Community College student research project pedagogy delivered online during COVID-19 and assessment of scientific principle learning

Dr. sunil Dehipawala, City University of New York, Queensborough Community College

Sunil Dehipawala received his B.S. degree from University of Peradeniya in Sri Lanka and Ph.D from City University of New York. Currently, he is working as a faculty member at Queensborough Community College of CUNY.

Dr. Dimitrios S. Kokkinos, City University of New York, Queensborough Community College

Dr. Dimitrios Kokkinos is an Associate Professor of Physics at Queensborough Community College of CUNY since 2017. He completed his Electrical Engineering degrees (BE, ME, PhD) at CUNY and undergraduate in Physics in Europe. He worked in industry for AT&T Bell Labs and Verizon Communications for 23 years as a telecommunications engineer specialized in fiber optical system research and development. He is teaching pre-Engineering Physics courses, conducts research in fiber sensors and mentors student research projects.

Prof. Vazgen Shekoyan

Dr. Vazgen Shekoyan is a professor of physics and his experiences include pedagogy, CubeSat, etc.

Dr. Rex Taibu

Dr. Rex Taibu has taught studio physics classes for several years. His teaching experience has shaped his research focus. Currently, Dr. Taibu is actively engaged in
1) promoting scientific inquiry attitudes in students through designing, implementing, and assessing innovative inquiry based physics labs.
2) conducting research regarding the role of language in conceptual understanding.
3) exploring cosmic rays (detection, data collection, and analysis).

Mr. George Tremberger Jr

George Tremberger serves as Lecturer in City University of New York Queensborough Community College Physics Department and his interests include pedagogy and astronomy.

Prof. Tak Cheung

Tak Cheung, Ph.D., professor of physics, teaches in CUNY Queensborough Community College. He also conducts research and mentors student research projects.

©American Society for Engineering Education, 2020
Community College student research project pedagogy delivered online during COVID-19 and assessment of scientific-principle learning

CUNY Queensborough Community College Physics Department

Introduction

The COVID-19 lockdown had induced new pedagogy. Around August 2020, Dean of Engineering at McMaster University, Dr. Ishwar K. Puri, called on faculty to create short learning modules to be put into a virtual depository and free up time for faculty to mentor challenge-based learning instead [1]. Meanwhile ASEE published a paper on August 2020 stating that the use of worked-example videos for blended learning is an efficient pedagogy [2]. When the lockdown continues onto Spring 2021, there would be a sizable collection of worked-example videos in virtual depository and faculty mentoring challenge-based learning would become reality. Recently, a PER (Physics Education Research) publication found that challenge-based learning was successful to increase the scientific literacy in the studied students [3].

Research projects using data available on the web would be suitable to be delivered online from professors to community college students. There have been articles describing the negative impact of COVID-19 lockdown on scientist parents and academic mothers in science careers [4, 5]. Although the reports focused on professionals, the negative impact of COVID-19 lockdown on engineering students with care giver responsibilities is expected to be similar. A pedagogy for online delivery of research projects must be flexible so as not to exclude student care giver participations, and yet should not demand excessive faculty instruction time. These two requirements would necessitate a pedagogy adjustment from the face-to-face situation.

Our community college is in New York City. Our community college has a diverse student population speaking more than 100 different mother tongues, summarized in the Queensborough Community College Institutional Research Fact Book. The range of student academic levels spans from those who transfer to top R1 universities to those who need remediation. When a student expresses interest in learning from doing a research project, a faculty offers some selections. For example, some of our technology students are interested in the construction of research equipment and a faculty offers some selections with the encouragement of “transferring to engineering majors”. Some of our students are interested in AI and statistical machine learning software and that their related projects focus on algorithm development with lab and/or simulated data. In the face-to-face situation prior to the COVID-19 lockdown, the data source was either data collection in a lab setting or data in a national archive. During the lockdown, using the data in national archives became the only choice in a community college with limited budget. Our experience showed that community college students in general would need face-to-face instructions on how to manage data pipelines. Our online delivery pedagogy found that data curation was necessary during the COVID-19 stringent lockdown rules in New York City. Note that the University of Illinois School of Information Sciences explained data curation as “Active
and on-going management of data through its lifecycle of interest and usefulness to scholarship, science, and education, ….. a critical step in scientific data digitization, sharing, integration and use …” (https://ischool.illinois.edu/research/areas/data-curation).

One of the most important elements in the learning of a physics related project must be an understanding of the engineering principles related to the data collection. The National Academy of Engineering, through its NAE Committee on the Diversity of the Engineering Workforce, sponsors an Enginergrl.org website that explains engineering principles [6]. When the "KISS (keep it simple stupid)" principle conflicts with the "keep the target user in mind" principle, an engineer makes a choice. The data curation process retains the engineering principles to prepare those community college students interested in engineering.

Implementation

Data curation examples on spatial data in stochastic calculus Fokker Planck equation, Solar Dynamics Observatory data, T-cell spatial-temporal data, exoplanet archive data, and Hubble quasar data are presented.

The stochastic calculus Fokker Planck equation could be applied whenever there is a histogram built on a large set of outcome data at one specific instant. Applications include the studies of electrical circuit noise [7], electrochemical noise [8], Langevin dynamics and free energy relationship in biological systems [9], etc. During lockdown, laboratory generated noise or fluctuation experiments would be difficult. Thus, the use of online fluctuation data for Fokker-Planck equation application studies would be acceptable. A large set of outcome data usually would display a Gaussian-like distribution. A Fokker-Planck approach in tracking the time evolution of the distribution could be applied. In mathematics, the Ornstein–Uhlenbeck stochastic process could be captured by the Fokker-Planck equation in terms of the underlying probability density function. A Solar Dynamics Observatory brightness data example was published by us in the 2014 ASEE Proceedings [10]. The solar bright area probability density function was modeled using a Fokker Planck equation. The reported data curation procedure was found to be suitable for online delivery during the COVID-19 lockdown. The open access of a Python Fokker Planck numerical solver from MIT would also be helpful for those student participants knowing the Python language and treating the numerical solver as a tool [11].

A 2019 report on the recruiting of students into engineering projects suggested that the telling of engineering disaster stories would be an effective recruitment method [12]. The study of space weather study has been gaining attention with renewed interest for astronaut missions. Space weather had affected the US Navy blockage mines along the coastal waters of Vietnam in 1972. The 1972 disaster due to space weather was reported in 2018. The telling of the 1972 disaster at the Scientific America reading level [13] and journal reading level [14] would generate affective learning. The August 2020 study that linked space weather to the Titanic event would provide another disaster story material for recruiting students to engineering projects [15, 16]. As Wylie explained [12], listening to “disaster stories” could induce self-deprecation, humility, teamwork and mutual learning.
Just like the Titanic event, a cancer event is among the category of “disaster stories”. A discussion of immune cells in relationship to cancer cells would generate interest among the student participants. The T-cell motion captured in a microscope is an important element in the study of the immune system. A data curation procedure was used on the data contained in a Youtube T-Cell video posted by Cambridge University, based on an Immunity Journal report [17]. The boundary interface between a cancer cell and CD8+ Killer T-cell (or Natural Killer NK cell) in a microscope image would carry information on membrane elasticity related to actin activity. In addition, the brightness fluctuations in the cells would carry information regarding the cell-to-cell communication. A student participant was asked to use the ImageJ software to extract position data from 52-sec frame (shown in Figure 1) to 61-sec frame, and to study the torque coupling interaction of the killer T-cell and the cancer cell. Regular T-cell motility videos have been posted as well. A 2020 August PNAS publication contained several regular T-cell motility MP4 videos [18]. An investigation of the apparent random walk would be an interesting project. The acellular slime mold Physarum Polycephalum motion has been videotaped, and an application of Langevin dynamics would explore the feedback mechanism beyond the reported area growth as a function of time [19].

![Killer T Cell: The Cancer Assassin video (Cambridge University).](https://www.youtube.com/watch?v=ntk8XsxVDi0)

Besides the “disaster stories” related projects, scientific projects relying on advanced engineering are also important. The data curation procedures of exoplanet archive data and Hubble quasars data are presented in Appendix A and Appendix B respectively.

A display of a graphical representation instead of a data table would be an elementary data visualization method. An animation of a projectile’s range and height in introductory physics would be a data visualization method, in addition to a graphing of the calculated trajectory. On the one hand, the representation of a simulation result in a video format is a common technique. On the other hand, an explanation of a video could be a showing of the numeric information.
Our PER (Physics Education Research) experience showed that the illustration of a wave equation can be facilitated by a reverse numeric examination of the Colorado University PHET visual simulation on waves [20]. The spatial temporal simulation represented in the visual display could be digitized for numeric information. The second partial derivatives on the spatial variation and temporal variation would give the wave speed value, consistent with the selected frequency and wavelength displayed in the PHET simulation. From the viewpoint of PER, the meaning of partial derivative would be fully explored when taking a spatial snapshot at one instance in time, and/or taking time data at a single fixed location. Using an experiential learning experience to discover the wave equation is called discovery learning in the PER pedagogy.

The viewing of a video produced by simulation is an elementary data visualization pedagogy. The next step of the extraction of the science in a numeric exercise with graphical information would produce a deeper data visualization pedagogy for the learning of the equations. An Excel example of using the Illustris TNG cosmology simulation data would be useful to strengthen the experiential learning of data visualization during the COVID-19 lockdown [21]. The studied MP4 video was on a simulation with an explanation that said “Formation of a massive late-type, star-forming disk galaxy: an individual system is tracked through time, and its stellar light is shown in a 500 kilo-parsec region. This galaxy exhibits rapid ongoing star formation in an extended, clumpy disk, until it experiences a late-time merger (at approximately $z \sim 0.3$) with a nearly equal mass companion”. The simulation video link was at https://www.tng-project.org/movies/tng/tng100_sb0_late-type_formation.mp4

![Figure 2 The labelling of Galaxy-1 and Galaxy-2 in the studied TNG simulation.](image)

The labeling of the two galaxies in the simulation video is shown in Figure 2. The reading of the positions of a galaxy in the video frames generated the position versus time data table. The plotting of the data table as a trajectory graph, shown in Figure 3, would be acceptable for an illustration of data visualization beyond video viewing. The graphical numeric information was the basic result of the TNG simulation in which the gravitational attraction between the two galaxies was shown. A novice would appreciate the video viewing while an expert would prefer the graphical numeric information. The simulation showed that the two galaxies came closer with an apparent moving center of mass, following apparent elliptical paths, which would not be obvious from the video viewing.
Figure 3: Galaxy-1 and Galaxy-2 simulated trajectories with 16 position-time pairs, black arrow indicating time direction.

When a studied video was not analyzed for each individual frame, the extracted numeric information could be interpolated, shown in Figure 4. For a simple projectile trajectory in introductory physics, the interpolation would be validated by using the Taylor series expansion. The showing of interpolation would deepen the understanding of data visualization and experiential learning of data processing. Data interpolation also has a direct application in our Brookhaven National Lab Synchrotron EXAFS data analysis, which requires background subtraction. The interpolation formula at the Excel level has been posted on a microwave engineering website [22]. For the showing of data visualization to student participants, an analysis of every 3-frames or 5-frames would be able to demonstrate the essence of the gravitational attraction with a center of mass in a many-body problem. The present paper is not saying that the reported TNG simulation video did not carry a calculation for every video frame. The use of interpolation as an introduction to data visualization and experiential learning of data processing was found to be a practical approach. The deep learning in terms of post simulation-animation debrief pedagogy could be delivered online using the conversion of video to graph process [23].
Figure 4: Galaxy-1 and Galaxy-2 simulated trajectories with data interpolation (arrow indicating time direction)

The online delivery of analysis techniques such as nonlinear data fitting, pairwise comparison using data carrying several significant figures, Bayesian statistics, etc. can be adapted from the face-to-face delivery mode in a straightforward manner when given ample preparation time. The non-intuitive Bayesian inference is an important technique. Bayesian inference has been used to analyze 1-D and 2-D motions in biomechanics [24, 25], as well as star motion [26]. In particular, the Youtube video on Bayesian statistics of infection rate has been a popular topic among STEM students during the COVID-19 pandemic [27]. The Youtube video explained the Bayesian inference equation and offered a numeric example in a disease rate scenario. Consider the following two pieces of information. The disease rate in the population is 0.1 %. If a person has that disease, the test is 99% accurate. What is the actual probability that a person tested positive has the disease? The answer of 99% based on the test accuracy is wrong.
Using the Bayesian equation illustrated in Figure 5, the correct numeric answer would be 
\( (0.99 \times 0.001) \text{ divided by } (0.001 \times 0.99 + 0.999 \times 0.01) \), which equals 0.0902, 9% after being tested positive. We can reconsider the above example with an addition information. A person was tested positive using two independent laboratories. What is the actual probability? The correct numeric answer would be 
\( (0.99 \times 0.0902) \text{ divided by } (0.0902 \times 0.99 + 0.9098 \times 0.01) \), which equals 0.907, 91% after being tested positive twice; which is still lower than the accuracy of the test at 99%.

**Assessment Designs**

An assessment rubric for engineering education could be designed with guidance from standard experiential learning assessment. The importance of scientific principle understanding in engineering education, described in the 2000 National Academy of Engineering Founder Award given to Townes and the 2019 National Academy of Engineering Gordon Prize Innovation in Engineering and Technology Education given to Benkeser, should also be included in an assessment rubric [28, 29].

The assessment consists of the three deliverables developed by McGill University experiential learning team in terms of content-process mixture, big picture perspective, and reflection [30], and an additional deliverable on scientific data resolution related to engineering and technology. An assessment rubric example for experiential learning is listed in Table 1.
Table 1: Assessment rubric example for experiential learning of solar brightness project

<table>
<thead>
<tr>
<th>Deliverable</th>
<th>Highly competent</th>
<th>Competent</th>
<th>Needs Improvement</th>
</tr>
</thead>
<tbody>
<tr>
<td>Content theory (20%)</td>
<td>Provided clear writing of the origin of the fluctuation in solar brightness</td>
<td>Contained one mistake.</td>
<td>Contained two or more mistakes.</td>
</tr>
<tr>
<td>Process activity (20%)</td>
<td>Completed data curation resulting in text format and conducted brightness pairwise comparison</td>
<td>Contained one mistake</td>
<td>Contained two or more mistakes</td>
</tr>
<tr>
<td>Big picture real world perspective (20%)</td>
<td>Provided two applications of pairwise comparison analysis used in engineering projects</td>
<td>Contained one mistake</td>
<td>Contained two or more mistakes</td>
</tr>
<tr>
<td>Reflection transferrable knowledge (20%)</td>
<td>Provided correct justification for pairwise comparison analysis from other projects</td>
<td>Contained one mistake</td>
<td>Contained two or more mistakes</td>
</tr>
<tr>
<td>Scientific resolution of instrumentation (20%)</td>
<td>Provided an assessment of the scientific resolution of the pairwise comparison analysis with the original resolution of the instrumentation data</td>
<td>Contained one mistake</td>
<td>Contained two or more mistakes</td>
</tr>
</tbody>
</table>

Discussion

The three deliverables of content-process mixture, big picture perspective, and reflection must be built with some data experience. In a lockdown situation, a theoretical discussion of data collection would not be enough for experiential learning, in our opinion.

On the one hand, data curation offers a solid foundation for student participants to land back on reality after the excitement in the recruiting process [12]. Despite the interest in reading the background literature, a research project has little meaning until it has data. Prior to the lockdown, taking data in a lab was the standard expectation. During the COVID-19 lockdown, data curation of public data is an efficient solution for community college student precipitants to
proceed to the analysis stage. Not all data would require curation. Synchrotron EXAFS data, Protein Data Bank (PDB) data, Genbank data, etc. are already in text format so data procedure is straight forward without a requirement on new pedagogy. Bear in mind though that a PDB dataset could be 1,000 files in a study of protein interaction. A recent report of genetic correlation with cortical thickness shrinkage in healthy subjects had recruited 23,000 individuals [31].

On the other hand, the new trend of using AI software is gaining attention among community college students, but they may have minimal notion of what the AI software entails. The sigmoid function neural network [32], radial basis function neural network [33] and Support Vector Machine (SVM) algorithm in Machine Learning [34] could be used for data analysis of photon pulse in our muon detection project [35]. Since some student participants may not be familiar with Matlab, Python, etc., Excel could be used. We found that the Youtube video on Excel neural network posted on Feb 2020 was useful as an introduction [36].

The AI data visualization software application still belongs to the “work in progress” category in our experience, although deep learning in cross-class of physics and computer science is straight forward to implement in the undertaking of a research project [37]. Commercial software, such as Neural Designer (Trademark), on data classification or pattern recognition using neural network algorithm can be purchased. However, the education value would be minimal when student participant usage is limited to the software operation mode.

On the one hand, writing a neural network in Excel is more educational in challenge-based learning pedagogy, even though the research value would not be enough for scientific publication due to the Excel limitation. On the other hand, Excel neural network add-in Python/VBA packages are also available [38, 39]. However, we found that the use of such Excel add-in packages was not enough to enable student participants to grasp an understanding of neural network applications in exoplanet search study [40], gravitational lensing study [41], cosmology three-body-problem study [42], protein structure stability study [43], computer science class protein orientation student project [44], population genetic inference study [45], genomic single nucleotide polymorphism study [46], etc.

When ABET accepts the online learning delivery mode after the COVID19 lockdown, then faculty would have more time to mentor challenge-based learning. The AI data visualization software approach would be feasible for the student participants in a community college in the context of the PER pedagogy. A team of student participants with different majors would enable the application of AI software in a research project. The pre-engineering majors could focus on the application objectives, the computer science majors could focus on Matlab, Python, etc. for neural network implementation. However, such a collaboration between physics department pre-engineering students, technology department students and computer science department students would not be easy to organize during the sudden COVID19 lockdown in a community college.

When the lockdown duration continues beyond Spring 2021, the research project delivery will continue online without data collection in a community college lab. A recent report showed that the improvement of grit is more related to goal commitment when compared to growth mindset
A “discussion-only” theoretical data collection in a research project during lockdown would be weak to encourage goal commitment. The availability of data curation pedagogy would be consistent with the aim of grit improvement for research students in a community college facing a necessary continued lockdown.

The weakness of the present report includes small sample size and limited data curation examples. The massive deployment of worked-example videos in virtual depository will free up faculty time to mentor challenge-based learning, which includes the undertaking of research projects. A community college instructor guiding a student participant in a research project remotely would need to decide on the KISS engineering principle versus the "keep the target user in mind" engineering principle [6], when given the choice of using data curation with simple analysis versus text format data with advanced analysis.

Conclusions

The present paper has provided a list of data curation examples suitable for the COVID19 lockdown online delivery of research projects. The paper discussed the aim of developing challenge-based learning pedagogy after the COVID19 lockdown, when worked-example videos in virtual depository becomes a norm in engineering education. When projects do not require data curation, there are more time for the student participants to focus on advanced analysis methods. The paper provided an assessment rubric example of experiential learning; and discussed the neural network applications for projects with data already in text format.

Acknowledgments

The authors thank Mr. Alexei Kisselev of CUNY-QCC for computer support. The authors thanks Dr. Eric Cheung of Montefiore Mount Vernon Hospital for discussion on medical science. The authors thank researchers sharing their data on the open web. The authors thank the three anonymous reviewers of the 2020 Fall ASEE Mid-Atlantic Conference hosted by Stevens Institute of Technology.

Appendix A

Exoplanet project data curation
Go to https://exoplanetarchive.ipac.caltech.edu/
Then select TESS Project Candidates
Then select the first entry, for example, TOI 10000.01
For exoplanet time series data in K2 targets
https://exoplanetarchive.ipac.caltech.edu/docs/data.html
https://exoplanetarchive.ipac.caltech.edu/cgi-bin/TblSearch/nph-tblSearchInit?app=ExoTibs&config=k2targets
Download time series data in text read by browser, copy-paste onto Excel
Then download the files at the bottom of the webpage. A student participant can use the NASA Fv-tool to read the FITS format and export to Excel, described in our 2014 ASEE presentation [10]. The NSAS FITS Liberator (https://www.spacetelescope.org/projects/fits_liberator/) and Harvard SAOImage DS9 (https://sites.google.com/cfa.harvard.edu/saoimageds9/home) can also read FITS format. Matlab, Python, etc. also support FITS format.

For Exoplanet archive data
https://exoplanetarchive.ipac.caltech.edu/docs/KELT.html
select interactive table
https://exoplanetarchive.ipac.caltech.edu/cgi-bin/TblView/nph-tblView?app=ETSSView&config=conf/KELTP
Appendix B

Hubble quasar data curation
Go to https://hla.stsci.edu/hlaview.html
Type in the object name, for example WFI 2033-4723, a gravitationally lensed quasar
https://hubblesite.org/contents/media/images/2020/05/4613-Image
Then download the FITS files and use the NASA-Fv tool to read and export to Excel.
References


Colorado University PHET simulation on waves. Last accessed Sep 21 2020 https://phet.colorado.edu/en/simulation/wave-on-a-string

The Illuvstri TNG project media https://www.tng-project.org/media/


Post-Simulation Debrief: Post-simulation discussion with students leads to deeper learning. The instructor should: Prepare question to ask during the debrief to ensure students see alignment between the simulation and the course goals. https://www.kent.edu/ctl/simulation-teaching-strategy


Derek Alexander Muller. YouTube channel Veritasium The Bayesian Trap, April 5 2017. https://www.youtube.com/watch?v=R13BD8qKcTg

National Academy of Engineering. 2019 Bernard M. Gordon Prize for Innovation in Engineering and Technology Education. For fusing problem-driven engineering education with learning science principles to create a pioneering program that develops leaders in biomedical engineering. https://www.nae.edu/203085/Dr-Paul-J-Benkeser


https://iopscience.iop.org/article/10.3847/1538-3881/ab0e12


https://academic.oup.com/mnras/article-abstract/494/2/2465/5823775


Sarah Gurev, Nidhi Manoj, Kaylie Zhu. Predicting Correctness of Protein Complex Binding Orientations. Stanford University CS229 class project 2018.

