computer Science Engineering, CMR Institute of Technology, Bengaluru- 560 037, India.
²Assistant Professor, Department of Mathematics, CMR Institute of Technology, Bengaluru- 560 037, India.

Corresponding author email address: pata17cs@cmrit.ac.in

Abstract. Data Science is the most trending interdisciplinary science that integrates the steps of data collection, preprocessing data, transforming data, storing data, data visualization, and hence extracting the insights from the data to serve stakeholder’s purposes. Python is commonly used and, along with being a versatile and open-source language, is a favorite tool in Data science studies. The vast libraries are being used to manipulate data and are very simple for even a beginner data scientist to understand. In the present work, we intend to apply the data science methodology to decision making and predictive analysis using the python programming language. We consider the problem of selecting the better mode of study concerning some of the impractical phenomena from physics for the exact understanding of the process. Data collection has been from an educational institute and the comparison has been made between theoretical learning and simulatory learning for selected topics from the vast fields like mechanics, thermodynamics, fluid dynamics, and radioactivity. The steps of data science methodology are germinated to achieve the insights into the data procured and the results are wangled concerning the teaching methodology that could be employed. In the present work, we undertake a comparative study between the theoretical and simulatory modes of teaching by exploring the modes individually through evaluating the responses imparted by a class of high school students. The analysis reported the more inclination of the student’s responses towards the simulatory methods when compared to the theoretical method of learning.

1. Introduction
The teaching-learning process in high school is equipped with theoretical learning as well as practical learning. There has been a drastic development in the teaching-learning methodologies from time to time. There are several research articles and books based on the possibilities of introducing various online-related activities in teaching as well as learning processes. A comprehensive perspective on higher education through responsible teaching with pragmatic discussion and organizing meaningful learning activities were explored in previous research articles [1].
Also, the role of moodle technology in physics learning to implement an online learning community was investigated in early research [2]. Designing and implementing the evaluating system to measure teacher performance was studied based on the student scores [3]. Discussion on the challenges faced in science education through theoretical, methodological, and practical techniques was made in past studies [4]. But these studies did not attempt to introduce the simulatory method of learning into their investigation.

Practical learning of a concept is more advantageous and thought-provoking among students when compared to theoretical learning. Practical learning is generally characterized as a task wherein learners observe and then execute, create, or manipulate materials or experience a demo if they are unable to execute it for safety or other reasons. It can be an enjoyable way of learning, which can improve by listening and writing. Using practical learning helps a student to gain hands-on experience and understand the methods used in the course from the field. Practical learning can make one feel confident and competent in the knowledge and skills [5]. Unfortunately, not all the concepts can be practically studied or even the demonstration is impossible for some of the concepts. In such cases, the teaching-learning processes can be made interesting by introducing the method of simulation. Simulation tools available can be used to study numerous topics of interdisciplinary sciences without actually experimenting. Simulation’s key benefits include: testing a system's actions without constructing it, findings are usually correct compared to the empirical model, helping to identify unexpected phenomena, system behavior, easy to perform "What-if" research [6,7]. In the recent world, simulation has taken a significant role in the fields of Physics and Medicine.

Physics is a natural science that covers the entire universe through its basic fundamental laws. Topics like radioactivity, quantum mechanics, thermodynamics, fluid mechanics, optics form some of the major branches of the subject Physics. High school physics covers some of the mentioned topics theoretically and practically. But the topics like radioactivity, fluid dynamics, and few concepts of thermodynamics cannot be examined practically as the practical study of these concepts requires sophisticated tools and types of machinery for their investigation. In these cases, the implementation of simulation tools would make the study more fascinating and as the simulation requires only the literary data, the safety of the student and his/her surrounding is maintained [8].

In the present study, we employ the Free and Open Source Software Python to simulate the concepts that are unachievable practically. As discussed in the previous paragraphs, the python programming language is very easy to handle due to its code flexibility by even a normal person without programming knowledge. Being a programming language, it can even simulate real-world processes. The Jupyter Notebook is an open-source web-based application that allows users to build and share live code, calculations, visualizations, and narration text files. Data screening and conversion, numerical simulation, mathematical modeling, visualization of data, machine learning, and much more are used. We use Jupyter Notebook hosted by the Anaconda platform to run the python code under our present study.

The present work comprises three upcoming sections. In Section 2, we define the purpose of our present study. Section 3 clears out about the methodology applied in the present exploration. Using the Data Science Methodology approach adds novelty and strength to our investigation. A brief introduction to Data Science Methodology and its stages are discussed in this section. In section 4 of this paper, we intend to analyze the results obtained and hence the conclusions were presented based on the analysis made, in the final section.
2. Purpose of the study
This research aimed at presenting certain concepts of physics (like radioactive decay, pressure-altitude relation, etc.,) in a high school class comprising students in the form of teams. The topics were explained theoretically and the same topics were investigated using the results obtained by simulating the governing models of these concepts. At the end of the theoretical presentation and the simulation results- used explanation, an assessment was conducted. Answers given by the same teams for the different modes of teaching were compared.

Based on the analysis made on the replies, the following questions were answered:
- Was there any change in the way of presenting the answers?
- How did the students use the python results to further their study of physics?
- Which mode of teaching resulted in a better understanding of the concepts?

3. Methodology
Data Science Methodology involves a series of interdependent steps starting from Business understanding reaching the model deployment and testing. It is a recursive scheme of mechanisms that governs data scientists via an approved series of steps to the ideal approach to addressing data science problems. This methodology guides the study systematically by validating each step involved in the process. Thus it implies a structure to our study. Multiple steps can be taken together, such as data requirements and collection steps, but the positive thing is that this structure points us to the most efficient exercise of our time and resources as we solve data science problems. This feature of data science methodology is unique and very efficient in problem-solving. In our present work, we only reach the stage of evaluation and the deployment stage is expected to be the future alterations in the mode of teaching. Table 1 forms the most important part of our present work as it explains the various steps involved in the data science methodology, as designed by John Rollins [9] corresponding to our present investigation.

![Decision Tree](image)

**Figure 1.** Decision tree

| Stages                 | Current scenario                                                                 |
|-----------------------|----------------------------------------------------------------------------------|
| Business Understanding | - Clarity around the problem to be solved<br>- To determine what data to be used to answer the question: “Can learning science topics be made |
inspiring by simulation methods?"

**Analytic Approach**

- A predictive model can be explained using decision trees, as in figure 1, to decide the efficiency of using simulation methods using free and open-source software (FOSS).

**Data Requirements**

- The identification of necessary data content required for the present study has been decided in a proper format.
- Sources to collect the data have been decided. That is, prior permission has been obtained from the institutional head to conduct the survey mentioned in the later section of the paper.

**Data Collection**

- A survey has been made to collect the data in an unbiased manner using Python programming language hosted using Jupyter Notebook.
- Gaps in the data have been identified and they were substituted by the replies made by other students.
- The time taken by each student to answer the question and also the responses for the final poll are recorded for our analysis, in the form of a CSV file.
- The details of the data set collected can be found in table 2.

**Data Understanding**

**Data Preparation**

- We use descriptive statistics and visualization techniques just after original data gathering to explain the collected data, analyze the quality of the data, and uncover preliminary information from the data. There could be a need for additional data collection to plug gaps.
- As our present study requires the collection of simple data, the data preparation stage is achieved with the understanding stage.

**Modeling**

- A training set has been used for the predictive model formed and a model has been created by parameter tuning techniques.

**Evaluation**

- A testing set is created to test the model created in the previous stage.

---

**Model deployment**

**Feedback**

---

The following are the steps involved in the present study:

a) Deciding the topics for the study, by the teacher
b) Exploring the sufficient theory required for simulation, by the teacher
c) Teaching student teams without using the simulation results
d) Taking up Assessment 1
e) Explaining the same concepts using the python simulation results
f) Taking up Assessment 2

Table 2. Details of the data set collected

| Serial Number | Data collected | Size of the data set | Type of the data collected |
|---------------|----------------|----------------------|----------------------------|
| 1             | Time taken by each student to answer the questions asked under the assessment section | 120 | Numerical |
| 2             | Responses to the final poll | 120 | Categorical |

a) Deciding the topics for the study:

For the current research, we chose the following topics for which teaching-learning is infeasible: Pressure-Altitude relation (Barometric formula), radioactive decay, fluid flow, and rocket motion. The source articles for the Mathematics behind these topics have been mentioned in table 3. In the below table, for serial number 1, \( P \) represents the pressure at a given altitude, \( P_0 \) is the average sea-level atmospheric pressure, \( M \) denotes the molar mass of the atmospheric air, \( h \) is the height above the sea level, \( R \) is the Universal gas constant, \( g \) represents the gravitational acceleration. For the case of radioactive decay, \( N \) represents the amount of radioactive material, \( N_0 \) is the initial amount of the radioactive material, \( \lambda \) is the reciprocal of average lifetime and \( t \) is the time parameter. Under serial number 3, \( z \) indicates the height of the fluid above the opening in the container, \( g \) represents the gravitational acceleration, \( R \) is the radius of the base of the cylindrical container, \( a \) is the radius of the opening, \( T \) represents the time at which the fluid level decreases to zero, \( H \) is the initial height of the fluid and \( t \) is the time parameter. And finally, for the case of rocket motion, \( m \) represents the mass of the rocket, \( v \) denotes the initial velocity of the rocket, \( u \) is the velocity of the rocket at time \( t \), \( m_0 \) is the initial mass of the rocket and \( m_1 \) is the final mass of the rocket.

Table 3. List of references for the Mathematical concepts referred for the present work

| Serial Number | Title of the topic | Differential equation | Mathematical model | Reference |
|---------------|--------------------|-----------------------|--------------------|----------|
| 1             | Pressure-Altitude relation (Barometric formula) | \( \frac{dP}{dh} = -\frac{Mg}{RT}P \) | \( P = P_0 e^{-\frac{Mg}{RT}h} \) | [10] |
| 2             | Radioactive decay | \( \frac{dN}{dt} = -\lambda N \) | \( N = N_0 e^{-\lambda t} \) | [11] |
| 3             | Fluid flow | \( \frac{dz}{dt} = -\frac{a^2}{R^2}(2gz)^{1/2} \) | \( T = \frac{R^2}{a^2} \left( \frac{2H}{g} \right)^{1/2} \) | [12] |
| 4             | Rocket motion | \( \frac{m}{dt} = u \frac{dm}{dt} \) | \( v = u \ln \left( \frac{m_0}{m_1} \right) \) | [13] |
b) **Exploring the sufficient theory required for simulation:**

It is seen that all the topics selected are governed by simple first-order ordinary differential equations. Physics teachers are aware of this form of differential equations and also a simple mathematical technique is required to solve these elementary equations. As high school mathematics includes only the basics of calculus, the methods of solving a differential equation are out of coverage for a high school student. Thus, using python, a simple program to solve these differential equations is developed and the same has been altered to all cases accordingly. The simple python program, incorporating Matplotlib plotting library and its numerical mathematics extension NumPy, to solve first-order differential equations has been taught to students along with the extensive advantages of the above-mentioned libraries.

```python
# Dependence of pressure with altitude
import numpy as np
from scipy.integrate import odeint
import matplotlib.pyplot as plt

# function that returns dp/dy
def model(p, h):
    M=0.02896
    g=9.807
    R=8.3143
    T=288.15
    dpdh = -((M*g)/(R*T))*p
    return dpdh

# initial condition
p0 = 101.325

# time points
h = np.linspace(0,120000)

# solve ODE
p = odeint(model,p0,h)

# plot results
plt.plot(h,p,lw=2,linestyle='--')
plt.xlabel('Altitude(m)')
plt.ylabel('Pressure(kPa)')
plt.show()
```

![Figure 2. Sample Jupyter Notebook](image-url)
Figure 3. Pressure vs Altitude

Figure 4. Amount of radioactive particles vs time in case of iodine

Figure 5. Amount of radioactive particles vs time for Uranium

Figure 6. The velocity of rocket vs time

Figure 7. Height of fluid vs time of flow
c) **Teaching student teams without using the simulation results**
Above mentioned topics were taught in class without any graphs, using only the standard classic theory available in textbooks of high school physics. Differential equations governing pressure altitude relation (Barometric formula), radioactive decay, mass-velocity relation in rocket science, and flow rate studies for fluid were explained in the class theoretically, as to how a normal class would run. As a high school student, we would recommend the student to remember only the final solution of these differential equations (as mentioned in any standard textbooks).

d) **Taking up Assessment 1**
The following questions were mentioned in the assessment sheet:
- Assuming the initial pressure to be 101.325 kPa, acceleration due to gravity to be 9.807 m/s², molar mass to be 0.02896 kg/mol, and the universal gas constant to be 8.3143 Nm/ (mol K), find the pressure at an altitude of 20000 m at 288.15 K.
- 80 grams of radioactive iodine was taken for study. The sample has a half-life period of 8 days. Find, how many grams of iodine would be left after 20 days has elapsed?
- Assuming the ideal rocket equation to be \( v = u \ln\left(\frac{m_0}{m}\right) \), find the velocity of the rocket after 40 seconds. Given \( u=3000 \) m/s, \( m_0=5000 \) kg and the fuel burn rate is 100 kg/s.
- Find the time at which the level of the fluid drops to 0.05 m above the vessel opening. Data: acceleration due to gravity to be 9.807 m/s², the base radius of the cylindrical vessel is 10 cm, and the radius of the opening on the vessel is 1 cm.

e) **Explaining the same concepts using the python simulation results**
Students were taught to use Jupyter Notebook, an open-source web application hosted by Anaconda, the world’s most popular data science platform. Basics of the Jupyter Notebook would be sufficient to run the python programs to solve the equations. After assessment 1, all the mentioned concepts were explained again using the results obtained using the python program. The following graphs were the results of the simulation of the four topics of our interest. Figure 2 shows the sample Notebook page with a python program run on it.

f) **Taking up Assessment 2**
The following questions were mentioned in assessment sheet 2 (the above graphs of figures 3-7 can be used to solve the following):
- Assuming the initial pressure to be 101.325 kPa, acceleration due to gravity to be 9.807 m/s², molar mass to be 0.02896 kg/mol, and the universal gas constant to be 8.3143 Nm/ (mol K), find the pressure at an altitude of 10 km at 288.15 K.
- Why do mountaineers usually suffer from nose-bleeding at high altitudes?
- 80 grams of radioactive iodine was taken for study. The sample has a half-life period of 8 days. Find, how many grams of iodine would be left after 30 days has elapsed?
- Assuming the ideal rocket equation to be \( v = u \ln\left(\frac{m_0}{m}\right) \), find the rate of change of velocity of the rocket. Given \( u=3000 \) m/s, \( m_0=5000 \) kg and the fuel burn rate is 100 kg/s.
- Find the height of the fluid after 7 seconds. Data: acceleration due to gravity to be 9.807 m/s², the base radius of the cylindrical vessel is 10 cm, and the radius of the opening on the vessel is 1 cm.
4. Analysis
Results of both the assessments exhibited a huge difference in the context of the presentation of answers as well as in the time taken to solve each problem. The time taken for a team to complete was an average of 42 minutes; but when the students attended assessment 2, they utilized only 8 minutes to answer all the questions of assessment 2; sample assessment answers can be found in figures 8 and 9.

The exact answer for question 1 under assessment 2 was 30.965 kPa. Also, students were able to see the pattern in which the change is happening in each of the cases. In the case of the radioactivity study, one of the students (student 1) of a team beautifully put forth a result just by analyzing the graph. His observation was “It is found that the half-life of iodine was less, actually negligible when compared to Uranium. This clearly shows that a longer half-life means more gradual decay of the material. It will emit fewer radioactive particles per minute, due to the same mass of radioactive material. Hence, anybody near the substance receives fewer toxic chemicals per minute. The lower dose would have less impact on wellbeing”. We were astonished to hear this line of inference from the student. Also, students were able to relate the bentness of each of these curves to the concept of the slope they learned in their mathematics class. Using this knowledge, they were able to answer question 4 of assessment 2.

Student 2 of another team expressed his views about figure 6 as “Graph for rocket motion clearly shows that as a rocket becomes lighter due to the loss of its mass on track, its velocity is raised and the curve has a reflected L- shape. Raise in velocity happens as a smooth curve.”

Student 3: “Although these graphs shouldn’t be compared with each other as they, individually, represent a unique process, I would like to put forth my views about these graphs.
Graphs 3, 4, 5 and 7 are decreasing graphs. But graph 6 is much straight when related to others, meaning that the fluid flow is faster.”

![Performance analysis for Assessment 1 & 2](image)

**Figure 10.** Comparison chart for the various assessments

These observations made by students made us realize the importance of the graphical results in caricaturing the phenomena under study. Figure 10 is the bar chart diagram which can be used for the data visualization of the data set collected. These bar charts display the average time taken by the students to answer the questions of assessment 1 and 2. For every topic under study, we can see that the blue bars representing the average time taken for answering, when the concept was taught theoretically, is extremely more when compared to the height of orange bars. This depicts the fact that the students were able to answer assessment 2 consuming very less time. Also, the responses made by the students as mentioned in the previous paragraph project the participation of students enthusiastically. Simulatory learning is thought-provoking when compared to the existing traditional theoretical methods of teaching-learning. The concepts chosen for our study cannot be studied practically in the laboratory at a high school level. The theoretical learning of these concepts is discrete but the simulatory results can aid students in a continuous understanding of the phenomena. Students can become proficient in the decipherment of the pattern following in each case. Additionally, students can study the phenomena at any intermediate time using the simulatory results, which is not easily possible by theoretical studies [14].

Finally, after the analysis, there was a poll on the efficacy of learning using the python programming language. Poll results were represented below as a pie chart, figure 11. As seen, 89% of the students preferred python whereas the rest 11% preferred to learn theoretically.
As a concluding remark, we would like to mention that our research does not go against the practical learning of any subject but computational learning is a safer way of learning. Where practical learning takes a back seat, computational learning comes forward to make the learning efficient.

![Figure 11. Survey results](image)

5. Conclusions
Teaching and learning processes define the world. There are different methods involved in teaching-learning processes. In the present work, we attempted to compare the traditional theoretical learning methods with the simulatory method of learning. In this regard, we conducted sessions for the high school students to explore the impact made on the students by the theoretical and simulatory modes of teaching. Data about the assessment results and survey results were collected. Several science topics were taught to the students theoretically as well as through the simulation results obtained using the python programming language. Two sets of assessments were taken up by the participants and the results obtained were recorded. The data set constructed was analyzed visually using comparison charts in the form of bar graphs and pie charts. The results suggested that, in the case of impractical topics of science, the simulation of the concepts would be more advantageous in understanding the concepts undoubtedly. Hence the data science methodology employed for the present investigation was efficient to conclude that the simulation results were more approaching when compared to the normal theoretical ways.

Ethical Statement
We confirm that the present research has been carried out following the principles outlined in the IOP Science ethical policy. A prior consent had been taken from the Head of the Institute to take up our research. Informed consent to participate in the study was obtained from all the participants. Also, the research work has been approved ethically by our Institutional Research Committee with the approval ID 2020-SEP-BS-001.

6. References
[1] Garrison, D. Randy; Archer W 2000 A Transactional Perspective on Teaching and Learning: A Framework for Adult and Higher Education. Advances in Learning and Instruction Series.
[2] Martín-Blas T and Serrano-Fernández A 2009 The role of new technologies in the learning process: Moodle as a teaching tool in Physics Comput. Educ. 52 35–44
[3] Ballou D and Springer M G 2015 Using Student Test Scores to Measure Teacher Performance *Educ. Res.* **44** 77–86

[4] Treagust D F and Duit R 2008 Conceptual change: a discussion of theoretical, methodological and practical challenges for science education *Cult. Stud. Sci. Educ.* **3** 297–328

[5] Chandos V, Davies G, Edmonds D, Ivens M, Pomeroy B, Vadera S and Rossiter A 2007 *Practice Makes Perfect: The Importance of Practical Learning* 1 The Social Market Foundation Members of the Board

[6] Lateef F 2010 Simulation-based learning: Just like the real thing *J. Emergencies, Trauma Shock* **3** 348–52

[7] Luo W, Pelletier J, Duffin K, Ormand C, Hung W C, Shernoff D J, Zhai X, Iverson E, Whalley K, Gallaher C and Furness W 2016 Advantages of computer simulation in enhancing students’ learning about landform evolution: A case study using the grand canyon *J. Geosci. Educ.* **64** 60–73

[8] Bilotta F F, Werner S M, Bergese S D and Rosa G 2013 Impact and implementation of simulation-based training for safety *Sci. World J.* **2013**

[9] Rollins J B 2015 Foundational Methodology for Data Science A 10-stage data science methodology that spans technologies and approaches *IBM Anal.*

[10] Lente G and Ősz K 2020 Barometric formulas: various derivations and comparisons to environmentally relevant observations *ChemTexts* **6**

[11] Sujaul Chowdhury P K Das 2018 Radioactive decay: numerical solution of differential equations using Euler and second order Runge–Kutta methods using Mathematica - Book chapter - IOPscience *Numerical Solutions of Initial Value Problems Using Mathematica* (Morgan & Claypool Publishers) pp 1–8

[12] Geeter C 2018 *One-Dimensional Unsteady Flow From A Cylindrical Draining Tank* vol 1

[13] Ganji D D, Gorji M, Hatami M, Hasanpour A and Khademzadeh N 2013 Propulsion and launching analysis of variable-mass rockets by analytical methods *Propuls. Power Res.* **2** 225–33

[14] Katsaliaki K and Mustafee N 2011 Applications of simulation within the healthcare context *J. Oper. Res. Soc.* **62** 1431–51