Integrated Data Science for Secondary Schools: Design and Assessment of a Curriculum

Emmanuel Schanzer
Bootstrap
USA

Nancy Pfenning
University of Pittsburgh
USA

Flannery Denny
Bootstrap
USA

Sam Dooman
Brown University
USA

Joe Gibbs Politz
University of California, San Diego
USA

Benjamin S. Lerner
Northeastern University
USA

Kathi Fisler
Brown University
USA

Shriram Krishnamurthi
Brown University
USA

ABSTRACT

We propose that secondary-school data-science curricula should be based on four key ingredients: two are technical (programming and statistics, with visualization sitting at their intersection), while two are human-facing (meaningful domains, and civic responsibility). We describe their relationship and argue for their importance.

Based on this, we then present the Bootstrap:Data Science curriculum, designed for integration into multiple disciplines and settings. It achieves this by (a) being designed as a set of remix-able lessons, and (b) letting classes and students choose personally meaningful datasets.

We also initiate the process of evaluating this curriculum. We create two assessment instruments, one focused on learning and the other on personalization and engagement. We provide very preliminary data gathered from students and teachers.

CCS CONCEPTS

• **Social and professional topics** → **K-12 education**; Information science education; • **Applied computing**;

KEYWORDS

data science; programming; statistics; application domains; civic responsibility; integrated curricula; engagement

ACM Reference Format:
Emmanuel Schanzer, Nancy Pfenning, Flannery Denny, Sam Dooman, Joe Gibbs Politz, Benjamin S. Lerner, Kathi Fisler, and Shriram Krishnamurthi. 2022. Integrated Data Science for Secondary Schools: Design and Assessment of a Curriculum. In *Proceedings of the 53rd ACM Technical Symposium on Computer Science Education V. 1 (SIGCSE 2022), March 3–5, 2022, Providence, RI, USA.* ACM, New York, NY, USA, 7 pages. https://doi.org/10.1145/3478431.3499311

SIGCSE 2022, March 3–5, 2022, Providence, RI, USA

ACM ISBN 978-1-4503-9070-5/22/03…$15.00
https://doi.org/10.1145/3478431.3499311

1 INTRODUCTION

Data science curricula are growing in popularity in several countries. The USA has both state- and nation-wide [6] initiatives to expand their use and to embed them in schools. Popular commentators have argued for their importance, and several curriculum providers have created initiatives to teach this material (§2).

However, the field is still in a relative fledgling phase, as standards-writers, curriculum designers, and others are trying to formulate what data science *is* and hence what a curriculum must cover. Many conversations are situated in efforts to modernize math education around Algebra 2 and Calculus [13, 16]. We posit that a quality data science curriculum should be broader and include four ingredients: programming, statistics, domains for application, and civic responsibility. (At the K–12 level, we view visualization as lying at the intersection of programming and statistics.) The first two cover core technical skills; the third is important to motivate students, to enable use in multiple contexts, and to have impact; and the fourth is necessary both to understand applications and to avoid the perils of using data without regard for their social impacts.

We present the Bootstrap:Data Science (BS:DS) curriculum, available from `www.bootstrapworld.org/materials/data-science`, that mixes the above four ingredients. BS:DS is ideally suited to secondary schools, especially ages 12–16 (US grades 6–10), though it has been used with both younger and older students. About 8000 students used it the 2020–21 academic year, in about 50 cities in the USA and Canada. There are two key aspects to the curriculum:

(1) It is designed for *integration*, i.e., for embedding into existing disciplines (e.g., mathematics, science, social studies) rather than as a stand-alone course. It can of course be taught stand-alone, but many schools lack free space in the curriculum or teachers with capacity to offer such courses.
(2) Because of integration, we cannot assume every teacher is an expert in all the required skills (especially both programming and statistics). Thus, the curriculum uses tools and pedagogies that help teachers to present the materials even as they develop their own skills. In return, it gives students agency through the *choice of data sets*, which can potentially increase engagement and personalize their education.

In addition to curriculum design, we are also interested in assessment, on which there is relatively little work for secondary-school

data science (§2). In this paper we briefly present two distinct instruments that reflect the above two aspects. One is designed to assess learning of the technical content, while the other examines student engagement through curriculum personalization. Naturally, we believe the questions presented here are only preliminary and will require several rounds of revision, but we hope they will help spark a conversation on this topic. We also present some very preliminary data obtained from deploying these instruments on small portions of our users.

2 RELATED WORK

Several data-science curricula have been proposed in recent years. Many are at the post-secondary level; these are outside our scope. IDS [20] is a full-year course for high-school students using R, as is Berkeley's Data8 course [22]. YouCubed's curriculum [24] is designed to replace Algebra 2 and uses Python. None of these are designed for use in integrated contexts at the secondary-school level. The Concord Consortium produces CODAP [4], a platform for teaching data analysis and visualization, but with limited support for programming, and no support for programming outside of data science contexts. Tuva Labs offers data science materials [21] that focus on data analysis and data modeling, but not on computer science (or programming beyond data science). The Data Science for Everyone initiative [6] links to a variety of other resources.

We have not found publicly-available data science assessment instruments that target the same integration we achieve, to compare with ours (in §5). Feldon and Litson [9] have a validated (but unpublished) assessment of data-visualization skills.

3 THE FOUR INGREDIENTS MODEL OF DATA SCIENCE

We take the position that a strong and equitable data science curriculum combines four ingredients:

- **domains of study** Data and questions about them arise in concrete *domains* such as healthcare, government policy, sports, social phenomena, and scientific research.
- **statistics and mathematics** Statistics and, broadly, mathematics (especially algebra) provide the analytical tools for answering questions about data.
- **computing** Computing enables transforming, combining, visualizing, and managing data before, during, and after analysis. It makes questions about data actionable by encoding them as programs.
- **civic responsibility** Civic responsibility helps students understand their roles as both producers and consumers of data and warns about the perils when analysis is done poorly, irresponsibly, or without attention to potential harms.

We use the term "ingredients" because we view this much as one might a cooking recipe: each teacher or district will adapt it to their tastes and needs. Different curricula can use different quantities of the ingredients so long as the dish remains recognizable. This flexibility allows data science to take root in a variety of subjects (math, social studies, science, computing, language, etc.), which provides flexibility for schools and districts. For instance, a data-enhanced lesson in social studies can reinforce summary statistics and visualizations while discussing civic responsibility. A lesson for science can emphasize data collection and preparation. A lesson with stronger computing emphasis can help meet CS standards around algorithms and social impacts. Each of these approaches can incorporate the four ingredients while simultaneously supporting other curricular goals. However, we believe curricula that completely drop one of these can fail in strength or equity.

In particular, many framings for data science are driven by math and statistics standards [2, 3]. While these are important, they also have significant weaknesses: they can easily fail to address the computational needs of data science (which include aspects of data engineering). While they may reference coding, which can capture the low-level operations needed to work with data, students also need high-level skills related to planning, design, and testing that are rarely covered in typical K-12 coding curricula. Thus, while the required coding skills may seem modest, math-centric framings that focus only on coding miss the many levels at which computing education matters for data science.

Other framings can and do give short shrift to civic responsibility (much less cultural responsiveness). Students will be creators, users, and even victims of data. They should understand how the same data can tell multiple stories, how applications use data, and the risks of misusing or over-relying on decisions recommended by algorithms. Such issues draw deeply on both statistical and computational concepts, but must be explored in rich contexts that are personally and culturally relevant to students and communities.

Of course, these ingredients can be satisfied in different ways. For instance, the computing requirement does not only need be satisfied by programming in Python or R. A variety of tools, including spreadsheets, can be used to accomplish similar ends while satisfying other needs or addressing the abilities of different audiences (e.g., younger students). Indeed, a fair degree of basic data science can be conducted with relatively few concepts and constructs mixed together in the right ways.

4 CURRICULUM OVERVIEW

BS:DS is a set of lessons that lead students who have no prior experience in programming or statistics to perform basic analysis of real datasets and present findings in written form. We discuss several facets of the curriculum before outlining the specific lessons.

Communication and Other Standards. The curriculum makes students ask questions, reformulate these over data, write programs, interpret output, and write up their findings in prose as a report. All these align well with skills in STEM standards such as NGSS [5]. They also support already-authentic practices in disciplines such as social studies, in which teachers are using BS:DS.

Tables. Currently, BS:DS focuses only on *tabular* data. While other data formats (like natural language) are of course important in data science at large, we focus on tables for several reasons. First, many important datasets are published as tables (e.g., CSV files). Second, they have several convenient user interfaces for creation (such as spreadsheets and even surveys). Third, they are unambiguous to parse, unlike formats like natural language. Furthermore, because of this lack of ambiguity, they do not *force* the introduction of probabilistic reasoning earlier than necessary in the curriculum.

As we discuss in §1, an important part of the curriculum is letting students personalize their study by choosing data of interest or meaning to them. Nevertheless, not all students are necessarily invested in or adept at this. The curriculum therefore comes with a family of curated datasets, currently 26 in number, that cover environmental issues, politics, art, sports, entertainment, and more. These naturally represent the interests of our *current* population, and keep growing as our user population does.

One difficulty when teachers or students try to use datasets they find relevant and accessible is that they may choose one that does not meet our curricular objectives. Our curriculum is written to assume that every table has these characteristics:

- It must be large enough to have real variability without taking too long to process. Teachers usually determine the former based on their statistics learning goals. In practice, the latter tends to mean 1000–5000 rows.
- It must have at least two categorical columns and at least three quantitative ones.
- It should have at least one high correlation, and ideally multiple of different strength and polarity.

All our curated datasets meet these criteria, and we work with teachers to help them find datasets with these properties.

The early parts of the curriculum (outlined in fig. 1) work with a specific (artificial) dataset about animals in an animal shelter. It is carefully designed to meet all the above criteria, and to provide a fixed reference around which course materials can be written. After students have learned the basics on this dataset, they are then free to experiment with one of their choosing (starting with lesson 13).

Data Quality. Real-world data science includes a fair amount of "data wrangling", such as cleansing a data set. This can involve ad hoc skills, as well as (sometimes) fairly advanced programming and statistics—more advanced than a secondary-school curriculum tries to cover. We make a conscious choice to avoid running into these issues early on (even though we can support this for those who want to explore it). Our curated datasets have properly cleansed data so that students can focus on basic skills without distraction.

A special case of this is the problem of missing values in data, which has a long and complex history. A proper understanding of missing data shows that they cannot be reconciled easily: as Date notes [7], there can be many *interpretations* of a missing value, each of which may require different treatment. Again, therefore, we avoid them in our curated datasets.

Programming. All programming is currently done in the Pyret [1] programming language.[1] Pyret is a Python-like language that eschews some of Python's strange features and design flaws [18] and embraces tables. It has an accurate handling of numbers (rather than directly exposing machine numbers), to reduce numeric surprises; built-in support for tables (including both syntax for defining them and visual output for displaying them); a SQL-inspired query language integrated into the language; direct interfacing with Google Sheets for importing data; and other features that make it especially well-suited to this curriculum, while retaining the syntactic simplicity and convenience of Python.

[1] A simplified version of this curriculum for 5th grade uses Google Sheets, and a parallel version using CODAP [4] is currently being written.

The use of a textual language may raise some concerns. Though we have considered the use of blocks, teachers report that they and their students often find greater authenticity in and enthusiasm for programming with text. Furthermore, the amount of programming in BS:DS is not very much. Finally, Pyret's error messages, which are the result of several research papers [14, 15, 23], are actually a useful pedagogic device in their own right. Therefore, choosing between textual and block programming in this setting is not trivial.

Lessons. The curriculum is organized into *lessons*, which individual teachers can "remix" into sequences that are meaningful to them. This flexible approach is important in an integrated setting, where teachers in different contexts may want to use varying quantities of content. Each lesson plan lists dependencies on other lessons; the curriculum suggests standard pathways. Figure 1 summarizes the lessons and how each interacts with the four ingredients. Domains of study are covered by personalization, either determined by the instructor to be meaningful to the subject (e.g., social studies), or chosen by the student to be meaningful to them.

What's Not Here? Naturally, given the size of a discipline like data science, many topics must be elided or handled in only very limited ways at the secondary level, especially in the 6–10 grade range that is our focus. Here are some notable elisions:

- Visualization is a large subject with its own methods and pitfalls. We limit ourselves to the visualizations present in secondary-school statistics standards, which are basic plotting and graphing functions. At that level, the main question is choosing the right one from a standard toolbox, whereas in general, visualization involves creating *new* forms of presentation that take into account disciplinary standards.
- The curriculum currently eschews continuous math, focusing on discrete mathematics. Probability theory typically only arises in US grades 11–12 (and even then, often only in Advanced Placement Statistics). Including this, however, would be a natural extension of what we do.
- As noted earlier, we currently largely sidestep data quality issues. While Pyret supports both data cleansing and handling missing values, our lessons do not currently exercise them: in an *integrated* setting, teachers rarely have time for more content (since they have their host discipline's materials to cover too). These would likely only be used in stand-alone data science courses, for which we are building out more materials. In particular, BS:DS leads into our college-level materials and textbook [10] which cover these issues as well as more advanced computing content.

5 TECHNICAL SKILLS INSTRUMENT

To assess student learning of the BS:DS technical skills, including whether students can accurately interpret data, we have developed (and are pilot-testing) an instrument on this topic. It features 11 questions that focus on programming, statistics, and visualization. Due to space limitations, here we only summarize the questions and their goals. The full instrument is available from `cs.brown.edu/research/plt/dl/sigcse2022/`.

Lesson	Stats/Mathematics	Computing	Civic Responsibility	Domains of Study
1	**categorical vs quantitative data**	questions that can be asked about datasets		
2		**introduction to programming** in Pyret, including basic data/operations		
3	domain and range of functions	images as datatypes, **manipulating basic data with programming** operations		
4	**pie charts and bar charts**	expressions to create charts from tables		
5	**interpreting pie and bar charts**	extracting rows and cells from tables		
6	**manipulating data tables to focus on a specific question**	using programs to order and filter rows, as well as to compute new columns		
7	**identifying patterns in expressions that access data**	developing functions from examples (abstraction)		
8	how domain/range, input/output examples, and patterns in expressions represent different aspects of functions	**step-by-step processes to develop functions** to manipulate and plot data (systematic design process)		
9		combining multiple table manipulations to **define interesting subsets of data** (decomposition and planning)		
10	considering **subgroups of a population**	if-else expressions		
11	**random samples and statistical inference**		the importance of good samples	
12	the value of considering multiple subgroups of population	**extracting populations based on precisely-stated criteria**	disaggregating sub-populations can expose different results for different populations	
13			an opportunity for teachers or students to study a dataset or question with civic consequences	**choose a dataset** to study: identify categorical/quantitative variables, meaningful subsets, and questions of interest
14	**distributions in datasets and histograms**			interpreting analysis results to learn about the domain of study
15	**data shape and skew**	creating multiple plots to understand properties of data		interpreting analysis results to learn about the domain of study
16	**measures of center**			interpreting analysis results to learn about the domain of study
17	**spread of datasets and box plots**			interpreting analysis results to learn about the domain of study
18		writing tests to establish that programs produce expected answers	describing **whether an analysis is trustworthy**, and using code to check it	
19	**scatter plots**			interpreting analysis results to learn about the domain of study
20	**correlations**		correlation versus causality	interpreting analysis results to learn about the domain of study
21	**linear regressions**			interpreting analysis results to learn about the domain of study
22			**inferences and bias from machine-learning algorithms**	
23			forms of **bias in analysis interpretation and reporting**; misleading readers with statistics	consider what constitutes validity and bias in a specific domain

Figure 1: Summary of the lessons in BS:DS and how they exercise the four ingredients. Lesson foci are in bold.

In all of the questions, respondents have the option of indicating there are no suitable options or that they don't know. The instruments are carefully designed to include distractors.

Types and Representations of Data. The first group of four questions concerns a table of sample data about made-up movies. They:

- Ask respondents to classify each column as categorical, quantitative, or neither.
- Ask respondents what *programming data type* would be most suitable to *represent* the columns.

This group illustrates how we bridge statistics and programming.

Data Quality. The next two questions are about this table:

Likes History?	Grade	Height (in.)
"yes"	6	5.5
"false"	6	"4ft 10in"
"nope"	9	"61 inches"
"no"	11	72

Respondents are asked whether or not each column is "ready to use to make meaningful charts or compute meaningful statistics". They are then asked to pick one column that they indicated was not ready (if any) and show its corrected version. These questions are designed to make sure respondents understand the perils of processing data that have not been normalized, and that they have some understanding of what a properly normalized datum might look like (even if fixed by hand).

Program Planning. The next question presents a sample table of hypothetical sales. It describes a desired task that requires several computational steps. The goal is to determine how well respondents can *plan* [19] a solution. Respondents are given six possible steps, including distractors, which they have to put in order (avoiding distractors), in the style of a Parsons problem [8].

Visualization Comprehension. The next two questions each show a graph and ask respondents to answer questions about it. The first shows four histograms and asks which is most likely to correspond to a particular interpretation. The second goes in the other direction: it shows a histogram and asks respondents to interpret it.

Table Interpretation. The next question shows a subset of rows from a table of hypothetical sales at a fruit store. It then lists several questions and asks respondents to determine which can and cannot be answered based on the table.

Visualization Interpretation. The last question shows a pair of plots based on data from IMDB, a movie database. One shows the average rating of the top-250 and bottom-250 movies (as determined by their rating). The other plots rating against the number of reviews for the top-250, and overlays a linear trend line. It then asks whether a series of statements is true or false based on these plots. Critically, some of the options are *causal* statements, which of course cannot be determined from these plots at all.

Pilot Deployment. We have been able to administer this instrument to two groups: teachers and students. However, due to lateness in the school year and COVID-19 pressures, we have only a dozen teacher responses and 23 student responses. These numbers are sufficiently small that any detailed interpretations would be meaningless. We can confirm that (a) some questions did obtain a wide range of responses (suggesting a lack of agreement among the population), (b) our distractors did trip up some respondents, (c) some questions identified clear cases of confusion, most notably about what to do with data that do not neatly fit the categorical–numeric split (such as the studio numbers where movies were filmed). A detailed discussion of these issues will not fit in the space available in this paper, and needs to be deferred to a future venue.

6 STUDENT PERCEPTIONS INSTRUMENT

With data science being touted as a way to let students work on personally-relevant projects (as BS:DS supports), understanding how students perceive their work and choices are important. We designed an instrument for this purpose and asked a handful of experienced USA-based BS:DS teachers (ones who still had bandwidth at the end of a semester under COVID-19) to help us pilot test it with students. We received 94 unique responses, all of which contained answers that responded meaningfully to the questions. Some questions were free form, while others used a 4-point Likert-type scale, ranging from "Strongly disagree" to "Strongly agree".

Grade Levels. Though BS:DS is primarily designed for grades 6–10, most of our respondents were outside this range, in grades 11–12 (ages 17–18): 13.8% in each of 9th and 10th, 35.1% in 11th, and 37.2% in 12th.

Pair Work. In BS:DS, classes are encouraged to engage in pairwork, but the choice is left to teachers. Even within a single class, a teacher may give students the choice of working alone or in pairs. We found that 42.6% worked with a partner. 52.3% of pair-working students (22.3% of all respondents) strongly agreed to working well with their partner, 42.5% of them (18.1% of all) agreed, and only one student each picked disagree or strongly disagree.

"I learned ..." When asked to respond to "I learned some programming from doing this", we see overwhelming support: 42.6% strongly agree and 54.3% agree, with only 3 students (3.2%) disagreeing. On "I learned some data analysis from doing this", we see even stronger support: 46.8% strongly agree and 53.2% agree, with nobody disagreeing. Finally, for "I learned some graphing (data visualization) from doing this", we again see largely positive support: 40.4% strongly agree, 53.2% agree, and 6 students (6.4%) disagree.

"What was your dataset about?" In most cases students chose one of our curated datasets, though in a handful of cases they either downloaded fresh data from the Web or created their own (e.g., water filters in their school district, or asking classmates how much time they spent watching the NBA every week).

Three categories of data dominated the responses. Topics related to food and health came first at 20.8%. Two were about New York City restaurants, but all the rest had a health dimension: nutrition in fast foods, sodas, (breakfast) cereals, snacking, and cancer rates. Equally popular was a cluster on crime and policing, mostly about stop-and-frisk and marijuana arrest rates (comparing Black versus white, rural versus urban, etc.): in short, politically important and sensitive topics (some of which may have been especially salient in light of 2020's Black Lives Matter protests). Close behind was the

category of sports and entertainment (16.4%), but spread between baseball, basketball, track-and-field, and video game reviews. A dataset of the "Top 100 Movies" attracted 7% of students, followed by a long tail: data about specific states, the environment, animals, college majors, and charter schools each had more than one user.

In short, we see a broad spread of topics that could reasonably be described as personally meaningful. More directly:

"I chose my dataset because…" Students could choose more than one option. Here, 42.6% said they chose it because it was something they already knew something about, while just as many (40.2%) chose it because "I don't know much about it, and I was curious". In addition, 16% said "It affects me personally". Numerous students used free-response to express specific interests such as representing a favorite hobby or activity, or "I am interested in the medical field". Still, small numbers contained a potentially negative response: 11.7% said they had no reason, 4.3% said it was their partner's choice, and 3.2% said they did not have a choice (presumably meaning the dataset was chosen by the teacher).

Overall, then, we see that the personalization is accompanied in general by a high degree of engagement and, most importantly, low degrees of *dis*-engagement. This suggests that the technique of providing a (limited) range of curated datasets still gives many students room for personal expression.

Data Analysis. Students were asked whether they found the data analysis interesting, whether they felt they had done real data analysis, and whether they were proud of the work they had done. The results are overwhelmingly positive: only one student disagreed with the first, 7 with the second, and 4 with the third (and in no case did a student strongly disagree). Additionally, 22 (of 94) students voluntarily shared a link to their report or software.

Narrative Responses. Finally, students were given the option of a free-form response on what they "liked or disliked about analyzing data". A total of 78 students (which we considered a quite high response rate) wrote in answers.

Numerous positive responses again captured interest in the dataset and personal or cultural meaning they derived from it. Some appreciated procedural aspects, such as the creation of graphs. Several described forms of productive failure [11], such as making mistakes and fixing them, or the challenge of finding trends.

Several, however, also described difficulties with time, preparation, or tools (e.g., the difficulty of programming textually compared to using only Google Sheets) that led to a disappointing experience.

Only a small number reported overall negative experiences: being unable to finish, excessive frustration (e.g., getting errors due to case-sensitivity), blandness of the chosen dataset, being overwhelmed by numbers, or problems specific to datasets (e.g., not enough categorical variables). One criticism may reflect a philosophical bent: "No matter what type of data you are trying to analyze it never gets really interesting because in the end its all the same, numbers, letters and categories. In short, it's boring."

Survey Size. While there are many more questions we could have asked to probe deeper into these issues, in our experience, surveys that take longer than 10–15 minutes are frowned upon by teachers or obtain low response rates. We therefore worked hard to keep the number of questions small and make them as close to multiple-choice as possible, using free-responses only where necessary. This leads to some loss of information in return for teacher willingness to distribute the survey and students responding to it. Nevertheless, there is value to a more detailed survey that probes these matters in more depth—ideally there would be some compensation for students and teachers to fill it out.

Reflection. Overall, we believe this instrument provides useful insights into student perceptions and engagement. The fact that we did not have to filter out any responses due to quality of answers is telling. Of course, it is possible that the respondents were those who felt most positive about their work. For this and numerous other reasons, this can only be considered a very preliminary, formative investigation. As such, we are more interested in qualitative and general quantitative senses of student experience with the pilot data; put differently, the results are useful for piloting, but do not rise to the level of *findings*.

We chose not to ask students to report their gender or race, especially because we were expecting a smaller sample size and did not want students to fear risk of identification. We acknowledge that responses on this instrument could vary considerably based on instructional context, teacher goals, and demographic factors. Recruiting a diverse population of students, teachers, and classrooms is part of our plan for taking this assessment to scale.

7 DISCUSSION, ADOPTION, CONCLUSION

This paper tries to spark discussion about how to position data science in K-12. Proposals to have Data Science replace Algebra 2 [13, 16] are generating many conversations in schools. The computing-education community also needs to be involved. Data and visualization concepts are already part of many CS standards, but more narrowly framed than in our four-ingredient model. With many districts already struggling to define viable plans to achieve "CS for All", there will be mutually-beneficial opportunities to teach CS and Data Science together, as long as proposed curricula respect the broader set of necessary ingredients. We hope this paper will spark and support conversations about this exciting opportunity.

Integrated curricula often cannot offer any one model of adoption; this is even more true when a curriculum, like BS:DS, can fit in so many subject contexts. Instead, we help teachers remix our lessons around their needs, available time, and expertise.

Deeper conversations are also needed regarding the ability of data science to personalize and support STEM learning, as well as general development of student understanding about data and civic responsibility (which our instruments do not yet cover). Our proposed instrument explores one level of student perception and engagement. More significant promise of data science education, however, goes deeper into enabling culturally-responsible education [12, 17], which our ongoing work examines.

ACKNOWLEDGMENTS

We thank the US National Science Foundation for support. We deeply appreciate the feedback of numerous teachers and students over the years. Jen Poole made many early contributions. We appreciate the detailed reviews; space precluded making all suggested improvements. Sam Dooman is currently at Down Dog Yoga.

REFERENCES

[1] [n.d.]. Pyret Programming Language. https://www.pyret.org/. Accessed: 2020-04-10.

[2] 2010. Common Core Standards for Mathematics. http://www.corestandards.org/Math/; retrieved August 14, 2019.

[3] Anna Bargagliotti, Christine Franklin, Pip Arnold, Rob Gould, Sheri Johnson, Leticia Perez, and Denise A. Spangler. 2020. *Pre-K–12 Guidelines for Assessment and Instruction in Statistics Education II (GAISE II): A Framework for Statistics and Data Science Education.* American Statistical Association.

[4] Concord Consortium. [n.d.]. CODAP. https://www.commonsense.org/education/website/codap.

[5] National Research Council. 2013. *Next Generation Science Standards: For States, By States.* The National Academies Press. https://doi.org/10.17226/18290

[6] data [n.d.]. DataScience4Everyone. https://www.datascience4everyone.org/.

[7] C. J. Date. 2005. *Database in depth: relational theory for practitioners.* O'Reilly.

[8] Paul Denny, Andrew Luxton-Reilly, and Beth Simon. 2008. Evaluating a New Exam Question: Parsons Problems. In *Proceedings of the Fourth International Workshop on Computing Education Research* (Sydney, Australia) *(ICER '08)*. Association for Computing Machinery, New York, NY, USA, 113–124. https://doi.org/10.1145/1404520.1404532

[9] D. F. Feldon and K. Litson. 2021. Assessment of data visualization skills. Prepublication.

[10] Kathi Fisler, Shriram Krishnamurthi, Benjamin S. Lerner, and Joe Gibbs Politz. 2021. *A Data-Centric Introduction to Computing.* Published on-line. https://dcic-world.org/, accessed 2021-11-07.

[11] Manu Kapur. 2008. Productive Failure. *Cognition and Instruction* 26 (2008), 379–424.

[12] Gloria Ladson-Billings. 1995. Toward a Theory of Culturally Relevant Pedagogy. *American Educational Research Journal* 32, 3 (1995), 465–491.

[13] Steve Levitt. 2019. America's Math Curriculum Doesn't Add Up. Freakonomics Radio episode 391, https://www.datascience4everyone.org/post/freakonomics-podcast.

[14] Guillaume Marceau, Kathi Fisler, and Shriram Krishnamurthi. 2011. Measuring the Effectiveness of Error Messages Designed for Novice Programmers. In *ACM Technical Symposium on Computer Science Education.*

[15] Guillaume Marceau, Kathi Fisler, and Shriram Krishnamurthi. 2011. Mind Your Language: On Novices' Interactions with Error Messages. In *SPLASH Onward!*

[16] Jay Matthews. 2019. Algebra II Just Doesn't Add Up. *Washington Post Perspectives* (Dec. 2019).

[17] Gholdy Muhammad. 2020. *Cultivating Genius: An Equity Framework for Culturally and Historically Responsive Literacy.* Scholastic Teaching Resources.

[18] Joe Gibbs Politz, Alejandro Martinez, Matthew Milano, Sumner Warren, Daniel Patterson, Junsong Li, Anand Chitipothu, and Shriram Krishnamurthi. 2013. Python: The Full Monty: A Tested Semantics for the Python Programming Language. In *ACM SIGPLAN Conference on Object-Oriented Programming Systems, Languages & Applications.*

[19] E. Soloway. 1986. Learning to Program = Learning to Construct Mechanisms and Explanations. *Commun. ACM* 29, 9 (Sept. 1986), 850–858. https://doi.org/10.1145/6592.6594

[20] Introduction to Data Science. [n.d.]. https://www.idsucla.org/.

[21] Tuva Labs. [n.d.]. Online Tools for Teaching and Learning Data Literacy. http://tuvalabs.com/.

[22] UC Berkeley. [n.d.]. Data 8: The Foundations of Data Science. http://data8.org/.

[23] John Wrenn and Shriram Krishnamurthi. 2017. Error Messages Are Classifiers: A Process to Design and Evaluate Error Messages. In *SPLASH Onward!*

[24] YouCubed. [n.d.]. High School Data Science Course. https://www.youcubed.org/resource/high-school-data-science-course/.