

3 | Using data in the research process

Jole Seroff

Traditional literacy, digital literacy, visual literacy, civic literacy, news literacy ... and now data literacy as well? It can take some creative thinking to figure out how all these skills fit together into a coherent curriculum for 21st century school libraries.

Many of us work with students at different junctures in their research, as they put together projects and papers or prepare debates and presentations. I've identified a few key moments in the research process that provide opportunities for interventions focused on building data literacy skills.

This chapter will outline flexible, modifiable activities you could use with your students to build background knowledge and learn to ask questions to make sense of the statistics they discover. These activities are meant to serve as starting points for incorporating more critical thinking about data in your research instruction. As you grow familiar with data literacy themes, you will likely find other connections and opportunities that are relevant for your own practice.

Building background knowledge with statistical benchmarking

The initial stages of the research process have always been particularly enjoyable for me. When students are just beginning a new project, everything is possible. They have so much to learn and don't yet know where their research might take them. Despite my own enthusiasm for this moment in the process, I've heard from colleagues and have seen in my own students that they often struggle to find a starting place for their research,

to define a topic clearly, or to narrow a topic appropriately (Kuhlthau 2004).

This can be an excellent moment to infuse your lessons with data literacy skills. Students may begin with a general area of inquiry, something quite broad, or even perhaps something too narrow for the requirements of their project. Whenever they are exploring a new topic, I encourage our students to begin with background reading, which I sometimes call “stepping stone” reading. I ask them to begin with general sources, novice-level sources, and focus their attention on getting a sense of the key aspects of their topic. I want our students to understand that the first reading they do is not meant to provide their final answer, but rather to help them better understand their question.

In other words, it will build a bridge from their starting point to deeper understanding. Background reading sheds light on what they will need to know to fully answer a question, and it may help to narrow or redirect a question.

To start their thinking, I might give students an article I have selected. Alternatively, I might point them to general encyclopedias or news coverage, or even ask them to try some Google searches and examine the snippets (or brief summary) displayed on the search results page in order to begin to get a sense of the topic. In order to focus their reading and stimulate active thinking, I ask students to look for these key things: *terms of art*, *experts and organizations focused on the topic*, and *benchmark statistics*.

Terms of art and statistical benchmarks

Terms of art refers to language that has a specific meaning within a particular field or profession. For example, *hypothesis* has a specialized meaning in a science classroom, just as *formula* does in a

math class. As they're getting familiar with a new topic, students can be confused by specialized terminology, so it is helpful to put them on the lookout for it through stepping stone readings. I ask them, "What are the terms used by people who know more about the topic than we do?" Once they have identified expert vocabulary related to their topic, they can use those terms in their future searches, and they're likely to get higher-quality results, including material written by or for experts in the field.

Similarly, identifying experts and major organizations associated with their topic and perhaps beginning to recognize whether these experts have a particular orientation or perspective with regard to the topic will help move them forward thoughtfully in their research. In the early stages of the research process students are building an understanding of their topic, including, perhaps, its leading thinkers, its controversial aspects, and its particular jargon. This knowledge will enable them to treat their topic with more nuance and comprehension.

How can we integrate data literacy into our practice of building background information on a topic? Gathering benchmark statistics (Best 2013), or reliable statistics against which we can weigh new data we encounter, provides an opportunity to bring data literacy practice into the process of background reading. A benchmark statistic is simply a standard or point of reference against which other things may be compared or assessed. For example, think how often you heard how many people voted for a particular candidate in a given year. If you know that there are approximately 325,000,000 people living in the United States (U.S. Census 2016), you can rapidly contextualize voter turnout. If you want to dig in further to these numbers, you might consider that not all Americans are eligible to vote, because of age and other factors. In October of 2016, *Politico* reported that the American electorate had surpassed 200 million registered voters (Goldmacher 2016). With that in mind, you have an even clearer grasp on the significance of the raw number of votes a candidate

received. As our students become more data literate, they must develop the habit of looking for similar benchmarks to help them make sense of the numbers they encounter.

Brainstorming

One engaging way to use benchmark statistics with students is through brainstorming. When you're introducing a new topic, ask students to imagine: what statistical benchmarks would help me get a grasp of this topic, or be useful for comparing and evaluating new information I encounter? Encourage students to focus their questions on numbers, amounts, and percentages. Remind them that statistics are descriptive of things that are countable.

Here's an example that could be of interest to teens: according to the Centers for Disease Control and Prevention motor vehicle crashes are the leading cause of deaths for teens in the U.S. (CDC 2016). Brainstorm: what kind of benchmarks will help contextualize this statistic?

For example, students may want to know:

- » **How many teens die in traffic accidents every year?**
- » **Do they suffer traffic fatalities at a higher rate than adults do?**
- » **What are the other major causes of deaths in teens and adults?**
- » **Are teens more likely to be drivers or passengers in fatal accidents?**
- » **Considering traffic accidents in general, how often do fatalities occur?**
- » **How effective are measures such as use of seat belts and extended driver's education at preventing traffic fatalities?**

This activity can be done as a 10-minute warm up or even as a homework assignment. This kind of group discussion can fuel and contextualize the early stages of research. By generating a list of questions and tracking down relevant benchmark statistics, students engage in active thinking. This process provides insight for shaping a research topic and clarifies directions for pursuing further research.

Once your students generate a good list of data-related questions, they can employ those questions as guides for further research. As they become more fluent in their use of data, they will understand that although there aren't always statistics available on every topic, many of the questions they pose likely do have accessible answers. You could spend a 15- or 20-minute chunk of your lesson searching for benchmark statistics with a focus on selecting sources for their accuracy and currency. Information literate students will need to build their capacity to identify sources that provide reliable, relevant data.

Questioning the source

It is precisely in order to counter short-cut thinking and biased assumptions that we want to guide our students to think more critically about data when they read independently and encounter data embedded in text. It is important for students to recognize that while there may be data referenced in a source, it may have been extracted from its original dataset, context, or source. It may be so decontextualized that the argument it was intended to support may not even be explicitly expressed. Traditional information literacy skills provide valuable structure for librarians and students working with data. Consider the *source* of the data, not the location where it was found. Ask yourself:

- » **Is the source reliable?**
- » **Does the institution behind the data have an agenda that is shaping the information it presents?**
- » **When you look at multiple sources, do you find a consensus or a range of ideas?**

Students should get in the habit of asking these and other reflective questions when they encounter data just as they do when they're reading for narrative arguments.

Sample lesson

Consider a recent 50-minute class with high school seniors, in which I challenged them to develop a set of critical questions around data by working with a preselected set of resources.

In this lesson, they were responsible for reflecting on how they analyze data when it's presented as evidence in news sources. As a group we worked toward developing some rules of thumb for useful questions to ask when thinking about data.

We examined materials focused on the current migrant and refugee crisis in Africa, Europe, and the Middle East. We began by reading one source together that provided some general context for the issue: making clear distinctions between migrants and refugees and clarifying why this distinction is so important. A *refugee* is defined by the United Nations as a person fleeing her homeland due to a well-founded fear of persecution; a refugee has a right to safe asylum, and she may not be forcibly returned to her homeland. *Migrants*, however, are persons relocating out of choice, often in search of economic opportunity. Migrants are not granted a right to asylum. Through the initial group reading students learn about this distinction, and they learn that there is frequently disagreement about which individuals qualify for the refugee designation and the rights it confers. Further, they learn that a single

individual may be considered a refugee when she reaches the first destination outside of her country of origin, but if she wishes to move again, she will be viewed as a migrant and may lose the rights of refugee status when she reaches her second destination. These and other complications pose real problems when it comes to calculating statistics related to this crisis.

Before breaking up into pairs to read and analyze on their own, students received a worksheet (see following pages) for taking notes, with some questions designed to focus their thinking on data they encountered. These questions directed students to examine the source of the data and to question the methods, if available, through which the data had been collected. Students were asked to look out for those discrepancies between the data presented in different sources and to hypothesize the reason for discrepancies. Further, I asked students to make note of things that surprised them and to keep a running list of new questions that came up for them as they were reading and analyzing.

I heard students engaging actively with data as I eavesdropped on the working pairs. They were wondering how displaced persons, who might want to escape official notice or who were not likely to have a fixed abode could be effectively counted. They were noticing their own surprise at the preponderance of refugees fleeing to other countries in Africa and the Middle East, and not just to Western Europe, which had been the focus of attention in Western news media. They were beginning to question whether statistics were being leveraged for political ends by leaders in countries such as Greece and Italy, that were hard-hit by the influx of refugees. They suspected that these leaders might stand to benefit from designating the new arrivals as migrants, with limited rights, rather than as refugees deserving asylum.

**What questions do you have about the source of the data?
(Perspective)**

-
-
-

**What questions do you have about how the data were
collected? (Methods)**

-
-
-

**What surprises you? (For example, numbers that seem
especially large or small, things that conflict with your
assumptions, etc.)**

-
-
-

**What data do you find that conflicts between the sources?
Where do the sources seem to disagree?**

-
-
-

What new questions do you have about this topic, based on what you have read?

-
-
-

What conclusions can you draw about this topic based on what you've read? (For example, complete the sentence, *These sources show me that...*)

-
-
-

If you have extra time, look online for reliable sources and try to answer some of the questions you have!

When we reconvened as a group, we focused our discussion on taking their observations and generating some rules of thumb for thinking about data more generally. For example, when students pointed out that most of the sources were from Western countries and that the single resource from the Middle East emphasized a different perspective, we were able to conclude that it is important to look at data published from a wide range of viewpoints, which may include geographically or geopolitically diverse sources.

Students also noticed that some discrepancies in the data could be accounted for by differences in how the sources were defining migrants and refugees, or which specific geographic areas the statistics were focused on. We were able to extrapolate from that observation that it's critical to notice how terms are defined and what precisely is being counted.

When students are charged to articulate guidelines themselves, they become more engaged and empowered as learners. Nevertheless, there is a struggle inherent in designing lessons that are driven by student inquiry: the more outcomes are student-generated, the less predictable they are. It can be challenging to start a lesson without knowing where it will end up. Ultimately I remind myself that it is worth taking that risk. When I see my students engaging actively, I know they are developing powerful skills in critical thinking.

As students research: Interrogating the statistic

Now students are ready to move from group discussion to independent research. When students encounter a statistic used as evidence for an argument, I encourage them to interrogate the statistic. We interrogate statistics by exposing them to scrutiny and asking good questions about what conclusions can and cannot be drawn from them.

An excellent entry point for interrogating statistics is to help students learn to identify the original source of their data, to the best of their ability. It can be challenging to trace information back to its original source, but quality news outlets will mention the source behind the data. If that information isn't included, that's a red flag, and data literate students should learn to recognize that.

Another strategy is to have students look at conflicting statistics and evaluate their sources. This will give students an opportunity to practice interrogating the statistics: Can they determine the source of the conflict between the statistics? It could be because the information comes from different geographical locations or different time periods, or that terms are being defined differently, as in the case when it can be difficult to determine whether an individual is a migrant or a refugee. Ultimately, students can practice making a judgment when statistics conflict – which source do they feel is more credible? They should be able to make an argument to support this conclusion, perhaps citing the size of study or the relative bias of the source.

An activity of this sort can also provide an opportunity to practice both information literacy and statistical literacy skills. Students will need to evaluate the source for accuracy and bias. And they will also need to evaluate whether the numbers really mean what they seem to on their surface.

Comparing statistics

I used the article “Black Criminals, White Victims, and White Guilt” (Simpson 2015) from the Accuracy in Media website for our next example. This is a website that frames itself as a media watchdog through a conservative lens seen by some readers as biased. The basic claim of this article is that criminal justice reform in the U.S. is unnecessary because the rates of police violence against blacks

have been overhyped by the media and are not actually a serious problem.

First, by applying traditional information literacy skills and looking into the organization behind this article, students will learn that this organization's stated goal is to battle what it considers a liberal bias in the media, which lends its reporting a definite perspective. When students add a data literacy lens, they will notice that the figures presented in the article are offered in raw numbers, rather than percentages or rates over time. Data literate students will pay special attention to the format in which numbers are presented. It may be the case that a particular choice was made in order to shape the reader's perception of the data. They'll keep this in mind when they compare data from multiple sources.

For contrast, I looked at an article from the *Washington Post*: "Aren't more white people than black people killed by police? Yes, but no" (Lowery 2016). A little research into the source reveals that the *Washington Post* is one of the country's leading newspapers. The article explains that while in raw numbers more white people than black people are killed by police, because black people make up a much smaller percentage of the total population they are significantly overrepresented in these kinds of deaths. According to the *Washington Post*, black Americans are 2.5 times as likely as white Americans to be shot and killed by police officers. It can be helpful for students to discuss which of these articles *seems* more compelling and why.

Practice examining arguments from opposing viewpoints builds critical thinking and information literacy skills. Students learn to read actively, considering the arguments put forth by each writer. Examining the use of data in these arguments gives students experience evaluating whether the selected data really supports the conclusion of the author.

Further along in their research, students often seek assistance crafting arguments and selecting evidence to support their claims. This is another powerful moment in the research process for building data literacy skills. From their large accumulation of research material, students must now select those statistics, information, anecdotes, and facts that best inform the argument. They begin to see that data is not merely information: it is potential evidence. When it comes to using data as evidence, what orientation do we want our students to have? What guidance can we give them?

Two interrelated concepts are useful for students to learn here: cherry-picking and confirmation bias. Finally, annotations give students an opportunity to share their thinking and analysis of sources with their instructors.

Cherry-picking

It is often said by skeptics and cynics that statistics can be used to prove anything. This notion seems validated when we see the same statistics cited to support opposing sides of an argument. Where this seems to be the case, cherry-picking, or selecting only examples that support your argument while ignoring counterexamples, may be the culprit. Cherry-picking is a critical concept for students to understand because it has become increasingly common in the world of partisan journalism, politics, social media, and blogs. Students should be on the lookout for this in sources they read and should be doubly aware of it in their own research and writing. Even when a student researcher discovers evidence that aligns with her thinking or seems ideal to support her argument, she should be sure she is not missing another side to an argument. Her research should encompass a full range of perspectives. Otherwise, she's in danger

of falling prey to confirmation bias. Further, just because statistics are offered that seem to support conflicting viewpoints, that doesn't mean that both sets of statistics are equally convincing. You can set up an activity in which students examine statistics used to support conflicting arguments and then evaluate the relative strength of the two arguments.

Confirmation bias

Confirmation bias is the natural tendency to favor information which confirms our existing beliefs and preconceptions. When we encounter information that aligns with our expectations, we are less likely to think critically or question the validity of the evidence. When students are seeking information, confirmation bias can pollute their search results. Students must learn to think carefully in order to avoid doing searches where the answer is already "baked in". Because the function of a search engine is to return pages that include the terms in the search, the researcher is likely to encounter material that aligns with the ideas represented in her search. For example, a search for [black crime statistics] will produce results that show links between blacks and crime, because that is literally what the query asked the algorithm to find. Students will likely find more inclusive, nuanced, or better-balanced results if they start by searching for [U.S. crime statistics] more generally and then look closely at that data for racial breakdowns within those statistics. This approach provides more context for the information, and context is a powerful tool to combat confirmation bias. When the researcher encounters information that conflicts with her assumptions, she has the opportunity to evaluate the evidence and reconsider her position.

Annotations

Cherry-picking and confirmation bias are abstract concepts that can be challenging for young scholars to master. How can we as instructors set up an environment that encourages them to engage more rigorously with sources? Annotations may be one answer. Annotations are short paragraphs that provide a student's analysis of a given source. Having students write short annotations for statistical information they include in their papers is a way to make the students' thinking transparent, and therefore easier for the teacher to assess and respond to. While not an onerous task, annotations help to build a habit of responsible use of data.

There are several online resources for discipline-specific annotation strategies, but for general purposes, most annotations include three parts, as outlined by the Purdue Online Writing Lab (n.d.): *summarization* (an overview of the key ideas), *assessment* (source evaluation and credibility), and *reflection* (student discussion of the role and relevance of this source to their overall research and project). Please see Figure 1 for details one might assign in each section of an annotation.

On a project-by-project basis, instructors could add additional criteria or questions to the annotation guidelines for a specific assignment. By doing so, they cue students to surface and confront their tacit understandings. Writing annotations can help students reflect on their assumptions, articulate their thinking, and share their understandings with educators. For more on annotations and annotated bibliographies, including a sample, see <https://owl.english.purdue.edu/owl/resource/614/1/>.



Annotation options for students

- **Summarization** – Students write a short overview of the main ideas and concepts of the source.
- **Assessment** – Students evaluate the source, including
 - » information quality and comprehensiveness
 - » currency of information
 - » reliability of information
 - » use of data
 - » who conducted the study and their authority, expertise, or known perspectives or political slant
 - » who funded the study and their authority, expertise, or known perspectives or political slant
- **Reflection** – Students consider how this source relates to the overall constellation of research conducted:
 - » How does this source fit into the emerging thesis of their paper or project? How does it influence their argument? What value does this source add to their thinking?

Figure 1: Three annotation strategies for students.

Conclusion

In this chapter, I have suggested a number of activities that integrate data literacy skills with traditional information literacy instruction throughout the research process. From the very first steps they take in researching a topic, active, critical thinking about data will help students gain a better understanding of a particular topic while developing powerful, transferrable skills. Incorporating data literacy in our instruction creates opportunities for practicing core information literacy skills such as source evaluation. As students get experience examining the way data is used to support arguments, they will build their capacity to incorporate statistical evidence in their own writing. The data literate

student is less susceptible to misinformation and better able to weigh evidence and form her own judgments. Ultimately the goal is to develop the habit of active, critical thinking about data.

Resources

- Best, Joel. 2013. *Stat-Spotting: A Field Guide to Identifying Dubious Data*. 2nd Ed. Berkeley, CA: University of California Press.
- Centers for Disease Control and Prevention (CDC). 2016. "Teen Drivers" Get the Facts." Oct. 13. Retrieved December 20, 2016, from https://www.cdc.gov/motorvehiclesafety/teen_drivers/teendrivers_factsheet.html.
- Goldmacher, Shane. 2016. "America Hits New Landmark: 200 Million Registered Voters." *Politico*, Oct. 19. Retrieved December 20, 2016, from <http://www.politico.com/story/2016/10/how-many-registered-voters-are-in-america-2016-229993>.
- Kuhlthau, Carol Collier. 2004. *Seeking Meaning: A Process Approach To Library And Information Services*. Westport, Conn. Libraries Unlimited.
- Lowery, Wesley. 2016. "Aren't More White People Than Black People Killed By Police? Yes, But No." *Washington Post*, July 11. Retrieved December 20, 2016, from <https://www.washingtonpost.com/news/post-nation/wp/2016/07/11/arent-more-white-people-than-black-people-killed-by-police-yes-but-no/>.
- National Safety Council. 2016. "Motor Vehicle Deaths Increase by Largest Percent in Fifty Years." Feb. 17. Retrieved December 20, 2016, from <http://www.nsc.org/Connect/NSC-NewsReleases/Lists/Posts/Post.aspx?ID=103>.
- Purdue Online Writing Lab. n.d. "Annotated Bibliographies." Retrieved December 20, 2016, from <https://owl.english.purdue.edu/owl/resource/614/1/>.
- Simpson, James. 2015. "Black Criminals, White Victims, and White Guilt." *Accuracy in Media*. Feb. 9. Retrieved December 20, 2016, from <http://www.aim.org/special-report/black-criminals-white-victims-and-white-guilt/>.



If you have five minutes:

- **Remind students that statistics and numerical information needs to be verified, too!**
- **Help students try to track a statistic back to its source: who gathered this information?** Ask if knowing this makes the students more or less confident about the reliability of the statistic.

If you have fifteen minutes:

- **Ask students to complete the worksheet from this chapter about an article you choose.**

If you have thirty minutes:

- **Practice contextualizing a statistic by generating (and answering!) a list of questions that will help you situate that statistic in relation to meaningful, familiar measures.** For example, according to the National Safety Council (2016), there were 38,000 traffic fatalities in the U.S. in 2015. Some questions students might generate include: Is there disagreement about this number? Is this higher or lower than 2014? What has the trend been in the last 5 or 10 years? What about vehicular injuries, or are we only interested in deaths? Do we know what percentage of these accidents involved alcohol and drugs? Are drivers in different age groups more or less likely to be involved in fatal accidents? etc.

If you have one class period:

- **Work with students to better understand the ways in which statistics are socially constructed.** College rankings hold inherent interest for many students. Consider the factors that go into published rankings. Do they represent the elements of a college experience that would be most meaningful to you? Try evaluating and re-ranking colleges based on your own criteria and weighting.
- **Explore the cultural construction of data by closely examining virtually any statistic.** Ask questions about how the terms of the statistic are defined, i.e., what do we

mean by *disabled* or *elderly*? Inquire who constituted the pool of respondents to a survey, and what other groups might have different opinions. Consider the limitations that may have made a perfect version of a given study impossible, and the concessions or adjustments which were likely made for practical reasons. Make sure your students have an opportunity to do this independently: find their own statistic and interrogate it. Does it meet their critical standards?

If you have multiple class periods:

- **Integrate statistical literacies into a larger research process.** Examine the role of numerical information at each step of the project. Pursue background knowledge to provide context for the statistics you encounter; build a list of benchmark statistics; evaluate sources for reliability; interrogate the cultural context that informs the gathering and reporting of statistics. Soon, critical thinking about data will become second nature for your students!

