EDC Oceans of Data Institute White Paper

Data Use in the Next Generation Science Standards

By Kim Kastens (kkastens@edc.org)

November 20, 2014; revised May 28, 2015

Suggested citation: Kastens, Kim. (2015, May). *Data Use in the Next Generation Science Standards* (revised edition) [White paper]. Waltham, MA: Oceans of Data Institute, Education Development Center, Inc. Retrieved from http://oceansofdata.edc.org/our-work/data-next-generation-science-standards

Introduction

The three purposes of this white paper are to:

- document the abundance of data-using in the Next Generation Science Standards (NGSS).[1]
- explore the distribution of data-using in NGSS across grade bands and disciplines.
- begin to explore how well the data-using skills called for in the NGSS align with the tasks and duties performed by big-data-enabled specialists (BDESs) in today's workforce.

The headline findings are as follows:

- Twelve percent (24/206) of NGSS performance expectations (PEs) probe Practice #4 Analyze and Interpret Data.
- Practice #4 is more common in elementary (13%) and middle school (15%) PEs than in high school (7%).
- Practice #4 is most common in Earth & Space Science PEs (16% of those PEs), least common in Physical Sciences (7%), and of intermediate abundance in Engineering/Technology (14%) and Life Sciences (13%).
- Data are found or implied in many PEs that are not targeted at Practice #4. Using a coding system that counted instances in which students collected, evaluated, or otherwise engaged with data, we found that approximately 46% of PEs involve some form of work with data.
- Under the broader coding scheme, data work is fairly evenly spread across the educational trajectory, comprising 46% of elementary PEs, 49% of middle school PEs, and 44% of high school PEs.
- There is substantial overlap between the NGSS practices and many of the duties and tasks that expert data scientists say they perform, notably in the areas of articulating questions, designing experiments and analysis plans, and using results from data analysis to draw inferences and answer questions.

Methods

Coding of NGSS

The heart of the Next Generation Science Standards (NGSS) is a set of 206 performance expectations (PEs). Each PE was created by linking one science and engineering practice (SEP) with a disciplinary core idea (DCI) and, in some cases, with a cross-cutting concept (CCC). Under each group of PEs, the NGSS shows a crosswalk of which SEPs, DCIs and CCCs were used to create each PE. The area containing the crosswalk is called the Foundation Boxes.

The NGSS presents the PEs in two orders. In the data spreadsheet that underlies this white paper, we used the order "arranged by Disciplinary Core Ideas." We used the NGSS's grade bands (K–5, middle school defined as grades 6–8, and high school as grades 9–12) and the NGSS's disciplinary divisions [Physical Sciences (PS), Life Sciences (LS), Earth/Space Sciences (ESS), and Engineering/Technology/ Applications of Science (ETS)].

The simplest way to consider the importance of data in the NGSS is to look at the abundance of performance expectations that include *SEP #4: Analyze and interpret data*, and this is done below. However, use of data is pervasive in science, and thus use of data can be found in other PEs. Defining what constitutes "data" turned out to be non-trivial, as had been previously reported by Manduca &

[1] NGSS Lead States. (2013). *Next generation science standards: For states, by states*. Washington, D.C.: National Academy Press.

Mogk.[2] We coded for four categories of data and two categories of non-data. The coding scheme is in Table 1.

Initial coding was done by research assistant Frances Straccia, with guidance from Kim Kastens. Straccia used a three-category code schema (unambiguous data, data requiring judgment call, unambiguous non-data). For PEs that Straccia found problematic, Kastens discussed each one with an EDC staff member with expertise in teaching in that grade band: Jeff Winokur for elementary, Marian Pasquale for middle school, and Ruth Krumhansl for high school. Based on notes from conversations with Straccia, Winokur, Pasquale, and Krumhansl, Kastens developed the final coding scheme described in Table 1 and finalized the coding for each PE.

Development and Coding of Occupational Profile

A modified DACUM process was used to develop the BDES occupational profile for the emerging profession of the BDES. The Oceans of Data Institute at EDC convened an expert panel of 11 such professionals representing a wide range of employment sectors for a two-day work session on August 15–16, 2014. Under the guidance of EDC facilitators, the group agreed on a definition of the target profession: "The big-data-enabled specialist is an individual who wrangles and analyzes large and/or complex data sets to enable new capabilities including discovery, decision support, and improved outcomes." Next, the group identified a set of tasks performed by such specialists, grouped them into a set of seven overarching duties, and arrayed the tasks and duties into a matrix. Finally, the group identified skills and knowledge, equipment/tools/supplies, and behaviors that are characteristic of a BDES. In the weeks following the expert panel convening, the elements of the draft BDES profile were shaped into a validation survey that was completed by approximately 100 additional big-data-enabled professionals. A few profile elements that were not affirmed by the validation survey were dropped, and the BDES profile was finalized. The final BDES profile is available for download at http://oceansofdata.org/our-work/profile-big-data-enabled-specialist.

For each task in the BDES occupational profile, the NGSS PEs were scanned to seek alignment. Because there are so many tasks, this was done holistically, without attempting to develop explicit inclusion and exclusion criteria for each task. Where alignment was found, it was noted as "1" in the spreadsheet, and the number of such hits was summed for each task. The findings are reported in terms of tasks: well-represented in the NGSS (>6 PEs), sparse in the NGSS (1–5 PEs), or absent in the NGSS (no PEs). Because the language of the NGSS and the BDES profile is so different, interpretation was involved in this matching process, and the reader should attend to the broad categories rather than reading too much into the exact numbers.

Findings

Distribution of Practice #4 by Grade Band and Discipline (Tables 2 & 3)

The NGSS maps each PE to one and only one of the eight SEPs, so the 206 PEs collectively present a total of 206 opportunities to probe students' mastery of a practice. However, the eight SEPs are unevenly represented (Table 2, right-hand column). Of the 206 PEs, 24, or 12%, are mapped to *Practice #4: Analyzing and interpreting data*. This is in the middle of the pack, along with *#3: Planning & carrying out investigations* (13%), and *#7: Engaging in argument from evidence* (12%). Far more strongly represented are *#2: Developing & using models* (21%) and *#6: Constructing explanations & designing solutions* (22%). The most lightly represented SEPs are *#1: Asking questions & defining problems* (5%),

[2] Manduca, C., & Mogk, D. W. (2002). *Using data in undergraduate science classrooms.* Retrieved from http://d32ogoqmya1dw8.cloudfront.net/files/usingdata/UsingData.pdf

#5: Using mathematics & computational thinking (7%), and *#8: Obtaining, evaluating & communicating information* (8%).

The SEPs are also unevenly distributed across the grade bands (Table 2, middle 3 columns). Of the elementary SEPs, 10 out of 76, or 13%, are mapped to *Practice #4*. This increases to 15% in middle school and then drops abruptly to 7% in high school. Two other data-relevant patterns emerge from the grade-band portion of Table 2: *SEP #5: Using mathematics and computational thinking* is almost absent in elementary and middle school (1% and 3% respectively) and then jumps to 17% in high school, suggesting that quantitative treatment of numerical data isn't really expected until high school. *SEP #3: Planning & carrying out investigations* is very strongly represented in elementary school (22% of PEs), and then drops off in middle and high school (9% and 7% respectively), raising the possibility that student data collection may mostly occur in elementary school in NGSS-aligned schools.

SEPs are also unevenly distributed across the NGSS disciplines (Table 3). For example, Physical Sciences is rich in opportunities to engage with *Practice #3: Planning and carrying out investigations*, devoting 25% (18/72) of its PEs to this single practice. *Practice #4: Analyzing & interpreting data* is found in 16% of the Earth/Space PEs, 14% of the engineering design PEs, 13% of the Life Sciences PEs, and only 7% of the Physical Sciences PEs.

Abundance of Data under the Broader Definition (Table 4 and Figure 1)

Seeking data across all of the practices, rather than only in *Practice #4*, reveals that data are quite a bit more abundant in the NGSS. Whereas *Practice #4* comprises only 12% of PEs, all four data categories together comprise 46% of the PEs. Even looking at only data categories 1 + 2—which involve no subjective judgment because the word "data" appears in the PE text, the SEP title, and/or the SEP text—data are involved in 25% of the PEs.

When considering all data categories together, we no longer see the dramatic drop-off of data-using in high school that we saw when only *Practice #4* was considered (Table 2). Data-use under this broader definition occurs in 46% of elementary PEs, 49% of middle school PEs and, 44% of high school PEs. At least some of the difference is because in high school, students are expected to go beyond mastery of data manipulation procedures and use data in more sophisticated ways, including constructing explanations (*Practice #6*), evaluating solutions (*Practice #6*), developing models (*Practice #2*), and supporting or testing claims (*Practice #7*). For example:

- HS-ETS1-3. "Evaluate a solution to a complex real-world problem based on prioritized criteria and trade-offs that account for a range of constraints, including cost, safety, reliability, and aesthetics, as well as possible social, cultural, and environmental impacts." The NGSS maps this to the *Evaluating solutions* side of *Practice #6.* We coded it as D-4, reasoning that prioritizing criteria would require use of some quantitative data on cost, safety, and/or reliability.
- HS-LS4-4. "Construct an explanation based on evidence for how natural selection leads to adaptation of populations." The NGSS maps this to the Construct Explanations side of *Practice #6*. We coded this as D-4, reasoning that at the high school level, evidence for natural selection should include examination of data, such as the well-studied Galapagos finches.
- HS-ESS2-6. "Develop a quantitative model to describe the cycling of carbon among the hydrosphere, atmosphere, geosphere, and biosphere." This is NGSS *Practice #2: Develop and use models*. We coded this as D-4, reasoning that building a quantitative model requires putting in quantitative parameters or relationships or constraints, which would be based in data. If it had said "use" a model, rather than "develop," the student would not necessarily have needed to engage with the underlying data, and we would have coded it as ND-1.
- HS-ESS1-5. "Evaluate evidence of the past and current movements of continental and oceanic crust and the theory of plate tectonics to explain the ages of crustal rocks." The NGSS maps this to

Practice #7: Engage in argumentation from evidence. We coded this as D-4, reasoning that the appropriate evidence would include data maps of earthquake locations, crustal age, etc.

These examples of D-4 at the high school level involve professionally collected data, data that the students did not collect themselves. This was generally the case when empirical evidence was called for in developing models, evaluating evidence, designing solutions, or constructing explanations at the high school level.

Comparison of NGSS and the ODI BDES Occupational Profile (Table 5, Figure 2)

The BDES profile includes seven “duties,” numbered 1–7 below. Each duty includes multiple “tasks,” designated with a number and letter. The discussion below takes up each duty in the order in which they appear in the BDES profile. Tables 5a, 5b, and 5c enumerate tasks that are well represented, sparsely represented, and absent in the NGSS.

1. Defines the Problem: There is a moderate degree of overlap between the NGSS and the BDES profile tasks in this duty. *Task 1B: Determine the stakeholders' needs* shows up in some of the ETS PEs, for example: “HS-ETS1-1. Analyze a major global challenge to specify qualitative and quantitative criteria and constraints for solutions that account for societal needs and wants.” *Task 1C: Articulate the question*, is well aligned with NGSS *Practice #2: Ask questions*. *Tasks 1E: Translates question into a research plan*, and *1F: Designs the experiment*, align well with the insistence of NGSS *Practice #3* that students shall plan as well as carry out Investigations. In our professional development work with middle and high school teachers around the NGSS, we have found that it can be a heavy lift to go from having students carry out an investigation to having them plan it as well. With respect to *Task 1G: Develops deep domain knowledge of data source*, one could consider that all of the disciplinary core ideas of the NGSS have the potential to build domain knowledge under various types of data.

2. Wrangles Data: The overlap between this duty and the NGSS is highly uneven. *Task 2D: Collects data*, is one of the best represented practices in the entire occupational profile, reflected in at least 24 of the 206 PEs. A skillful teacher could find many opportunities to require *Task 2I: Synthesizes data*, in an NGSS-infused curriculum, but it shows up explicitly only rarely, for example, in 3-ESS2-2, Obtain and combine information to describe climates in different regions of the world. The only form of *Task 2L: Perform data visualization* explicitly mentioned is graphing, as in 5-PS1-2: Measure and graph quantities to provide evidence that regardless of the type of change that occurs when heating, cooling, or mixing substances, the total weight of matter is conserved. The other tasks of *Duty 2: Wrangle data*, are absent from the NGSS. Two that seem of potentially high value are *2G: Identifies outliers and anomalies*, and *2N: Documents the process*. Both are found in many existing instructional materials and would be a good fit for an NGSS-aligned instructional program.

3. Manages Data Resources: This data science duty is completely absent from the NGSS. In the NGSS, data is portrayed as a means to an end (improving a design solution, for example, or constructing an explanation), rather than as a valuable resource to be protected and managed.

4. Develops Methods and Tools: Only one of the BDES profile tasks in this duty is represented in the NGSS: *Task 4E: Runs simulations*. There are four PEs aligned with this task, for example, HS-ETS1-4, Use a computer simulation to model the impact of proposed solutions to a complex real-world problem with numerous criteria and constraints on interactions within and between systems relevant to the problem. Interestingly, all four of them were coded as ND-1 rather than in one of the data categories, because we weren't convinced that students would need to dig all the way down to the empirical data to use or revise the computational model/simulation, and we had chosen to exclude model output from the “data” category.

5. Analyzes Data: Of the seven BDES profile duties, this one is the best represented in the NGSS, which is not surprising in that the NGSS devotes an entire practice to *Analyzes & interprets data*. It is a great strength of the NGSS that every single PE that begins “Analyze…data….” aligns with BDES profile

Task 5H: Answers the question (e.g., insights drawn from results). In other words, every one of the PEs that probes *SEP #4* requires students to answer meaningful questions about the real world, about a physical system, or living system or earth system. There are no PEs along the lines of "calculate the mean and standard deviation of this data," no items that exist purely in the frame of reference of the data without reference to the real world of which the data is merely a representation. As for the other tasks in this duty, it is difficult to know exactly which tasks would be exercised without seeing the data and the prompt, but our judgment is that the majority of the "Analyze….data…" performance expectations would call for *Task 5A: Develop analysis plan* and *5B: Applies methods and tools.* The "method" may be as simple as calculating an average, and the "tool" as simple as a sheet of graph paper, but they are methods and tools nonetheless. *Task 5D: Evaluates results of the analysis* (e.g., significance, effect, size) comes into play in a smaller number of PEs in which students are asked to judge whether an observed effect is sufficient, for example, MS-ETS1-3: "Analyze data from tests to determine similarities and differences among several design solutions to identify the best characteristics of each that can be combined into a new solution to better meet the criteria for success.*" Task 5G: Compares results with other findings* is sparsely represented, and the other tasks are absent.

6. Communicates Findings: Analyzing the NGSS PEs revealed a modest number of instances of *Task 6F: Prepares visualizations*; *6H: Articulates conclusions*; *6I: Contrasts alternative approaches and past results*; and *6J: Provides recommendations based on results*. However, for this particular duty, our methodology understates the amount of time spent communicating findings in an NGSS-infused classroom. Many, if not most, of the investigations that would be common in NGSS-aligned instruction culminate in expressing one's findings in words and graphics, so we would expect to see plenty of *6B: Compiles report*; *6C: Describes the problem, methods, and analysis*; and *6K: Tells "data story" to convey insight.*

7. Engages in Professional Development: Students aren't professionals so this duty is not, strictly speaking, relevant to them. But *Task 7A: Seeks out mentors* and *7G: Mentors others* would be good to foster in K–12. As schools and instructional designers figure out how to implement the NGSS's cross-cutting concepts, we may see the emergence of *7H: Engages in cross-discipline training*.

Discussion & Recommendations

- It is not well known among the K–12 education community that the NGSS PEs are unevenly distributed across the eight science/engineering practices, from a low of 5% of PEs devoted to *Practice #1: Asking questions and defining problems* to a high of 22% of PEs devoted to *Practice #6: Constructing explanations & designing solutions*. Instructional designers may want to take this balance into account. It remains to be seen how the state assessments will weight the various practices.
- The community of people trying to teach with and design curricula around the NGSS need to have a very serious conversation around the term "evidence" in collaboration with experts in epistemology. The NGSS PEs use this term liberally, but it seems to mean different things in different contexts. In particular, "evidence" seems to be tightly coupled to data in some PEs but not so in others. A student- and teacher-friendly typology of what kinds of evidence are recognized in science would be a profound contribution to the field.
- There is a gap between the kinds of small, student-collected data sets that are envisioned in *Practice #3: Plan and carry out investigations*, and the kinds of scientists' data sets that will be needed in many cases to evaluate solutions (*Practice #6*), develop quantitative models (*Practice #2*), and test claims (*Practice #7*). Research and development is needed to understand and then develop curricula and teacher supports that build these challenging practices upwards from the data skills and understandings that students developed in hands-on investigations with small data sets.
- A science education aligned with the NGSS would provide a foundation for many aspects of a career in data science. Relative to current practice, the NGSS represents an improvement in several data-

science relevant directions. These include its insistence on having students plan as well as carry out investigations that include collecting and interpreting data, the use of data to evaluate and refine design solutions, and the insistence that analysis of data must lead to insights about the real world outside the frame of reference of the data.

- Some aspects of data science are poorly represented in the NGSS. Some of these could best be left to higher education and specialized training, but others could be usefully infused into K–12 education, including the rudiments of data management (e.g., the ethics of data about people) and more attention to data quality.
- Science class is not the only K–12 venue for learning about and with data. It would be useful to do a similar analysis with the math portion of the Common Core State Standards, and with appropriate social studies and computer science standards documents. Science classrooms are an excellent setting in which to build data-using skills, but a coordinated multidisciplinary approach across the entire K–12 system would be better.
- Keep in mind that the NGSS are aspirational; the NGSS should not be read as a description of the students who will be pouring forth from the K–12 system in the immediate future. There remains a large and poorly mapped gap between the *status quo* and NGSS-aligned instruction, and then a further gap between NGSS-aligned instruction and the full suite of capabilities needed to fuel the big-data economy of the 21st century. More work is needed to “unpack” the tasks articulated by experts and better understand the foundational skills that undergird the ability to work with big data, so as to advance instruction from today’s status quo towards NGSS-aligned teaching and learning.

Acknowledgments

I thank Frances Straccia for developing the spreadsheet, deciphering the practices associated with each PE, and doing the initial coding of PEs. For providing judgment calls about how specific problematic PEs would most likely be taught in current teaching practice, I thank Jeff Winokur (elementary PEs), Marian Pasquale (middle school PEs), and Ruth Krumhansl (high school PEs). The occupational profiling workshop was ably facilitated by Joe Ippolito and Joyce Malyn-Smith. This study was funded by EDC’s Oceans of Data Institute.

Table 1: Schema for coding NGSS performance expectations

Data-1	PE and SEP both use the word "data."
Data-2	PE does not use the term "data" but the SEP does. Moreover, the syntax of the SEP requires data use, e.g., "Conduct an investigation to **produce data** to serve as the basis for evidence that meet the goals of an investigation."
Data-3	PE does not use the term "data" but the SEP does. However, the syntax of the SEP allows the use of data or other options, e.g., "Support an argument with evidence, **data,** or a model." In such cases, we relied on the judgment of an experienced teacher as to whether data would be likely to be used in current teaching practice.
Data-4	Neither PE nor SEP uses the term "data," but another wording suggested that the student was engaged in making methodical, purposeful, empirical observations of reality or interpreting such observations as made by others. In making this judgment: • knowledge from every day experience was not coded as data (e.g., birds make nests, things drop down when released) • output from computer simulations was not coded as data.[3] • Keyword **"statistics":** usually requires data. • Keyword **"graph"** (used as a verb): implies data is what's being graphed • Keyword **"patterns"** usually implies patterns in data. • Keyword **"observe"** or **"observation":** we looked for indication that the students are making methodical observations of the real world and are recording the results, i.e., are making first inscriptions. • Keyword **"information":** we considered the likelihood that for this grade band the information would be at least partly in the form of graphs, maps, tables of numbers, etc. • Keyword **"evidence"** presented great difficulty. We considered the likelihood that for this grade band, the evidence considered would be at least partly in the form of graphs, maps, charts, or numbers—as opposed to narratives or purely verbal statements. "Empirical evidence" sounds like it should mean data, but is not always clearly separable from plain "evidence" in actual PEs
ND-1	None of the above indicates that students are using data.
ND-2	PE does not use the term "data." The SEP does use the term "data" but the syntax of the SEP allows the use of data or other options (e.g., "Support an argument with evidence, **data,** or a model) and our teacher-coder's judgment was that most teachers would not teach to this PE using data. The NGSS's "clarification statement" or "assessment boundary" can help to separate ND-2 from Data-3 by suggesting whether or not the standards writers envisioned that students would be collecting or considering data.

[3] Whether or not output from computational models should be called "data" is controversial among scientists. See Manduca & Mogk (2002), op. cit.

Table 2: Distribution of PEs across the practices by grade band

Practice (SEP)	Elementary	Middle School	High School	K–12
1. Asking questions & defining problems	7% (5/76)	5% (3/59)	3% (2/71)	5% (10/206)
2. Developing & using models	16% (12/76)	27% (16/59)	23% (16/71)	21% (44/206)
3. Planning & carrying out investigations	22% (17/76)	9% (5/59)	7% (5/71)	13% (27/206)
4. Analyzing & interpreting data	**13% (10/76)**	**15% (9/59)**	**7% (5/71)**	**12% (24/206)**
5. Using mathematics & computational thinking	1% (1/76)	3% (2/59)	17% (12/71)	7% (15/206)
6. Constructing explanations & designing solutions	20% (15/76)	20% (12/59)	25% (18/71)	22% (45/206)
7. Engaging in argument from evidence	12% (9/76)	14% (8/59)	11% (8/71)	12% (25/206)
8. Obtaining, evaluating & communicating information	9% (7/76)	7% (4/59)	7% (5/71)	8% (16/206)

Note: Numerators are the number of PE's using the indicated SEP and falling in the indicated grand band. Denominators are the total number of PE's in that grade band.

Table 3: Distribution of PEs across the practices by disciplines

Practice (SEP)	Earth & Space Science (ESS)	Engineering, Technology & Applications of Science (ETS)	Life Sciences (LS)	Physical Sciences (PS)
1. Asking questions & defining problems	2% (1/57)	21% (3/14)	2% (1/63)	7% (5/72)
2. Developing & using models	23% (13/57)	14% (2/14)	22% (14/63)	21% (15/72)
3. Planning & carrying out investigations	7% (4/57)	7% (1/14)	6% (4/63)	25% (18/72)
4. Analyzing & interpreting data	**16% (9/57)**	**14% (2/14)**	**13% (8/63)**	**7% (5/72)**
5. Using mathematics & computational thinking	7% (4/57)	7% (1/14)	8% (5/63)	7% (5/72)
6. Constructing explanations & designing solutions	21% (12/57)	29% (4/14)	24% (15/63)	19% (14/72)
7. Engaging in argument from evidence	12% (7/57)	7% (1/14)	19% (12/63)	7% (5/72)
8. Obtaining, evaluating & communicating information	12% (7/57)	0% (0/14)	6% (4/63)	7% (5/72)
Fraction of PEs	28% (57/206)	7% (14/206)	31% (63/206)	35% (72/206)

Note: In the main part of the table, numerators are the number of PE's using the indicated SEP and falling within the indicated discipline; denominators are the total number of PE's in that discipline. In the bottom row, numerators are the number of PE's in that discipline and denominators are the total number of PE's in the NGSS.

Table 4: Distribution of data/non-data PEs across grade bands

Data use category	Elementary	Middle School	High School	K-12
Data-1	9% (7/76)	20% (12/59)	4% (3/71)	11% (22/206)
Data-2	20% (15/76)	7% (4/59)	10% (7/71)	13% (26/206)
Data-3	1.3% (1/76)	0% (0/59)	3% (2/71)	2% (3/206)
Data-4	16% (12/76)	22% (13/59)	27% (19/71)	21% (44/206)
NonData-1	50% (38/76)	51% (30/59)	55% (39/71)	52% (107/206)
NonData-2	4% (3/76)	0% (0/59)	1% (1/71)	2% (4/206)
All Data	46% (35/76)	49% (29/59)	44% (31/71)	46% (95/206)
All NonData	54% (41/76)	51% (30/59)	56% (40/71)	54% (111/206)

Notes: Data use categories are as defined in Table 1. Numerators are the number of PEs in the indicated grade band that were coded in the indicated category. Denominators are the total number of PEs in that grade band.

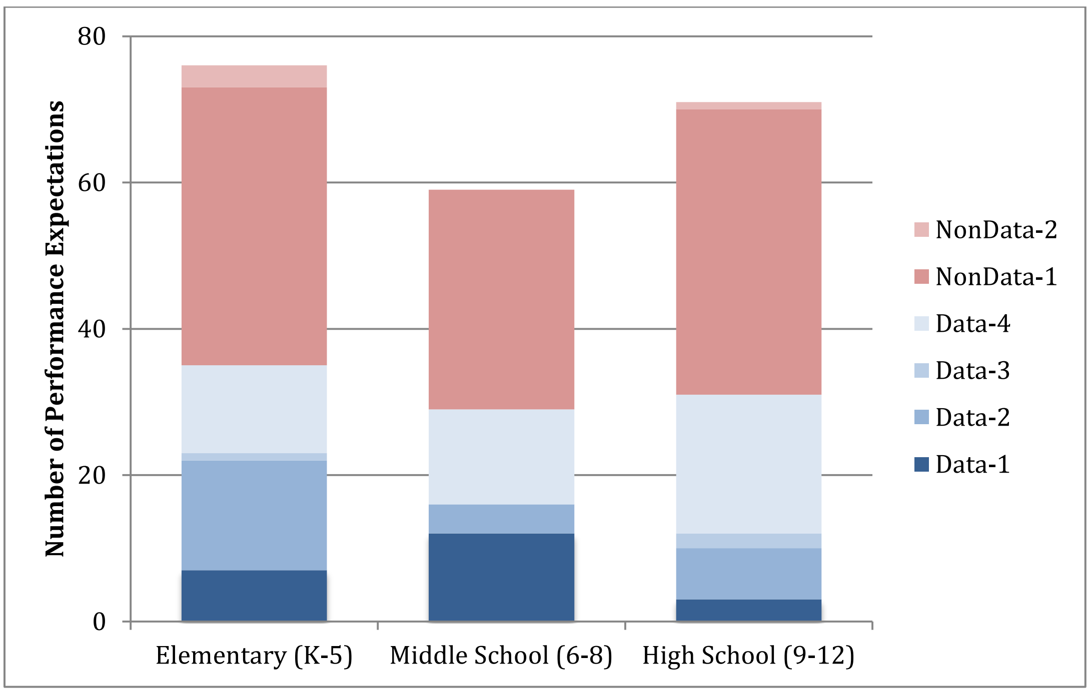

Figure 1: Distribution of data/non-data PEs across grade bands shown graphically. Middle school has fewer PEs because it covers fewer grades. For Data-1 and Data-2, the wording of the PE and/or the SEP leaves no doubt that students are required to engage with data. For Data-3 and Data-4, some judgment was required on the part of the coders, either because data was one of several options offered (Data-3) or because use of data had to be inferred from other terminology such as “measure,” “observations,” “empirical evidence,” or “statistics” (Data-4).

Table 5a: Occupational profile tasks that are well represented in the NGSS

1. Defines the Problem
1B. Determines stakeholders' needs
1C. Articulates the question
1E. Translates question into a research plan
1F. Designs the experiment
1G. Develops deep domain knowledge of data source
2. Wrangles Data
2D. Collects data
3. Manages Data Resources
n/a
4. Develops Methods and Tools
n/a
5. Analyses Data
5A. Develops analysis plan
5B. Applies methods and tools
5D. Evaluates results of the analysis (e.g., significance, effect, size)
5H. Answers the question (e.g., insights drawn from results)
6. Communicates Findings
6B. Compiles report **
6C. Describes the problem, method, and analysis **
6I. Contrasts alternative approaches and past results
6K. Tells "data story" to convey insight **
7. Engages in Professional Development
n/a

Note: **These tasks were not found explicitly stated in the NGSS PEs. They have been elevated to the "well-represented" category because we think they would be common within investigations carried out under the "Plan and carry out investigations" SEP.

Table 5b: Occupational profile tasks that are sparsely represented in the NGSS

1. Defines the Problem
n/a
2. Wrangles Data
2J. Synthesizes data
2K. Defines new metrics/attributes based on accessible data
2L. Performs data visualization
3. Manages Data Resources
n/a
4. Develops Methods and Tools
4E. Runs simulations
5. Analyzes Data
5G. Compares results with other findings
6. Communicates Findings
6F. Prepares visualizations
6H. Articulates conclusions
6J. Provides recommendations based on results
7. Engages in Professional Development
n/a

Table 5c: Occupational profile tasks that are absent from the NGSS

1. Defines the Problem
1A.Identifies stakeholders
1D. Aligns study to organizational goals and objectives
1H. Discerns whether big data is needed to solve problem
1I. Identifies resources (e.g., experts, software)
1J. Performs gap analysis
1K. Assesses risk and bias involved in conducting study/project
1L.Communicates cost/risks of study to stakeholders
1M. Negotiates plan, including deadlines and budgets
1N. Creates requirement document (sign-off)
2. Wrangles Data
2A. Performs data exploration
2B. Identifies data
2C. Creates the data dictionary
2E. Assesses the extent/methods to clean the data
2F. Maps data across heterogeneous sources
2G. Identifies outliers and anomalies *
2H. Cleans data
2I. Transforms data
2M. Writes software to automate tasks
2N. Documents the process *
3. Manages Data Resources
3A. Manages data life cycle
3B. Conducts capacity planning of resources
3C. Complies with legal obligations
3D. Applies ethical standards *
3E. Identifies tools that may be needed for purchase or modification
3F. Protects data and results *
3G. Determines access to the data
3H. Designs ETL workflow [extract, transform, load: read data from one database, transform it into another form, and load it into another database]
3I. Implements ETL workflow
3J. Stores the data *
3J. Upserts the data [insert or update depending on condition]
4. Develops Methods and Tools
4A. Researches current methods/ models
4B. Extends existing methods/ models, if possible
4C. Selects tools/software/ programming environment
4D. Develops new methods/ models
4F. Iterates correctness and scalability of methods/ models *
4G. Validates methods/ models with test cases
4H. Disseminates methods/ models for peer review
4I. Documents methods/ models
5. Analyzes Data
5C. Conducts exploratory analysis (e.g., identifies anomalies, outliers, bias in sampling; visualizes)
5E. Estimates precision and accuracy of answer
5F. Determines level of confidence in results *
5I. Submits preliminary findings for peer review
5J. Documents preliminary findings
6. Communicates Findings
6A. Selects documentation media (e.g., dashboard, PowerPoint, e-mail)

6D. Identifies limitations (e.g., data use, data application methods) *
6E. Scopes data narrative based on time, depth, and method
6G. Guides interpretation

7. Engages in Professional Development
7A. Seeks out mentors
7B. Stays current on emerging technologies, data, types, and methods
7C. Attends relevant big data conferences
7D. Contributes new knowledge to the field
7E. Maintains professional library
7F. Participates in professional organizations
7G. Mentors others
7H. Engages in cross-discipline training
7I. Articulates value of big data activities to other departments/ functions of the organization
7J. Articulates evolving role of big data in supporting organizational goals

Note: * These tasks are present in the BDES profile but not in the NGSS PEs, and we think they would be valuable additions to the K–12 curriculum, in an age-appropriate form.

Figure 2. Comparison of NGSS and the ODI BDES Occupational Profile

DUTIES	TASKS											
1. Defines the Problem	1A. Identifies stakeholders	1B. Determines stakeholders' needs	1C. Articulates question	1D. Aligns study to organizational goals and objectives	1E. Translates question into research plan	1F. Designs experiment	1G. Develops deep domain knowledge of data source	1H. Discerns whether big data is needed to solve problem	1I. Identifies resources (e.g., experts, software)	1J. Performs gap analysis	1K. Assesses risk and bias involved in conducting study/ project	1L. Communicates cost/risks of study to stakeholders
	1M. Negotiates plan, including deadlines and budgets	1N. Creates requirement document (sign-off)										
2. Wrangles Data	2A. Performs data exploration	2B. Identifies data	2C. Creates data dictionary	2D. Collects data	2E. Assesses the extent/methods to clean the data	2F. Maps data across heterogeneous sources	2G. Identifies outliers and anomalies	2H. Cleans data	2I. Transforms data	2J. Synthesizes data	2K. Defines new metrics/ attributes based on accessible data	2L. Performs data visualization
	2M. Writes software to automate tasks	2N. Documents process										
3. Manages Data Resources	3A. Manages data life cycle	3B. Conducts capacity planning of resources	3C. Complies with legal obligations	3D. Applies ethical standards	3E. Identifies tools that may be needed for purchase or modification	3F. Protects data and results	3G. Determines access to data	3H. Designs ETL workflow	3I. Implements ETL workflow	3J. Stores data	3K. Upserts data sources	
4. Develops Methods and Tools	4A. Researches current methods/ models	4B. Extends existing methods/ models, if possible	4C. Selects tools/ software/ programming environment	4D. Develops new methods/ models	4E. Runs simulations	4F. Iterates correctness and scalability of methods/ models	4G. Validates methods/ models with test cases	4H. Disseminates methods/ models for peer review	4I. Documents methods/ models			

Abundant in NGSS (1B, 1C, 1E, 1F, 1G, 2D) · Potentially (implicitly) abundant in NGSS · Sparse in NGSS (2J, 2L, 4E) · Absent from NGSS

DUTIES	TASKS											
5. Analyzes Data	5A. Develops analysis plan	5B. Applies methods and tools	5C. Conducts exploratory analysis (e.g., identifies anomalies, outliers, bias in sampling; visualizes)	5D. Evaluates results of the analysis (e.g., significance, effect, size)	5E. Estimates precision and accuracy of answer	5F. Determines level of confidence in results	5G. Compares results with other findings	5H. Answers the question (e.g., insights drawn from results)	5I. Submits preliminary findings for peer review	5J. Documents preliminary findings		
6. Communicates Findings	6A. Selects document-tation media (e.g., dash-board, PowerPoint, e-mail)	6B. Compiles report	6C. Describes problem, method, and analysis	6D. Identifies limitations (e.g., data use, data application methods)	6E. Scopes data narrative based on time, depth, and method	6F. Prepares visualizations	6G. Guides interpre-tation	6H. Articulates conclusions	6I. Contrasts alternative approaches and past results	6J. Provides recommen-dations based on results	6K. Tells “data story” to convey insight (e.g., talks to CEO)	
7. Engages in Professional Development	7A. Seeks out mentors	7B. Stays current on emerging technologies, data types, and methods	7C. Attends relevant big data conferences	7D. Contributes new knowledge to the field	7E. Maintains professional library	7F. Participates in professional organizations	7G. Mentors others	7H. Engages in cross-discipline training	7I. Articulates value of big data activities to other depart-ments/ functions of organi-zation	7J. Articulates evolving role of big data in supporting organizational goals		

Abundant in NGSS | Potentially (implicitly) abundant in NGSS | Sparse in NGSS | Absent from NGSS