



NATIONAL CONFERENCE ON SCIENCE EDUCATION

ATLANTA 23

MARCH 22-25



**SAVE WITH
EARLYBIRD
PRICING**

**DEADLINE:
JANUARY 27**

NSTA / Connected Science Learning / Connected Science Learning November–December 2022 /

Data Detectives Clubs: A Collaborative Approach to Data Science Through Epidemiology

EMERGING CONNECTIONS

Data Detectives Clubs: A Collaborative Approach to Data Science Through Epidemiology

[Connected Science Learning November–December 2022 \(Volume 4, Issue 6\)](#)

By Laura Martin, Janice Mokros, Nav Deol-Johnson, Pendred Noyce, and Jacob Sagrans



As the COVID pandemic highlighted a stark need for better scientific and data literacy, we launched the afterschool Data Detectives Clubs as a way to immerse 10- to 14-year-olds in authentic COVID data. The 15- to 20-hour Data Detectives Clubs are structured around an adventure novel, *The Case of the COVID Crisis*, which introduces readers to epidemics across space and time. Each Club session features a chapter that is accompanied by discussion, activities, and exploration of infection and vaccine data using CODAP, the Common Online Data Analysis Platform (Concord Consortium 2020). In addition to learning about data and data representation, youth become engaged in the science and begin to identify with the work of epidemiologists (PEAR 2022).

The activities are conducted in an informal environment. Strong STEM programs after school can feed into regular school time activities because they have an impact on students' engagement and career interest, identity, attitudes, and skills (Hull et al. 2022). In addition, afterschool settings provide an important route to reaching many underrepresented youth through such programs as 4-H, Boys and Girls Clubs, YMCA, and Girls Inc.

With funding from the National Science Foundation, Data Detectives Clubs represent a collaboration among several organizations that focus on underserved populations. Science Education Solutions and Tumblehome co-directed the project, Concord Consortium developed the software and computer-based activities, and Imagine Science recruited and oversaw the Club sites. All organizations meet regularly with one another and, less often, with project evaluators.

Imagine Science is a collaborative effort among four leading national youth organizations (4-H, Boys and Girls Clubs, YMCA, and Girls Inc.), formed to bridge the STEM gap by stimulating the imaginations and confidence of historically underrepresented youth (Imagine Science n.d.). Under the aegis of Imagine Science, the national network of Innovation Sites partner with one another to share curricula, training opportunities, outcomes data, and other assets. Together they seek to identify best practices and engage hard-to-reach youth. For this project, a dedicated site coordinator from Imagine Science served as a key member of the project leadership team and worked with all the Club sites, Club administrators, and Club leaders.

Program Components

The work we report addresses current concepts and technologies that are not usually taught to students in the middle school years, who are the target audience. To provide an overview of the program and its components, see this [short video](#) (Mokros et al. 2021).

The goal of the program is to help middle school youth better understand epidemiology and data analysis through reading, discussion, science activities, and use of selected technology tools. Specifically, participants develop data fluency, an understanding of how to address their own questions of data sets, and confidence in their ability to study urgent epidemiological challenges.

The Setting

Each Data Detectives Club serves between 12 and 20 youth. The project provides books and activity supplies. For part of the session, the Club leader guides the children in discussions or group activities. During the data exploration part of the session, each young person works on a computer, sometimes shared (during the period of school closures, schools provided children with computers and internet connection at home). Leaders present the data questions and walk between stations to assist. Lessons last between 60 and 90 minutes and take place over varying numbers of sessions. However, we found that at least 12 hours of exposure led to stronger positive outcomes for the youth, as measured by a final survey (PEAR 2021, 2022).

The Program Leaders Training and Guide

Because it is committed to building a community of practice for excellence in STEM programming, Imagine Science invests in leaders through professionalized training and curriculum webinars—and also (when funding allows) with stipends and materials, as it does in this case.

For Data Detectives Clubs, leaders are given six hours of professional development by project staff, which walks them through the activities and trains them on the use of CODAP. A detailed guide for

leaders outlines the concepts and objectives of the program and each chapter. The guide has

- an introduction and overview,
- detailed instructions for carrying out the activities and for using the software,
- links to podcasts and other resources including an audio version of the chapters,
- written summaries of the chapters, and
- a materials list.

The guide and training also target how to share artifacts of the youth work with the research team through screenshots and PDF files.

In principle, the Clubs follow the activities outlined in the guide. In some cases, as when adapting to local schedules, leaders vary the procedures (e.g., skipping an activity).

The Book

The focal program component is an 11-chapter children's adventure book, *The Case of the COVID Crisis* (Noyce 2021) published by Tumblehome, that touches on epidemics and pandemics throughout history, as well as the evolution of the COVID pandemic since the project's inception (Noyce 2021). Later chapters focus on vaccination and on how the COVID pandemic has changed over time. Information is presented through a fictional adventure story as opposed to factual narrative. See this [link](#) for an overview of the book and its concepts.

To encourage reading, each youth is given a book to take home and a book to use at the program site. Some leaders dedicate time for individual reading, read chapters aloud, or read summaries of chapters aloud. Not all children read the whole book (though many do) but they report enjoying the story in any case (PEAR 2022).

Podcasts featuring the main characters in the book and summarizing each chapter provide a way to review the important content (see this [link](#) for an example of a podcast for Chapter 4).

Activities

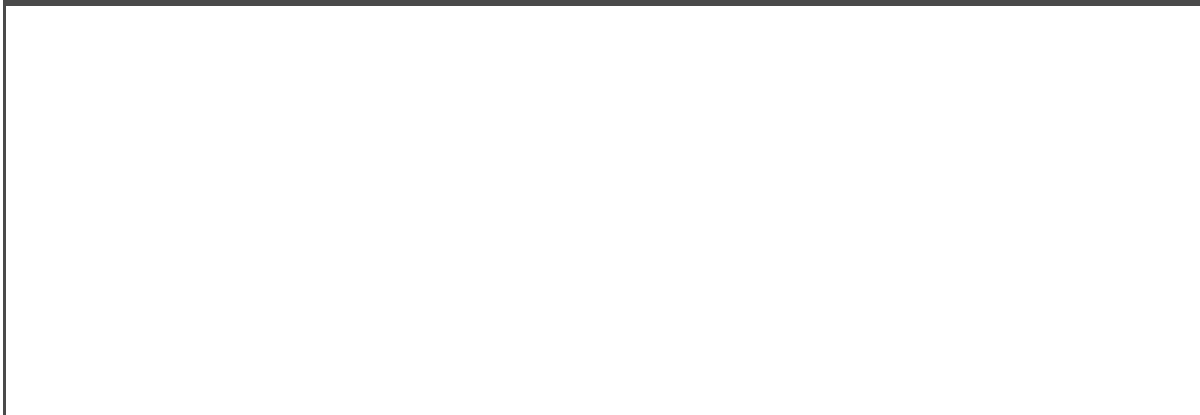
The Data Detectives Clubs activities involve regular use of data sets and data tools, hands-on activities, and discussions about pandemic-related issues. Designed by Science Education Solutions, Tumblehome, and Concord Consortium, the data activities involve CDC data sets, a graphing tool (CO-DAP; Concord Consortium n.d.) and a simulation tool (NetLogo). Many of the CDC COVID data sets we work with change on a weekly basis, with new variables (such as number/rate of vaccinations for specific age groups) added regularly. Developing data activities has required identifying basic data sets on historic pandemics and COVID, and adjusting them to be developmentally appropriate for 10- to 14-year-olds.

In addition to data and science activities, youth regularly engage in discussions on issues such as misinformation, the need for clear policies during a pandemic, and decision making on both an individual and collective level. We build on the youth development expertise of site leaders in providing safe environments for discussions of a sensitive nature.

Below are some examples of activities included in the program, with a focus on ones that have been most popular with youth:

- **The contagion game.** This game, which models the spread of an infectious disease, is one of the first activities that the youth participate in. One person gets a cup filled halfway with vinegar and the rest of the group get cups filled halfway with water. Accompanied by music, participants share liquids with the person nearest them when the music stops. After about four iterations, placing litmus paper in each person's cup indicates how many got "infected." The group is often surprised at the ease of spread.
- **Matching graphs on JamBoard.** Youth are given four electronic "sticky notes" with clues and eight graphs. The graphs show four different countries' rates of new daily COVID infections over time as well as four graphs showing cumulative cases over the same time span. The challenge is to match the country with its two corresponding graphs. An example of a "sticky note" clue is "COVID first appeared in China but it was quickly contained; new case numbers remained low for the next two years." Youth need to determine which graph presents a picture of new cases over time compatible with the clue, and then which of the cumulative infection cases graphs is compatible with the first graph.
- **NetLogo simulations embedded in CODAP.** NetLogo includes a program that simulates spread of disease. It allows the user to set starting parameters such as how many people are immune, the survival rate, how contagious the illness is (its R-naught), and how long the illness lasts. Youth set their parameters and run the simulation, which appears as little characters or "peeps" colliding on the screen. Peeps change color as they are infected or recover, and they disappear if they die. Players finish each simulation run with a final number for how many people were infected and how many remained safe. When they run the simulation with the same parameters several times, they see that contagion can be somewhat random/variable, and they do not get the same outcomes each time.

FIGURE 1. SCREENSHOT OF NETLOGO SIMULATION.





Final Projects

We experimented with different summative activities. In the earliest programs, we asked youth to design public service announcements based on population data from the CDC. Later, in addition to making the graphs, youth were asked to develop presentations to a hypothetical local council recommending whether a large crowded venue or schools should be open, based on data models. Our last iteration has youth

- create a sonification of daily COVID infection rates using kazooos to reflect changes in rates and matching them to graphs,
- work with data from two counties and make recommendations about resource allocation, and
- create and discuss a wish line to talk about their own experience (cognitive, social, and emotional) with Data Detectives work and with COVID and pandemics.

Career Connections

We emphasize career connections in the program. Most chapters of the book feature historical and current professionals working in fields as disparate as virology, vaccine development, and wastewater virus detection; at least a dozen such specialists populate the book. In addition, virtual “visits” with epidemiology workers are arranged by the project team and by individual sites so that kids can ask questions about infectious disease and related work with data. Professionals who have conducted virtual visits include an infectious disease doctor, director of state department of public health, an epidemiology modeler (who also is featured as a character in the book), and Jackson Laboratory researchers working on genetic sequencing and vaccine development. One Club leader was able to arrange in-person visits with a head chemist from the Pennsylvania Department of Health and also with a local pediatrician. We discovered that most afterschool leaders do not have enough time to arrange local in-person career visitors, so we rely primarily on a bank of professionals who can visit virtually.

Table 1 maps out the project components of the Data Detectives Clubs during 2021–22.

TABLE 1. CHAPTER TOPICS AND RELATED ACTIVITIES	

Chapter	Topics	Place(s) visited	Activities
1	Spread of measles among unvaccinated. Social distancing and other limiting factors.	USA	Computer disease spread simulation. Litmus paper activity showing contagion.
2	1918 flu. Data tracking and prediction/inference. Course of an infectious disease (rise and fall).	USA	CODAP graphing of 1918 flu. Video animation intro to viruses.
3	Spread of COVID on cruise ships. Variability and data smoothing.	Off the coast of Japan	More CODAP. Examining COVID infection trends. Role-play advice to authorities.
4	Bat-spread Nipah virus. Epidemiological detective work. Transfer between species.	Bangladesh	Virus mutation card game. Second virus video animation.
5	Smallpox history and eradication through worldwide vaccination.	Colonial US and modern Bangladesh	Exploring data on smallpox eradication. Discussing social factors affecting eradication timeline.
6	Ebola. Ring vaccination. Misinformation. Importance of public trust in government.	Congo (DRC)	Discussion about vaccines and public trust. Contact tracing activity.
7	COVID-19 in different countries. Policies, government communication, and social trust.	France and Taiwan	Compare countries' COVID-19 course of infection. Matching COVID graph to country.
8	Impact of COVID on different communities related to race/ethnicity.	Navajo Nation; USA	Compare rates of COVID among different groups to their % of the overall population.
9	How are vaccines developed/tested? What is vaccine efficacy?	DRC; USA	Return to disease spread simulation to model effect of vaccination.
10	Considering COVID vaccination priorities. COVID variants.	USA	Creating vaccination priorities. Simulation of COVID variants.
11	Wastewater monitoring. How is wastewater used to monitor COVID trends? What might the future have in store with COVID?	USA	"Dog Urine" testing activity. Wastewater COVID data exploration.

Discussions

Finally, each session has suggested discussion questions where youth can explore the personal associations they have with COVID and the content of the activities, including the book. Youth regularly engage in discussions on issues such as misinformation, the need for clear policies during a pandemic, and decision making on both an individual and collective level. Leaders are already skilled at providing safe environments for discussions of a sensitive nature, and we build upon their expertise in youth development.

Method

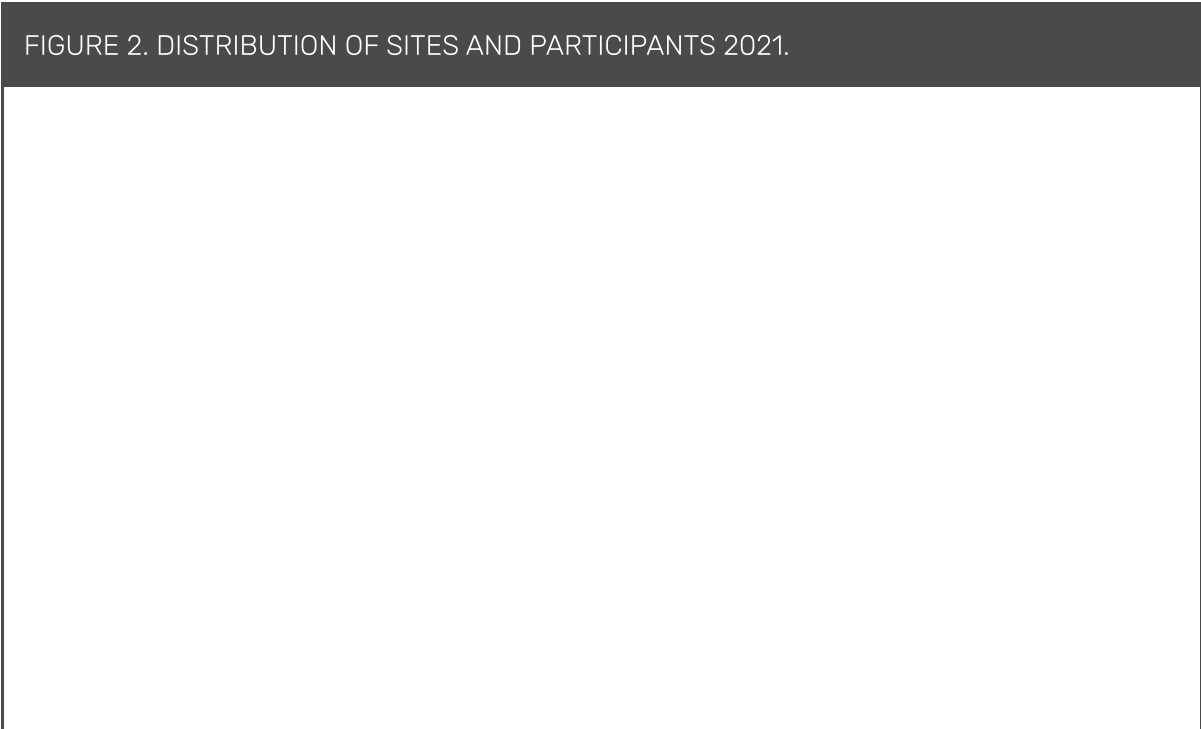
Approximately 655 school-age students (grades 4–8) in 47 sites in 10 communities across nine states in the country participated in the program. Fifty site coordinators for Imagine Science were recruited among employees of the 36 affiliated agencies within the national Imagine Science network and were invited to participate in the program during the summer or fall of 2021 and the spring or summer of 2022. Nine of the Data Detective Clubs took place during school hours or immediately after school in a school setting. One of the Clubs was a class in an expeditionary learning school and another was a specialty science summer camp program. The rest were held either virtually or in person in the sites where the YMCA, Boys and Girls Clubs, 4-H, and Girls Inc., offer programs. Seventy percent were summer programs.

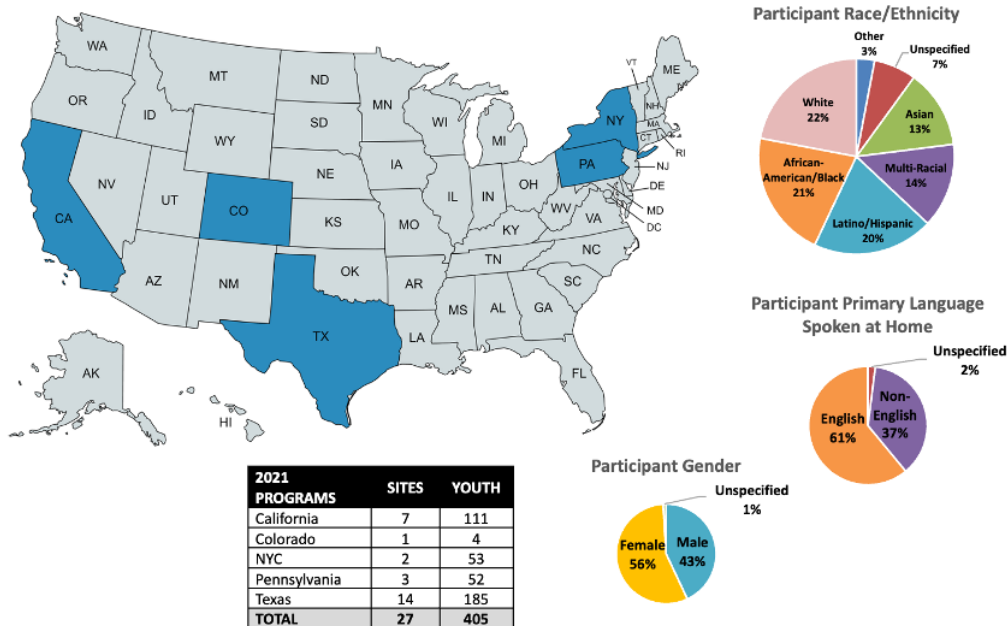
Site recruitment includes informational sessions and sharing program materials. Program sites submit interest surveys and complete an assessment prior to selection. Program selection is based on access to the targeted participant age group, staff capacity to participate in training and research, and technological resources to support implementation.

Once selected, the leaders receive books and materials, and they participate in the training described above. Imagine Science conducts regular check-ins, and project partners made themselves available for assistance during dedicated office hours.

Among the leaders who participated in 2021, 54% lived in the communities where they worked, and 80% had experience conducting science activities with youth (SLP4I 2022).

As seen in the graphic below, the project reached a diverse audience of youth in 2021: 22% white/non-Hispanic Caucasian, 21% African American, 20% Latinx/Hispanic, 14% multiracial, 13% Asian/Asian American, with 56% girls and 43% boys. In 2021, 37% of the youth served came from homes where the primary language is not English.





Click [here](#) for larger image

Evaluation Results

To gain a solid understanding of how the program was implemented and factors that might be associated with its success, our evaluation partner Strategic Learning Partners for Innovation (SLP4i 2022) used mixed methods and culturally responsive measures to collect data from youth, leaders, and project partners. Key findings from the evaluation were that youth understood the goals of the program, enjoyed the program components (story and characters, data mapping, guest speakers), felt that science is accessible to all, and gained a better understanding of science and related careers. Other findings showed youth needed support for understanding large numbers and applications of graphing in the real world.

Program leaders noted high interest in the story concepts and in subsequent discussions among participants; they appreciated the professional development time to practice with the software; they had suggestions for organizing the materials; and they recommended taking more time with implementing the program. The detailed Program Leader's Guide was highly valued. Ten of the leaders elected to run the program twice.

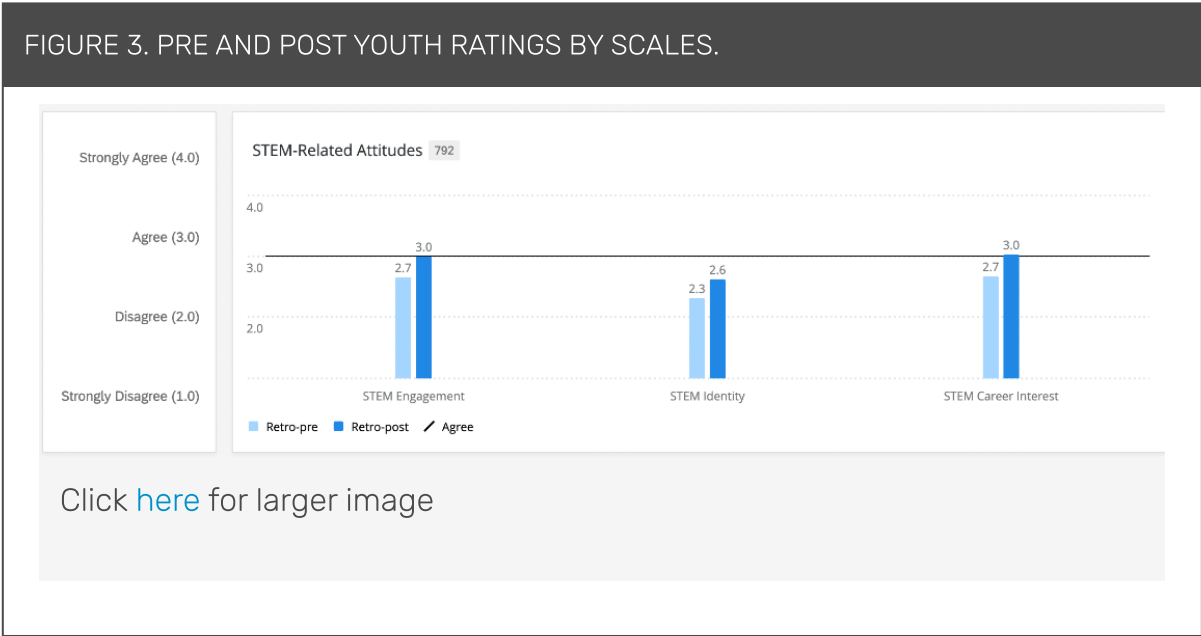
Research Survey Results

To understand the impact of the program on participants' attitudes about STEM—and about data and epidemiology in particular—we worked with Partnerships in Education and Resilience (PEAR), who ad-

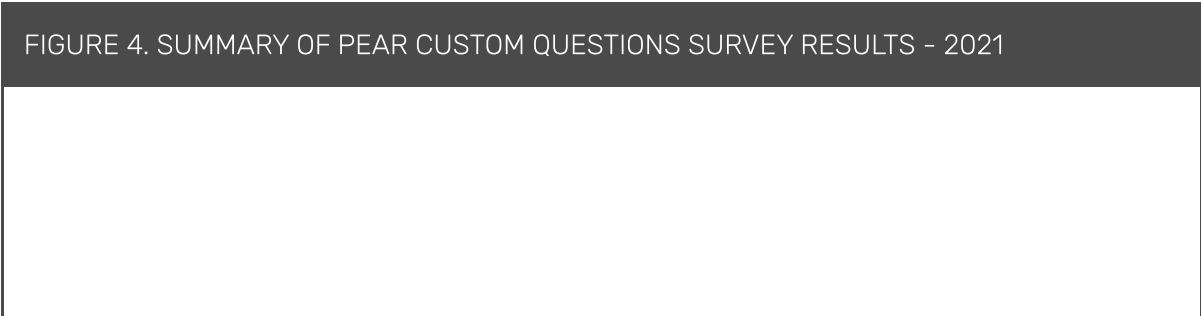
ministered a validated instrument that has been employed with thousands of youth in afterschool programs (the Common Instrument Suite; Noam et al. 2020). This survey, which uses a “retrospective pre-post” design, was administered to 200 participants in the Data Detectives Clubs during 2021. This design empowers participants to rate where they were at the beginning of the program and where they stand now with respect to variables of interest. The scales involve STEM engagement, STEM identity, and STEM career interest. Figure 3 shows the results of these scales among youth in 2021. The horizontal line represents an average answer of “Agree” (3 on a 4-point scale).

In addition to the validated scales, our team added 14 customized questions to the PEAR survey to study engagement, identity, and career and general interest in data science and epidemiology topics specifically, as well as attitudes toward reading and community impact of data science.

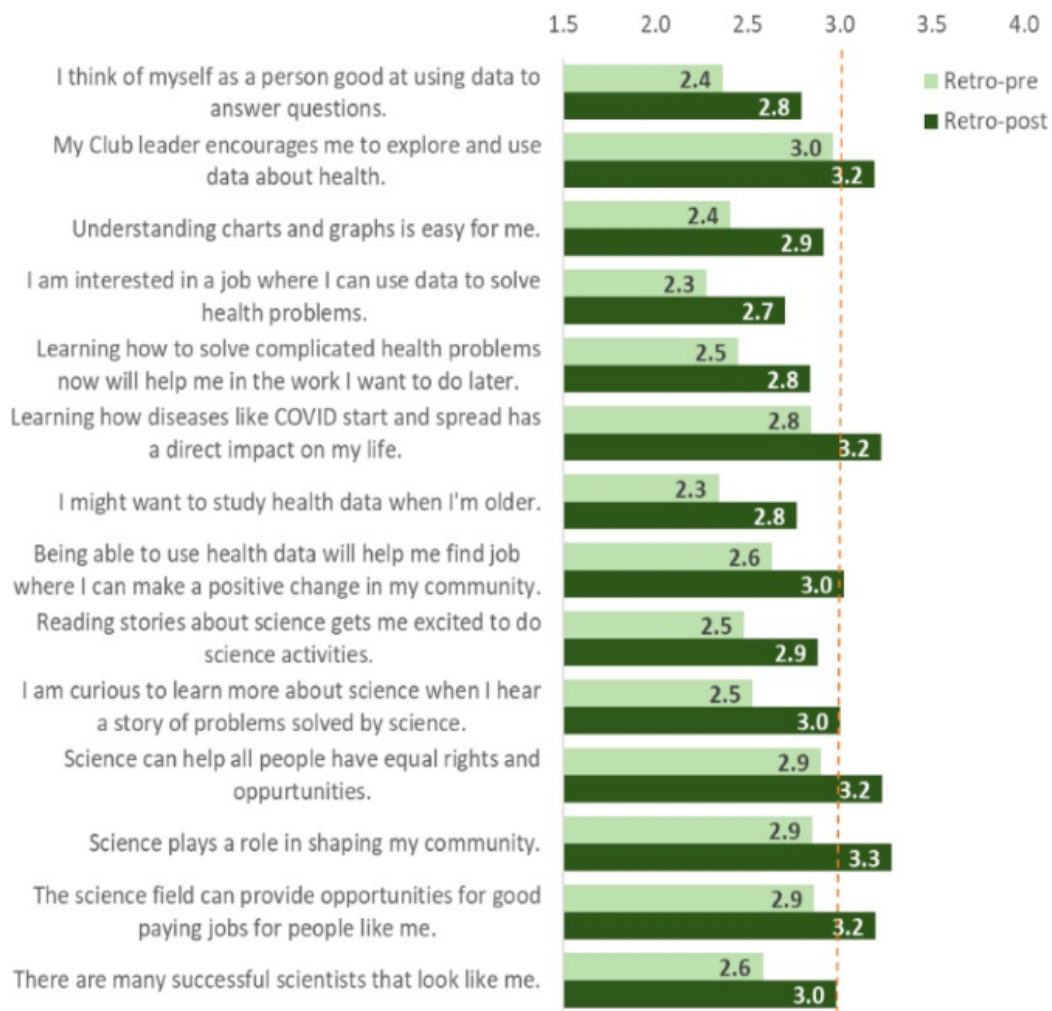
Results from 2021 showed that youth demonstrated statistically significant positive changes on all three scales on the PEAR-validated survey (see Figure 3) as well as positive changes on our 14 custom survey items. Program-related gains also exceeded those in the national comparison group on all three scales. There were no significant differences within age, grade, gender, hours of participation, or race and ethnicity. Results from the 2022 participants are forthcoming.



Below, we present self-reported changes in students’ interests, attitudes, and understandings as they related to the project (see Figure 4; PEAR 2022).



Custom Questions Added to CIS PEAR Student Survey (N=200)
 (1=Strongly Disagree, 2=Disagree, 3=Agree, 4=Strongly Agree)



click [here](#) for larger image

Research on Products Collected From Youth

To dig more deeply into what youth understood about data as a result of their participation in the program, we examined a sample of “artifacts” collected from participants. Specifically, we looked at graphs that youth produced in CODAP, the way they used NetLogo to determine how disease spreads when various percentages of the population are immune, and how they used data to support recommendations they made as part of a final project. These products were analyzed using scoring rubrics.

Overall, participants grasped the use of time on the x-axis. They needed a little more time to explore the functions of CODAP such as the way to smooth out data points and highlight different slopes. Participants also needed some additional support representing multiple variables and plotting attribut-

es. Participants understood that they could run a simulation multiple times with the same starting variable and get somewhat different results each time due to variation. At the same time, students understood that infection rates would vary as a function of the number of people who were immune at the beginning of the simulation.

We also found that preparing a set of recommendations in the context of producing a Public Service Announcement was not a successful vehicle for participants to make data-based recommendations. The context did not encourage youth to articulate the connections between the data and recommendations beyond standard ones such as wearing masks, keeping your distance, and washing your hands. However, students had fun creating their PSAs: One poster showed Big Foot, “the original social distancer.”

Based on early analysis of the final project, we revised instructions to clarify that recommendations needed to be based on data. With this revision, more youth (65%) correctly labeled and identified appropriate graphs and integrated data into their recommendations for the final project that asked them to argue for closing or opening up a venue to the public. The presenters to a “local council” were more successful in showing their understanding of graphs and tables. One young entrepreneur argued that a concert venue should be closed but that he would provide a multi-platformed digital alternative with student discounts!

Discussion

The partners who designed Data Detectives Clubs have capitalized on the resources, services, and close connections of the unique umbrella organization Imagine Science. The Clubs have benefitted particularly from two aspects of Imagine Science’s supportive framework: First, Imagine Science provides extensive staff development as well as ongoing support. Second, Imagine Science conveys to sites an expectation that a STEM intervention will be substantial enough (15–20 hours or more) to have a real impact on youths’ interest in STEM and STEM careers. In other words, by working with Imagine Science, the Data Detectives Club program has an opportunity to achieve maximal impact.

At the same time, the Data Detectives program has offered something of value to the Imagine Science sites: a shared, comprehensive, multi-component curriculum. Instead of scrambling to find activities and materials, leaders during the pandemic could focus more on their own learning and on responding to youths’ specific social and emotional needs.

The Data Detectives program itself encompasses several innovative components. First, the immediacy of the topic has encouraged the program to integrate technical learning and skill-building with social and emotional support. While afterschool leaders are mostly comfortable with their social support function, they are not all comfortable with manipulating and recording data. The second innovation, then, is introducing staff to integrated literacy and data literacy activities and to new technologies in a

way that can fit into an afterschool setting. Along with hands-on activities and discussions, these elements provide a multi-channel opportunity for youth learning.

Finally, because the program builds on a timely topic of significant interest and consequence, constant refinement and innovation are built into the program plan. Content has shifted as developments occur in the scientific understanding of COVID, its prevention and treatment. For instance, the topics of racial disparities in COVID's impact, new vaccine and COVID detection technologies, and variants have been added for successive groups in response to the pandemic's progress.

LESSONS LEARNED

Lessons Learned

- Epidemiology and data science are not targeted in schools, so after school is a good environment to introduce them.
- Club facilitators are not STEM professionals, but they can still do effective STEM activities.
- There is value to have narrative connect activities, even for those who read or hear only a brief summary of the text.
- Youth love simulations; it is easy for them to work with data tools.
- Career visits are exciting for youth; their greatest growth was in measures of career awareness.
- Kids enjoy reading, but it is sometimes difficult to get them to read.

Conclusion

This project reaches underserved youth with meaningful STEM opportunities related to epidemiology and offers afterschool leaders significant professional development opportunities. Epidemiology and related data science concepts are usually not taught in schools, much less in afterschool settings. Out-of-school settings—being mostly free of subject department expectations and state curriculum standards—can offer an ideal environment for interdisciplinary learning, such as integrating epidemiology with literacy, data, career exploration, and important socio-emotional learning. Such settings can allow youth to explore current and pertinent topics and allow staff to hone their skills in providing engaging hands-on and technology-based activities.

Acknowledgment

This work was funded by the National Science Foundation under grant DRL-2048463. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.

Activity Materials

- *The Case of the COVID Crisis* book can be ordered at <https://bit.ly/37dRfg8>
- The Program Leaders Guide is free and available by requesting it at <https://forms.gle/texvds3x-RQVwbasq6>
- CODAP and NetLogo are free and are available at <https://codap.concord.org> and <https://ccl.north-western.edu/netlogo/>

Laura Martin is Director of Research for CIDSEE at Tumblehome Inc., in Scottsdale, Arizona. **Janice Mokros** is a Senior Research Scientist at Science Education Solutions in Los Alamos, New Mexico. **Nav Deol-Johnson** is National Programs Operation Manager at Imagine Science in Long Beach, California. **Pendred Noyce** is Director at Tumblehome Inc., in Boston, Massachusetts. **Jacob Sagrans** is a Senior Research Associate at Science Education Solutions in Los Alamos, New Mexico.

References



Careers

Computer Science

Equity

Interdisciplinary

Life Science

Middle School

Informal Education

Start a New Discussion



Start a new discussion about this article in the New Teachers forum!



Start a new discussion about this article in the General Science and Teaching forum!



Start a new discussion about this article in the Evaluation and Assessment forum!



Start a new discussion about this article in the Life Science forum!



Start a new discussion about this article in the Informal Science forum!

[Visit the Forums](#)

You may also like



Leadership Matters

Journal Article

[Leadership Matters!](#)



The Poetry of Science

Journal Article

[Solar Power](#)



The Early Years

Journal Article

[Preschoolers “On the Move” with Energy](#)



Journal Article

[Energy](#)

“Energy is liberated matter, matter is energy waiting to happen.” — Bill Bryson, A Short History of Nearly Everything ...