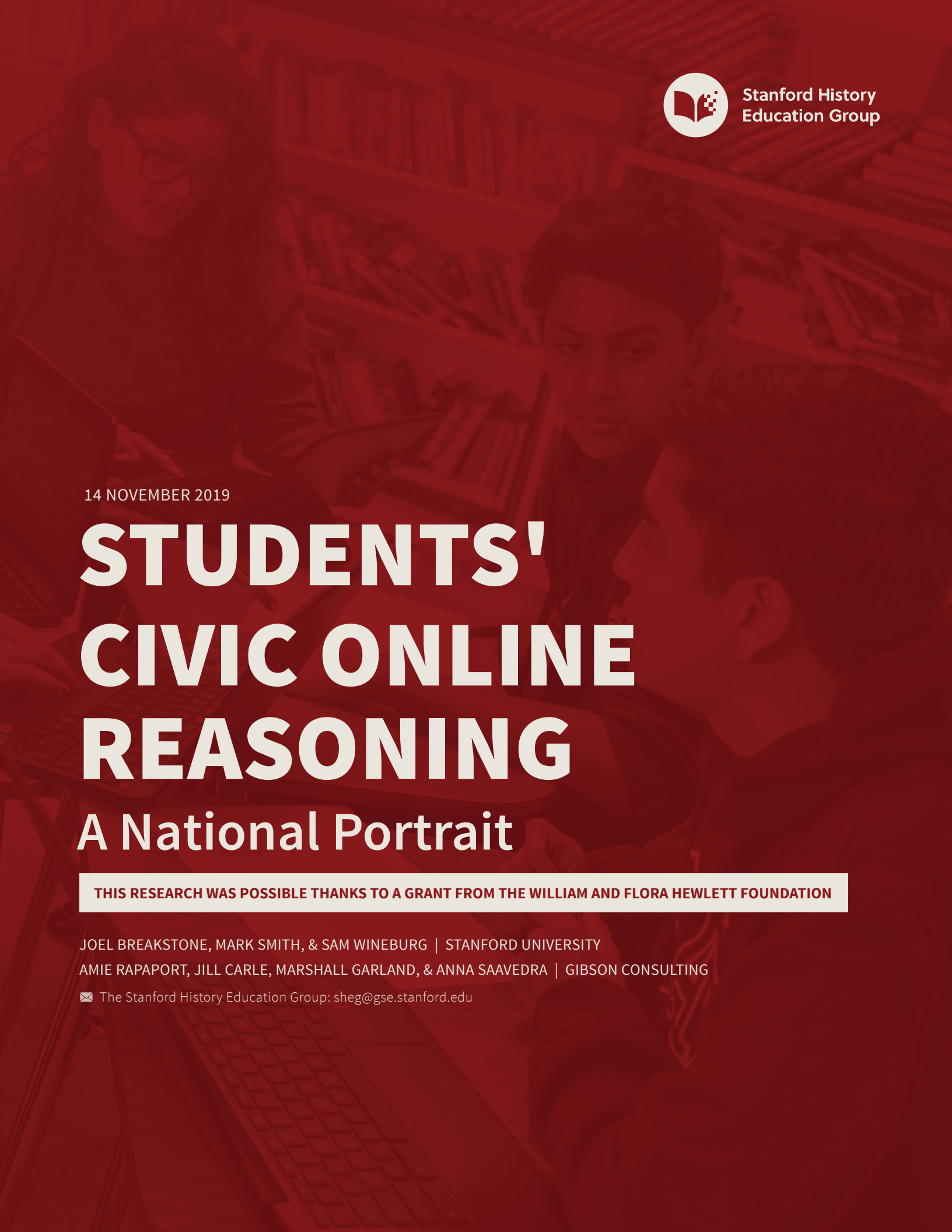


Stanford History Education Group

14 NOVEMBER 2019

STUDENTS' CIVIC ONLINE REASONING

A National Portrait

THIS RESEARCH WAS POSSIBLE THANKS TO A GRANT FROM THE WILLIAM AND FLORA HEWLETT FOUNDATION

JOEL BREAKSTONE, MARK SMITH, & SAM WINEBURG | STANFORD UNIVERSITY

AMIE RAPAPORT, JILL CARLE, MARSHALL GARLAND, & ANNA SAAVEDRA | GIBSON CONSULTING

✉ The Stanford History Education Group: sheg@gse.stanford.edu

CONTENTS

Executive Summary

The next presidential election is in our sights. Many high school students will be eligible to vote in 2020. Are these first-time voters prepared to go online and discern fact from fiction?

Overview

In November 2016, the Stanford History Education Group released a study showing that young people lacked basic skills of digital evaluation. Since then, a whole host of efforts—including legislative initiatives in 18 states—have sought to address this problem.

National Survey

From June 2018 to May 2019, we administered an assessment to ***3,446 students***, a national sample that matches the demographic profile of high school students in the United States. The six exercises in our assessment gauged students' ability to evaluate digital sources on the open internet.

The results—if they can be summarized in a word—are troubling:

- ***Fifty-two percent*** of students believed a grainy video claiming to show ballot stuffing in the 2016 Democratic primaries (the video was actually shot in Russia) constituted "strong evidence" of voter fraud in the U.S. Among more than 3,000 responses, only three students tracked down the source of the video, even though a quick search turns up a variety of articles exposing the ruse.
- ***Two-thirds*** of students couldn't tell the difference between news stories and ads (set off by the words "Sponsored Content") on Slate's homepage.
- ***Ninety-six percent*** of students did not consider why ties between a climate change website and the fossil fuel industry might lessen that website's credibility. Instead of investigating who was behind the site, students focused on superficial markers of credibility: the site's aesthetics, its top-level domain, or how it portrayed itself on the About page.

Nearly all students floundered. Ninety percent received no credit on four of six tasks.

Reliable information is to civic health what proper sanitation and potable water are to public health. A polluted information supply imperils our nation's civic health. We need high-quality digital literacy curricula, validated by rigorous research, to guarantee the vitality of American democracy.

Education moves slowly. Technology doesn't. If we don't act with urgency, our students' ability to engage in civic life will be the casualty.

INTRODUCTION

In November 2016, fears of fake news gripped the nation. Following the presidential election, headlines warned of the "Digital Virus Called Fake News" (*New York Times*, 2016) and asked, "Are You Ready for a War on Fake News?" (Bray, 2016). Television stories detailed "Fake News Stories Thriving on Social Media" (Tapper, 2016) and "The Fight Against Fake News" (Todd, 2016).

Research conducted by the Stanford History Education Group (SHEG) contributed to the collective unease. In the midst of the November 2016 media frenzy, SHEG released a report showing the problem extended well beyond fake news (McGrew, Breakstone, Ortega, Smith, & Wineburg, 2018; McGrew, Ortega, Breakstone, & Wineburg, 2017; Stanford History Education Group, 2016). SHEG collected nearly eight thousand responses to tasks in which students examined a range of online sources. From middle school to college, students struggled to perform even the most basic evaluations of digital material. Middle school students confused online ads with news stories. High school students trusted a photograph posted anonymously to a social media site. College students thought a website on employment policy was trustworthy even though it was actually run by a public relations firm. *The Wall Street Journal* (Shellenbarger, 2016) ran a half-page story about our findings. Dozens of stories—from *The New York Times* and *The Washington Post* to interviews with *Die Zeit*, the BBC, and Italy's *la Repubblica*—followed. We fielded questions from policymakers, educators, lawmakers, tech executives, and journalists. All were concerned by students' lack of preparation for dealing with the digital landscape.

Efforts to address "truth decay" (Kavanagh & Rich, 2018) took a variety of forms. Lawmakers across the country drafted legislation (including California Senate Bill No. 830, which cited our 2016 study) to address digital illiteracy (Wineburg, 2018b). New organizations sprung into action to create resources to meet the challenge.

Fast forward three years. The 2020 presidential election is just a year away. Many current high school students will be first-time voters. Are they prepared to find trustworthy sources in an age of digital overabundance?

This question motivated our project. We asked 3,446 high school students, a sample that matches the demographic profile of high school students across the country, to complete tasks that measured their ability to evaluate digital sources. The results—if they can be summarized in a word—are troubling.

BACKGROUND

Reliable information is to civic health what proper sanitation and potable water are to public health. Thriving democracies need citizens who are able to access, evaluate, and use reliable information to participate in public discourse (Hobbs, 2010; Mihailidis & Thevenin, 2013; Knight Commission, 2009). Research has found that young people examine online campaign materials at higher rates than adults (Kahne, Lee, & Feezell, 2012), a finding that presents immense opportunities for democratic participation on one hand and formidable challenges on the other. If youth consume information without the ability to assess its credibility—without the ability to find out who is behind a given cause and what that person's motives might be—they are easy marks for rogues of all stripes. On the internet, traditional gatekeepers and hallmarks of authority are largely absent. It is thus imperative to democratic functioning that students know how to assess

the quality of information on which to base their opinions and decisions (Metzger, 2007; Metzger, Flanagin, & Medders, 2010).

Few tools exist to measure students' competency in the digital realm. Curricula that teach students about the internet rarely include tasks that ask them to evaluate real sources. Instead, students are given multiple-choice questions, true/false items, and, in some cases, hypothetical scenarios. From such proxy measures, it is hard to know what students actually do when facing digital materials in real time.

The paucity of assessment options spurred us to develop a range of new measures. As part of this process, we set out to identify expertise in the evaluation of online sources. We asked three groups to evaluate a series of websites. We observed Stanford freshmen, university professors from four different institutions, and fact checkers from the country's leading news outlets as they navigated unfamiliar sites (Wineburg, 2018a; Wineburg & McGrew, 2019). The fact checkers' approach differed markedly from the undergraduates and professors. When they landed on an unknown website, fact checkers left it quickly and opened new browser tabs to search for information about the trustworthiness of the original source. We refer to this approach as *lateral reading*. In contrast, the students, as well as the academics, typically read *vertically*, spending minutes examining the original site's prose, references, About page, and top-level domain (e.g., .com vs. .org)—features that are all easy to manipulate. Landing on an unfamiliar site, fact checkers learned about it, paradoxically, by leaving it. They read less, learned more, and reached better conclusions in less time.

Fact checkers kept three questions in mind as they evaluated sources: (1) Who's behind the information? (2) What's the evidence? (3) What do other sources say? These questions are at the heart of what we refer to as *civic online reasoning* (COR)—the ability to effectively search for, evaluate, and verify social and political information online (McGrew et al., 2017). We emphasize the civic dimension of our work because digital literacy has become a prerequisite for civic engagement (Kahne, Lee, & Feezell, 2012; Knight Commission on the Information Needs of Communities in a Democracy, 2009; Mihailidis & Thevenin, 2013). Moreover, we sought to distinguish our efforts from the broader field of media literacy, which covers topics ranging from the creation of web videos to the dangers of sexting. The scope of COR is more focused: it trains attention on the skills, knowledge, and competencies that students need in order to make sound decisions about social and political issues.

The purpose of this study was to explore whether the intense concern about information literacy since 2016 has had an effect on students' digital abilities. Are young people today, over three years after our original study, prepared to make choices based on the digital information they consume?

RESEARCH QUESTIONS

This research addressed two related questions:

Question 1

How do students across the United States perform on assessments of civic online reasoning?

Question 2

Are there differences in how students perform based on region, locale (urban, suburban, rural), and demographic characteristics?

RESEARCH TEAM

This project was a collaboration between two organizations. Members of the Stanford History Education Group (SHEG), a research and development effort based at Stanford University's Graduate School of Education, worked with staff from Gibson Consulting, an education research group in Austin, Texas. Although it was a collaboration, each group took primary responsibility for different aspects of the project. SHEG was responsible for project conceptualization, assessment development, rubric creation, and reporting. Gibson led recruitment, scoring, and data analysis.

METHODS

Recruitment

We first gathered public data on all districts in the United States. Using the most recent Common Core of Data from the National Center for Education Statistics, we downloaded all district and school data for the 2015-16 school year. We separated all districts into 12 unique strata: four geographic regions (Northeast, Midwest, South, West) and three locale types (urban, suburban, and rural/town). We then recruited districts from each stratum, with the goal of acquiring student responses from each. All traditional, regular instruction, public and

Table 1:
Participating Districts by Region, Locale, State, and Enrollment

Region	Locality	State	Enrollment
Northeast	Urban	PA	22,359
Northeast	Suburban	NJ	15,878
Northeast	Rural	ME	2,089
Northeast	Rural	VT	2,034
Midwest	Urban	MI	45,455
Midwest	Suburban	OH	20,322
Midwest	Suburban	MN	21,029
Midwest	Rural	MN	5,122
Midwest	Rural	WI	5,100
South	Urban	NC	12,115
South	Suburban	OK	15,942
South	Rural	TX	427
South	Rural	TX	2,486
West	Urban	CO	29,682
West	Suburban	CA	4,074
West	Rural	OR	6,579

Source: National Center for Education Statistics Common Core of Data (2016-17).

Table 2
Student Race/Ethnicity in Participating Districts

Region	Locality	American Indian	Asian	Black/African American	Hispanic	White
Northeast	Urban	0.2%	3.6%	52.9%	3.1%	32.7%
Northeast	Suburban	0.1%	4.0%	5.2%	15.1%	73.1%
Northeast	Rural	0.3%	5.3%	1.6%	1.8%	91.0%
Northeast	Rural	0.1%	1.3%	2.5%	3.7%	88.9%
Midwest	Urban	0.2%	1.7%	80.7%	14.6%	2.5%
Midwest	Suburban	0.2%	11.7%	4.0%	3.1%	76.8%
Midwest	Suburban	0.3%	16.2%	23.7%	8.7%	44.4%
Midwest	Rural	0.6%	1.6%	1.3%	3.0%	89.3%
Midwest	Rural	1.2%	4.6%	1.8%	4.5%	85.6%
South	Urban	0.3%	15.0%	11.2%	15.7%	51.5%
South	Suburban	4.6%	3.1%	6.1%	13.9%	58.5%
South	Rural	0.2%	0.9%	6.6%	15.2%	71.9%
South	Rural	0.6%	0.6%	1.0%	30.0%	66.3%
West	Urban	0.5%	3.2%	1.1%	17.8%	73.8%
West	Suburban	1.2%	1.3%	0.6%	13.7%	76.1%
West	Rural	6.5%	1.2%	0.9%	19.3%	65.6%

Source: National Center for Education Statistics Common Core of Data (2016-17).

charter schools with students enrolled in grade 12 were included in the sampling scheme.

Within each stratum, we targeted districts with relatively higher levels of student diversity based on a measure of racial/ethnic heterogeneity.[1] By starting with more heterogeneous districts, we increased our chances of recruiting a diverse group of students. Racial and ethnic heterogeneity was not the only consideration. We also prioritized districts where we had existing relationships, and we sought to include districts from multiple states. Initial outreach included a recruitment letter, a FAQ document, a summary of the project, and a link to a study-specific recruitment webpage on the SHEG website. If a district expressed interest, we worked with district staff to answer questions and ensure compliance with district research protocols. We followed up with districts until we secured at least one district in each stratum. In some cases, multiple districts participated within a given stratum.

Participating Districts

We recruited 16 districts in 14 states. Participating districts varied considerably in

[1] We calculated the multi-group racial/ethnic diversity for each school and district with Simpson's Diversity Index formula. We used Reardon and Townsend's (2018) Stata -seg- command. The measure reflects the probability that two students selected from a random draw in a given entity (e.g., a school or district) are members of the same racial/ethnic group.

Table 3
Percent of Title I Schools, Students with Limited English Proficiency, and Students Receiving Special Education Programs

Region	Locality	Title I Schools	Limited English Proficiency	Special Education
Northeast	Urban	89.7%	4.0%	21.9%
Northeast	Suburban	0.0%	1.3%	17.2%
Northeast	Rural	0.0%	1.0%	14.4%
Northeast	Rural	83.3%	0.7%	17.7%
Midwest	Urban	92.3%	12.3%	17.0%
Midwest	Suburban	0.0%	2.2%	11.4%
Midwest	Suburban	100.0%	10.5%	14.2%
Midwest	Rural	100.0%	1.2%	15.5%
Midwest	Rural	38.5%	2.1%	14.0%
South	Urban	10.0%	10.4%	10.2%
South	Suburban	60.0%	4.3%	17.2%
South	Rural	100.0%	0.7%	10.5%
South	Rural	100.0%	11.1%	9.5%
West	Urban	9.6%	5.1%	*
West	Suburban	0.0%	1.7%	9.5%
West	Rural	68.2%*	4.6%	14.5%

Source: National Center for Education Statistics Common Core of Data (2016-17).

**The total count of students receiving special education services was missing for the urban district in the West in both 2016-17 and 2015-16 NCES data. Title I status was missing at the school level for 2016-17 for the rural district in the West, so we used data from 2015-16 instead.*

size, serving anywhere from 427 students to over 45,000. Table 1 includes a list of all districts with their 2016-17 enrollment figures, organized by their region and locale.

Districts represented a diverse selection of the country's public schools. Districts ranged in their levels of racial/ethnic diversity, with the percent of students enrolled who were White ranging from 3% to 91%. The percent of students enrolled who were Hispanic ranged in participating districts from 2% to 30%, and the percent of students enrolled who were Black/African American ranged from less than one percent to over 80%. Table 2 displays the student racial/ethnic composition of the participating districts.

Other characteristics of students enrolled in participating districts varied (see Table 3). In some districts, no schools were eligible to receive federal support under Title I, while in others all schools in the district were Title I eligible (based on the concentration of students who received free or reduced-price lunch through the National School Lunch Program). In some districts, English language learners represented 1% of students; in others 12% of students had limited English proficiency. The percentage of students receiving

special education services under the Individuals with Disabilities Education Act (IDEA) also varied, ranging from less than 10% to over 20%.

Teacher Recruitment and Registration

Individual teachers were selected to participate through a multi-step process. First, we invited teachers in at most three schools per district to participate. If a district had three or fewer such schools, we included all of these schools; if a district had four or more such schools, we randomly selected three to participate.[2] All 9th to 12th grade social studies teachers in selected schools received study information and registration instructions.

Interested teachers registered for the study online, providing their name, contact information, and details about each course section they taught. We randomly selected two sections from each teacher's schedule to ensure that teachers did not select only the highest performing (or otherwise non-randomly chosen) class. As a token of appreciation, we provided teachers with a $50 gift card for use with students in those sections. Teachers could administer the tasks to all of their sections, but we provided the gift cards only for the two sections we selected. Once registered, we sent instructions and worked with district staff to ensure students would have open access to the internet while completing the tasks.

A total of 122 teachers from 43 schools in the 12 participating districts registered. These 122 teachers taught between 1 and 8 course sections, with a median of 5 course sections. Across these course sections, the median number of students enrolled was 25, though course section enrollment ranged from 1 to 60 students.

Assessment Instrument

The assessment included six constructed-response tasks developed by SHEG (cf. McGrew et al., 2018). Taken together, the tasks measure the kinds of skills students need to be civically engaged in a digital age. For example, one task asked students to weigh the strengths and weaknesses of a tweet from a political advocacy group. Another asked them to evaluate whether a video posted on social media represents strong evidence of voter fraud in the 2016 Democratic presidential primaries. Table 4 outlines the tasks included in the survey and provides a brief description of each. (See Appendix A for copies of the assessments.)

After students completed the tasks, we asked them to provide information about a range of demographic characteristics, including: age, grade, gender, race, ethnicity, primary language spoken at home, grades in school, mother's education, internet habits, and free/reduced lunch participation. (See Appendix A for the complete survey.) Students were not required to answer the demographic questions in order to submit their responses to the tasks.

Administration

The assessment was administered using Qualtrics, an online survey platform. Participation was voluntary, and students were allowed to opt out. Students who agreed to participate used an internet-connected computer during their regularly scheduled social studies class. Students were allowed to use the internet at any time during the test administration, and they were reminded that they could search online to complete the tasks. Qualtrics presented each

[2]There were two exceptions. A district from Pennsylvania has a large proportion of magnet schools. To ensure we selected schools that represented the district, we first stratified by magnet/traditional school type and then randomly selected two schools from each type. In an Ohio district, some teachers taught in multiple schools. Some teachers from the three schools we selected also administered the assessment with students in a fourth school that we had not selected.

Table 4
Description of Tasks on COR Assessment

Item	COR Question(s) Addressed	Description
Evaluating Video Evidence	What's the evidence? Who's behind the information?	Evaluate whether a video posted on Facebook is good evidence of voter fraud.
Webpage Comparison	Who's behind the information?	Explain which of two websites is a better source of information on gun control.
Article Evaluation	What do other sources say? Who's behind the information?	Using any online sources, explain whether a website is a reliable source of information about global warming.
Claims on Social Media 1	What's the evidence? Who's behind the information?	Explain why a social media post is a useful source of information about background checks for firearms.
Claims on Social Media 2	What's the evidence? Who's behind the information?	Explain how a social media post about background checks might not be a useful source of information.
Homepage Analysis	Who's behind the information?	Explain whether tiles on the homepage of a website are advertisements or news stories.

task in the same order for all students. Students were not allowed to return to items after they had submitted their responses. No time limits were imposed for completing the tasks.

Participating teachers were provided with detailed instructions for administering the assessment. Teachers were allowed to provide technical assistance to students as they completed the tasks, but they could not help students evaluate sources, interpret questions, or compose their responses.

Analytic Sample

In order to create a sample that matched the demographics of American high school students, we first collected assessments from 6,445 students from 79 classrooms in 31 schools. Table 5 shows teacher and student participation by region and locale. Teachers were allowed to administer the assessment to as many students as they wished. As a result, students in some regions and locales were overrepresented in the sample. We used the 2016-17 Common Core of Data from the National Center for Educational Statistics to calculate the percentages of public school students in each geographical region of the United States (Midwest, Northeast, South, and West) and by locale within each region (rural, suburban, and urban). Next, we generated a random sample of 2,937 students that was proportionate to the distribution of students in each region of the United States. We refer to this group as the *analytic sample*.

We then compared the racial and ethnic composition of the analytic sample to the population of public school students in the United States. We found that the proportions of Hispanic and Black/African American students were lower than the population of public school students in the nation at large. To increase the number of students in these categories, we

Table 5
Teacher and Student Participation by Region and Locale

Region	Locality	Teacher *N*	Student *N*
Northeast	Urban	7	292
Northeast	Suburban	12	1,161
Northeast	Rural	5	282
Midwest	Urban	2	173
Midwest	Suburban	22	2,104
Midwest	Suburban	3	268
South	Urban	4	309
South	Suburban	5	459
South	Rural	5	332
West	Urban	3	316
West	Suburban	6	411
West	Rural	5	338
	Total	**79**	**6,445**

returned to the set of completed assessments and added the remaining 509 students who described themselves as Hispanic or Black/African American to the analytic sample. Adding these students resulted in a final analytic dataset of 3,446 students that better matched the racial/ethnic makeup of American public schools. Table 6 details the racial and ethnic composition of our analytic sample and compares it with the population of public school students in the United States.

Increasing the number of African American and Hispanic students changed the composition of the sample by region/locality, but the impact was small. Table 7 shows the geographic distribution of the final analytic sample of 3,446 students. Students from suburban schools in the Northeast and Midwest were slightly overrepresented and students from the South were slightly underrepresented, but the overall analytic sample remained reflective of the geographical distribution of students in the United States.

In order to ensure that our sample was reflective of the broader population of students in the country, we also examined self-reported grade level, gender, and free/reduced lunch eligibility for students in the analytic sample. Table 8 lists the number and proportion of students in each classification.

Rubric Design and Scoring Procedures

Two independent raters from Gibson Consulting scored student responses using a three-level rubric (Beginning - 0; Emerging - 1; Mastery - 2).[3] Estimates of interrater reliability were high, with Kappa coefficients between .93 and .97 for each task.

Rubric design for this project was informed by our prior research and development work on civic online reasoning assessment (cf. McGrew et al., 2018). In Mastery responses, students evaluated online content effectively by investigating

[3]Student responses were excluded if the students indicated they could not view the website featured in the assessment or if answers were nonsensical (e.g., jumbles of letters, numbers, or words).

Table 6

Analytic Sample Compared to U.S. Enrollment by Race and Ethnicity

Race/Ethnicity	Proportion of US Enrollment	Analytic Sample *N*	Analytic Sample	Analytic Sample without Missing
Asian or Pacific Islander	5%	112	3%	4%
American Indian/Alaska Native	1%	29	1%	1%
Black or African American	15%	306	9%	11%
Hispanic	26%	698	20%	25%
Other	-	42	1%	2%
Two or more races	3%	224	7%	8%
White	49%	1,373	40%	49%
Prefer not to say	-	28	1%	-
Missing	-	634	18%	-
Total	**100%**	**3,446**	**100%**	**100%**

Table 7

Final Analytic Sample Compared to U.S. Enrollment by Region and Locale

	Proportion of US Enrollment	Students in Analytic Sample	Proportion of Analytic Sample
Region			
Northeast	15.8%	732	21.2%
Midwest	21.4%	852	24.8%
South	37.6%	1,084	31.4%
West	25.1%	778	22.6%
Total	**100.0%**	**3,446**	**100.0%**
Locale			
Urban	30.7%	967	28.1%
Suburban	42.4%	1,659	48.1%
Rural	26.8%	820	23.8%
Total	**100.0%**	**3,446**	**100.0%**

Table 8

Analytic Sample by Gender, Grade Level, Free/Reduced Lunch Status, and Mother's Education

	% US Enrollment	Analytic Sample *N*	Analytic Sample	Analytic Sample without Missing
Gender				
Male	52%	1,283	37%	43%
Female	48%	1,443	42%	49%
Non-binary/self-describe/prefer not to say	-	230	7%	8%
Missing	-	490	14%	-
Total	**100%**	**3,446**	**100%**	**100%**
Grade Level				
8th	-	17	0%	1%
9th	27%	820	24%	28%
10th	26%	491	14%	17%
11th	24%	867	25%	29%
12th	24%	761	22%	26%
Missing	-	490	14%	-
Total	**100%**	**3,446**	**100%**	**100%**
National School Lunch Program Status				
Eligible/Participant	52%	850	25%	29%
Not eligible/Non-participant	48%	1,475	43%	50%
I don't know	-	533	15%	18%
Prefer not to say	-	84	2%	3%
Missing	-	504	15%	-
Total	**100%**	**3,446**	**100%**	**100%**
Mother's Education level[a]				
Less than high school	-	222	6%	8%
High school	-	506	15%	17%
Some college	-	460	13%	16%
Bachelor's degree	-	736	21%	25%
Advanced degree	-	473	14%	16%
I don't know	-	411	12%	14%
Prefer not to say	-	148	4%	5%
Missing	-	490	14%	-
Total	**-**	**3,446**	**100%**	**100%**

Source: National Center for Education Statistics Common Core of Data (2016-17).

[a] Common Core of Data does not include data on mother's education for U.S. public school students.

Table 9
Percent of Responses at Each Score Level by Task

Scored Response	Evaluating Evidence	Webpage Comparison	Website Evaluation	Social Media 1	Social Media 2	Homepage Analysis
Beginning	77.4%	92.1%	96.8%	92.8%	95.4%	68.7%
Emerging	13.9%	6.7%	1.4%	6.9%	2.7%	26.6%
Mastery	8.7%	1.2%	1.8%	0.3%	1.9%	4.7%
Total (*N*)	**3,119**	**2,678**	**2,994**	**2,431**	**2,431**	**2,969**

the source of information, by interrogating the evidence presented, or by seeking out information from other reliable sources. Emerging responses were on the right track but were partially incorrect or did not fully articulate sound reasoning. Beginning responses reflected incorrect or irrelevant strategies for evaluating online information. (See Appendix B for a sample rubric for Task 1.)

RESULTS

Research Question 1: Overall Student Performance

Overall, students struggled on all of the tasks. At least two-thirds of student responses were Beginning for each of the six tasks. On four of the six tasks, over 90% of students received no credit at all. Out of all of the student responses to the six tasks, fewer than 3% earned full credit. Claims on Social Media question 1 had the lowest proportion of Mastery responses, with less than 1% of students demonstrating a strong understanding of the COR competencies measured by the task. Evaluating Evidence had the highest proportion, with 8.7% earning full credit. (See Table 9 for the distribution of scores by task.)

Sample Task: Website Evaluation

The Website Evaluation task had the highest proportion of Beginning scores, with 96.8% of students earning no points. The question assessed whether students could engage in lateral reading—that is, leaving a site to investigate whether it is a trustworthy source of information. Students were provided a link to the homepage of CO_2 Science (co2science.org), an organization whose About page states that their mission is to "disseminate factual reports and sound commentary" on the effects of carbon dioxide on the environment. (See Figure 1.)

Students were then asked whether this page is a reliable source of information. Screen prompts reminded students that they were allowed to search online to answer the question. The few students who earned a Mastery score used the open internet to discover that CO_2 Science is run by the Center for the Study of Carbon Dioxide and Global Change, a climate change denial organization funded by fossil fuel companies, including ExxonMobil. The Center for the Study of Carbon Dioxide and Global Change also has strong ties to the American Legislative Exchange Council, an organization that opposes legislative efforts to limit fossil fuel use.

A student from a rural district in Oregon wrote:

> *I do not believe this is a reliable source of information about global warming because even though the company is a non-profit organization, it receives much of its funding from the "largest U.S. oil company–*

Figure 1
Image of the CO_2 Science Homepage

> *ExxonMobil–and the largest U.S. coal mining company–Peabody Energy" (Greenpeace). Moreover, Craig Idso, the founding chairman of The Center for the Study of Carbon Dioxide and Global Change, was also a consultant for Peabody Energy. It is no wonder this organization advocates for unrestricted carbon dioxide levels; these claims are in the best interest of The Center for the Study of Carbon Dioxide and Global Change as well as the oil and mining companies that sponsor it.*

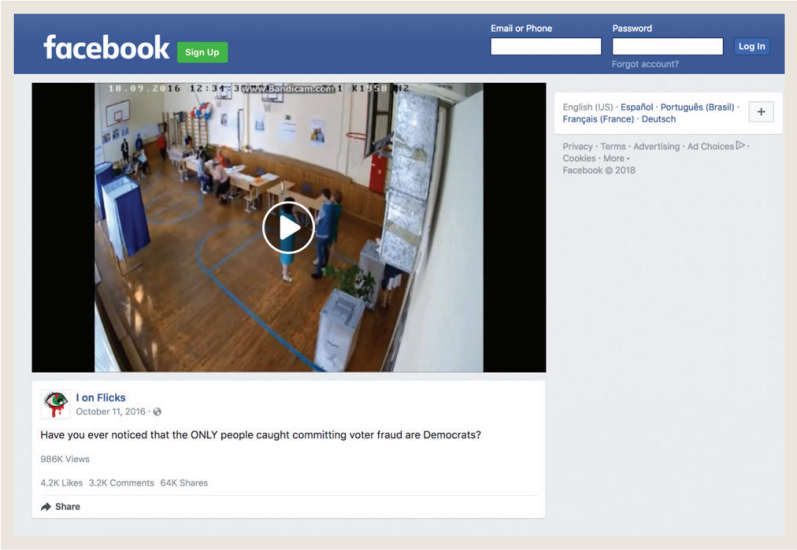
Another student from suburban Oklahoma responded:

> *No, it is not a reliable source because it has ties to large companies that want to purposefully mislead people when it comes to climate change. According to* USA TODAY, *Exxon has sponsored this nonprofit to pump out misleading information on climate change. According to the* Seattle Post-Intelligencer, *many of their scientists also have ties with energy lobbyists.*

Both students adeptly searched for information on the internet about who was behind the site. Both concluded that the site was unreliable because of its ties to fossil fuel interests.

Responses like these were exceedingly rare. Less than two percent of students received a Mastery score. Over 96% of student responses were categorized as Beginning. Instead of leaving the site, these students were drawn to features of the site itself, such as its top-level domain (.org), the recency of its updates, the presence or absence of ads, and the quantity of information it included (e.g., graphs and infographics). A student from suburban New Jersey wrote: "This page is a reliable source to obtain information from. You see in the URL that it ends in .org as opposed to .com." This student was seemingly unaware that .org is an "open" domain; any individual or group can register a .org domain without attesting to their motives. A student from the urban South was taken in by CO_2 Science's About page: "The 'about us' tab does show that the organization is a non-profit dedicated to simply providing

Figure 2
Image of Facebook Post for Evaluating Video Evidence Task



the research and stating what it may represent. In their position paper on CO_2, they provide evidence for both sides and state matters in a scientific manner. Therefore, I would say they are an unbiased and reliable source."

Accepting at face value how an unknown group describes itself is a dangerous way to make judgments online.[4]

Sample Task: Evaluating Evidence

The Evaluating Evidence task gauged students' ability to evaluate the trustworthiness of a social media post. The post presented students with a video on Facebook from a user named "I on Flicks" and asked students if this video was "strong evidence" of voter fraud during the 2016 Democratic primaries (see Figure 2). The video includes four clips of poll workers surreptitiously stuffing ballots into bins. The video is silent, but the captions tell viewers that the clips depict Democratic 2016 primary elections in Illinois, Pennsylvania, and Arizona. The post accompanying the video reads: "Have you ever noticed that the ONLY people caught committing voter fraud are Democrats?" None of this information is true. The clips actually show voter fraud in Russia, not the United States.

Over half of the students in our sample (52%) were fooled by the video. These students concluded that the video provided "strong evidence" of voter fraud in Democratic primary elections. A student from a rural district in Ohio took the video at face value: "Yes, it shows video evidence of fraud in multiple different states at

[4]Students often displayed a confusion about the meaning of top-level domains such as .org. While there are many .org's that work for social betterment, the domain is a favorite for political lobby groups and "astroturf" sites (groups that cast themselves as grassroots efforts but which are actually backed by powerful political or commercial interests). For-profit companies can also be listed as .org's. Craigslist, a corporation with an estimated $1 billion in revenue in 2018, is registered as craigslist.org. Nor is nonprofit status a dependable marker of an organization's credibility. Of the 101,962 applications the IRS received in 2015, 95,372 were granted tax-deductible status—an approval rate of 94%. An analysis by the Stanford University Center on Philanthropy and Civil Society examined the kinds of organizations that receive approval in a single year and reached this conclusion: "Obtaining recognition by the IRS as a public charity is an embarrassingly easy thing to do. It is hardly an exaggeration to say that when it comes to oversight of the application process to become a public charity, nearly everything goes" (Reich, Dorn, & Sutton, 2009, p. 4).

multiple different times." Another student from an urban district in Pennsylvania wrote, "Yes, because the video showed people entering fake votes into boxes. I'm assuming that they were Hillary votes."

A quarter of the students rejected the video but could not provide a relevant reason. Students focused on irrelevant features such as the lack of audio, the video's grainy quality, or insufficient narration from "I on Flicks" about what was happening. A student from suburban California reasoned, "The video provided no audio and gave little description of how the footage proves that there was voter fraud." Others accepted the veracity of the clips but rejected the video because there were too few examples. One student from suburban Oklahoma wrote, "The video only shows a few specific instances, and is not enough to blame the whole Democratic party of voter fraud." Still others struggled to interpret what was depicted in the clips, ignoring the source and origin of the video altogether. A student from Oklahoma explained, "I have no idea what is going on exactly and I see people put papers in the box, but there are cases where the votes are sneaked in. I am not sure this is a strong case of faking votes for a person, but it is a case for vote fraud."

Only 8.7% of students rejected the video and provided a relevant explanation. These students wrote that they had no way to know if "I on Flicks" was a reliable source or whether the video actually depicted Democrats in the United States. A student from an urban district in the West questioned the video's origin: "It does not provide enough evidence because we don't know who the people are in the video, and who the people are that made the video. There is very little context and information given, besides the place, and a little bit of commentary making claims to convince you." Another student from the suburban Midwest correctly identified a problem with the evidence: "Although the post depicts (what we assume are) votes being shredded (something that hits home in the US, where representation is sacred), the video never proved that all of these votes were in support of a particular side, as the Facebook post suggested." Of the more than 3,000 responses, only three students actually tracked down the source of the video. Unprompted, a student from suburban California skillfully engaged in lateral reading: "Doing some research on the internet, I found the exact same video as shown above. However, the video which I found, which was presented by BBC news, showed that the cam footage had nothing to do with the 2016 Democratic primary elections. Instead, BBC presented the cam footage as that of the Russian 2016 election. According to the announcer of BBC news, the video showed that the people in Russia—not America—were committing voter fraud."

Research Question #2: Differences Across Demographic Characteristics

Descriptive Portrait. After completing the six tasks, students responded to a series of questions about their race, ethnicity, gender, socioeconomic status, location, and academic characteristics. Below, we disaggregate student scores by these demographic variables.

Before conducting statistical tests, we wanted to understand patterns that emerged from the raw data. We examined the proportions of students who scored *either* Mastery or Emerging for each task. We did this because the vast majority of students earned no credit and few students earned Mastery scores, resulting in little variance. Combining Mastery and Emerging scores allowed

us to see patterns that we might have otherwise missed had we compared Mastery scores alone.

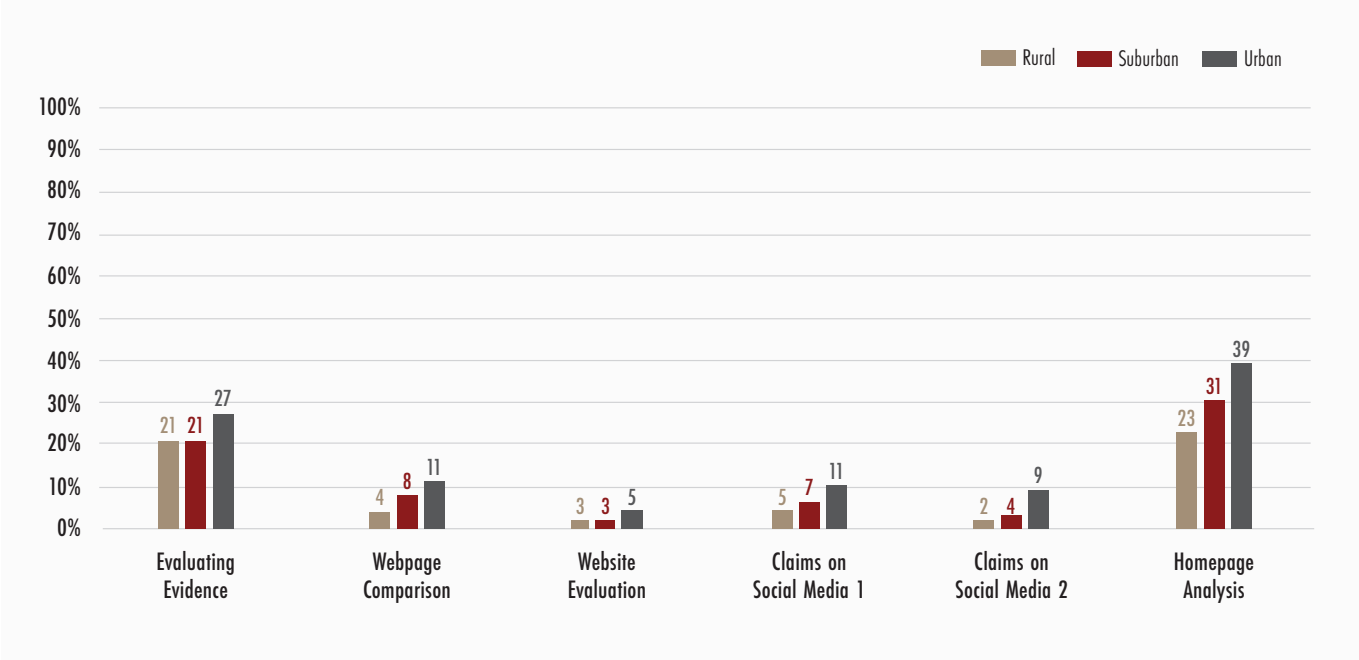
Performance by Geographic Region and Locale. The data revealed differences in the proportion of students earning Emerging/Mastery scores by locale (urban, suburban, or rural). Students in urban districts outperformed their peers in suburban and rural districts on all six tasks; students in suburban districts scored higher than rural students on four of six tasks. Figure 3 provides an overview of the percentages of students scoring Mastery or Emerging by locale for each task.[5]

There were also differences in how students from different regions of the country (Midwest, Northeast, South, West) performed. On all six tasks, students in participating districts from the South and West had higher proportions of Emerging/Mastery responses than students from districts in the Midwest and Northeast. (See Figure 4.)

Performance by Demographic Characteristics. We also observed differences in students' scores by race and ethnicity. Figure 5 shows the proportion of students scoring Emerging/Mastery by race and ethnicity for each task. A fairly consistent pattern of scores by race and ethnicity emerged across all six tasks. Students who identified as Asian/Pacific Islander had the highest proportion of responses scored as Mastery or Emerging on all six tasks; students who identified as Black/African American had the lowest proportion of responses in these scoring categories. The proportion of Mastery and Emerging scores for Hispanic, multiracial, and White students fell between those of Asian/Pacific Islander and Black/African American students for all six tasks.

The data also revealed differences in scores based on whether students reported receiving free or reduced lunch at school, a rough indicator

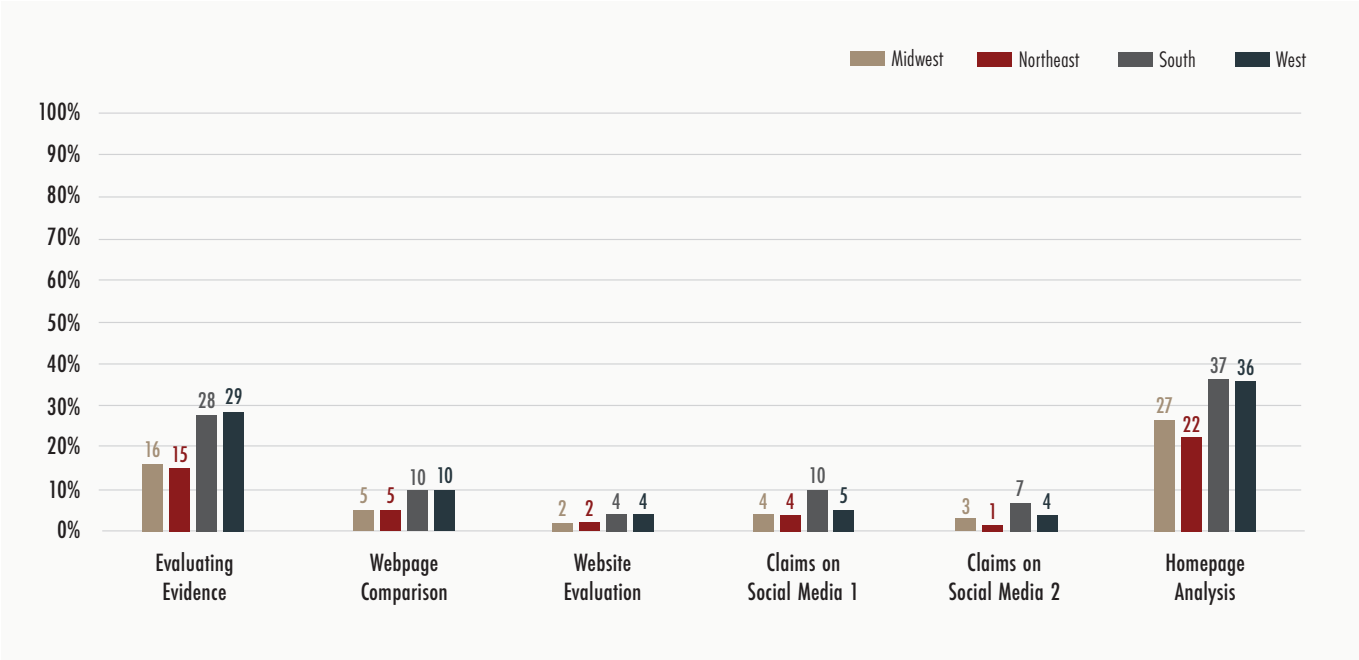
Figure 3
Percent Scoring Emerging or Mastery by Locale and Task



Note: Figure includes scores from 3,446 total students, but numbers vary by task. See Appendix C for student counts by task.

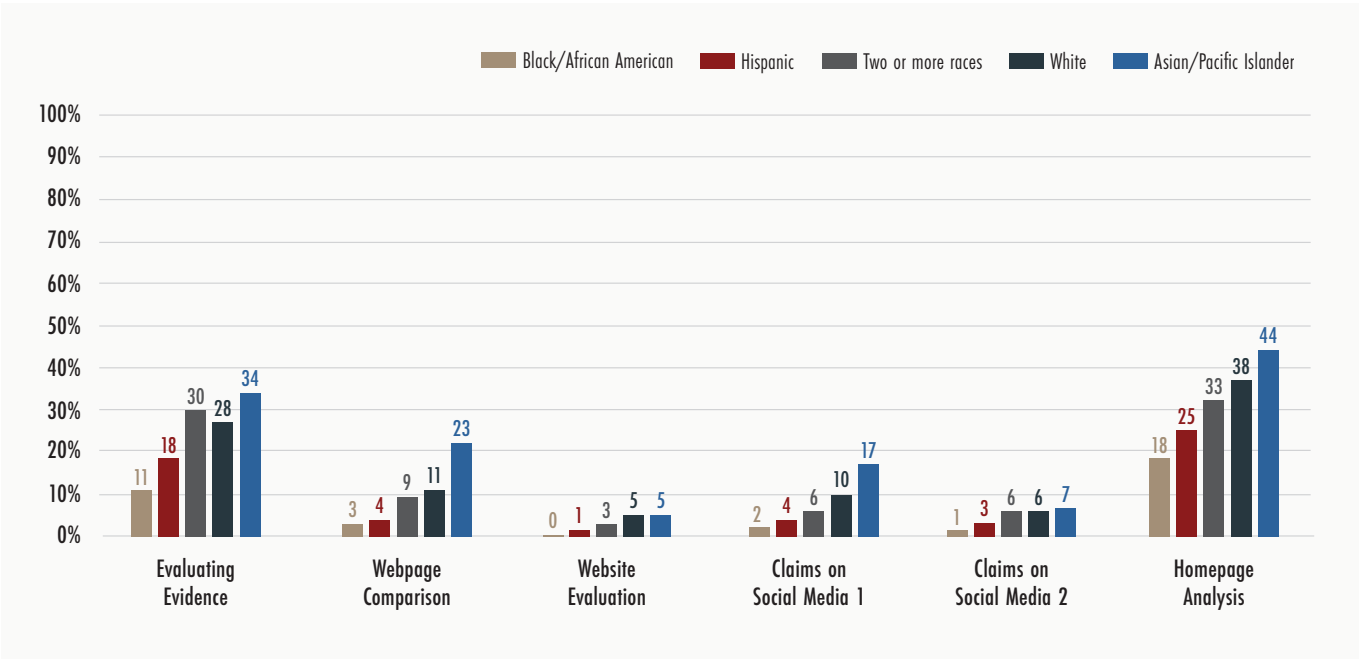
[5]Figures only include student subgroups with 50 or more respondents. Appendix C includes results tables with all response categories, regardless of group size.

Figure 4
Percent Scoring Emerging or Mastery by Region and Task



Note: Figure includes scores from 3,446 total students, but numbers vary by task. See Appendix C for student counts by task.

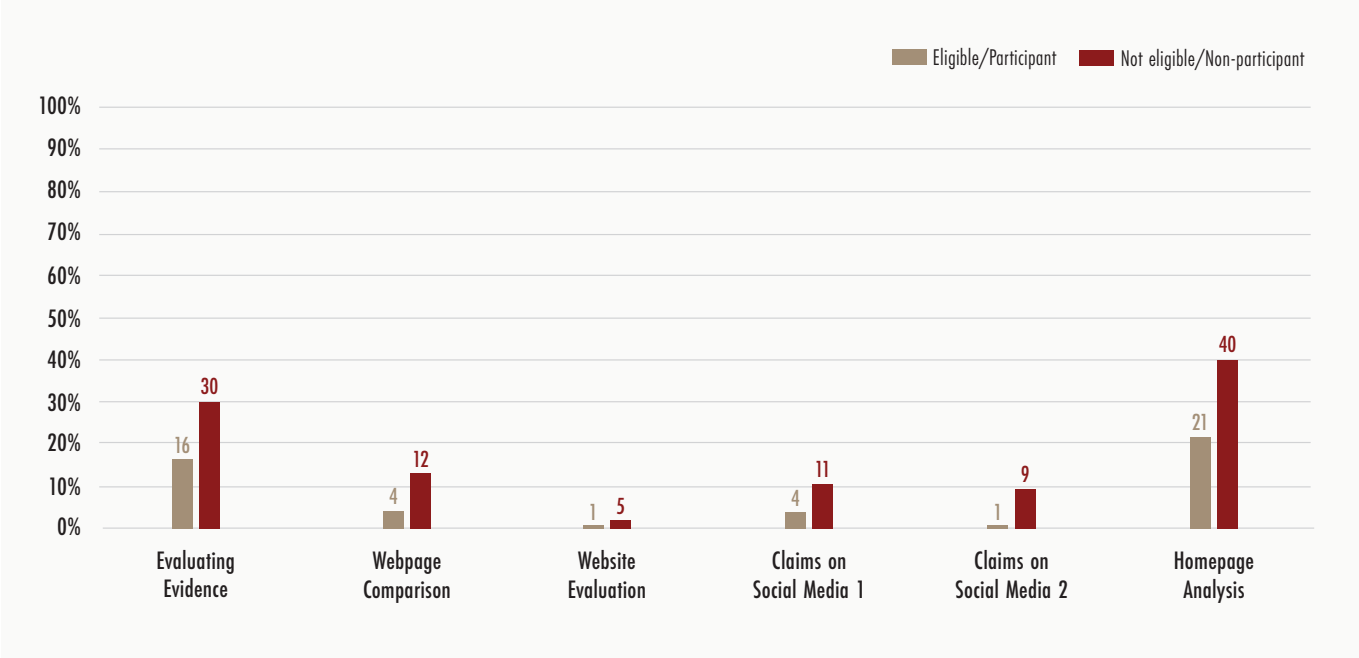
Figure 5
Percent Scoring Emerging or Mastery by Race/Ethnicity and Task



Note: This figure includes 2,490 students who selected a race/ethnicity category. However, the number of responding students varied by task. Excluded categories included "Other," "Prefer Not to Say," and "Missing." See Appendix C for student counts by task for all available response options for this question.

Figure 6

Percent Scoring Emerging or Mastery by Free/Reduced Lunch Status and Task



Note: This figure includes 2,325 students who selected a Free/Reduced Price Lunch category. Students who selected "I Don't Know" or "Prefer Not to Say" are not included. The number of students varied by task. See Appendix C for student counts for this task by response option.

of family income. Figure 6 shows the disparity in scores between students who reported receiving free and reduced lunch and those who did not. Across all six tasks, the proportion of Emerging/Mastery responses were lower for students who reported receiving free/reduced lunch.

Student performance also differed by parental education. Students who reported that their mother had higher levels of education were generally more successful (see Figure 7). Students whose mother had an advanced degree had the highest proportion of Mastery and Emerging scores on all six tasks; students whose mother had at least a bachelor's degree had higher scores than students whose mother had a high school degree or less on each task.

No clear patterns emerged for scores of male and female students, but students who selected "non-binary/prefer to self-describe" scored consistently lower than their male and female peers. Figure 8 shows the proportion of students whose responses were categorized as emerging/mastery by gender.

We did, however, see differences in student scores based on the primary language spoken at home. See Figure 9 for a comparison of student performance by language spoken at home. Students who reported that English was the primary language spoken at home scored higher on all six tasks than students who reported any other language as the most commonly spoken.

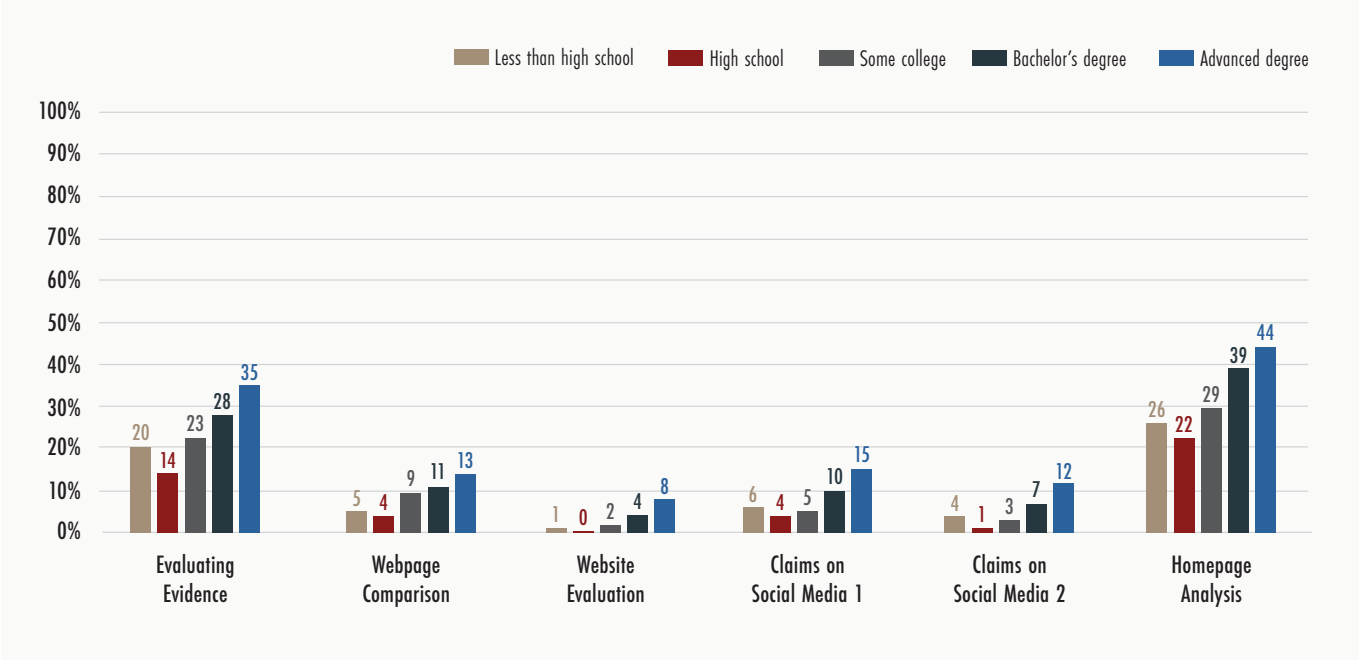
Performance by Academic Characteristics

The data showed differences in student scores based on grade level (9th-12th), with students in higher grades scoring higher than their peers in lower grades. Figure 10 shows the proportion of students scoring at Mastery or Emerging for grades 9-12. Students outperformed their peers at lower grade levels on all six tasks. In no instance did a lower grade outperform an upper grade.

Students' self-reported grades were also associated with their performance. Figure 11 shows the proportion of students earning Emerging/Mastery scores by self-reported grades. Students who reported higher grades typically did

Figure 7

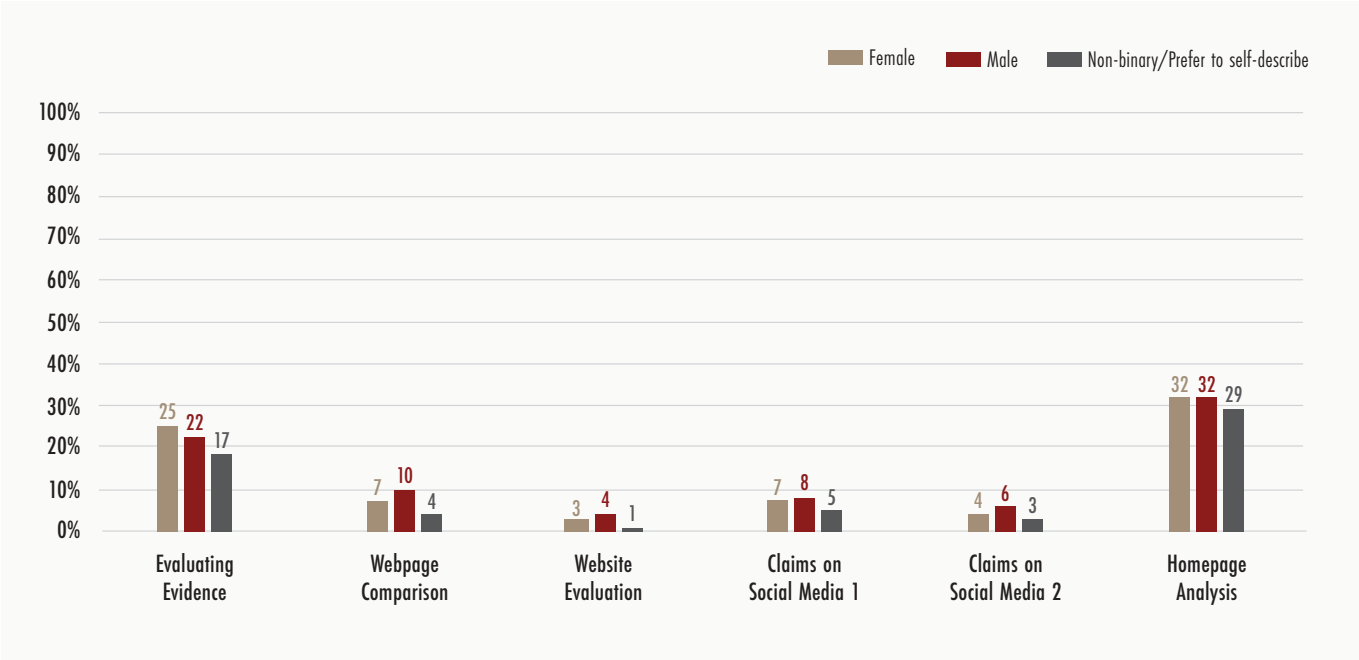
Percent Scoring Emerging or Mastery by Mother's Highest Level of Education and Task



Note: This figure includes 2,397 students who selected an education category though the number of responding students varied by task. Excluded categories include "I Don't Know," "Prefer Not to Say," and "Missing." See Appendix C for student counts by task for all available response options for this question.

Figure 8

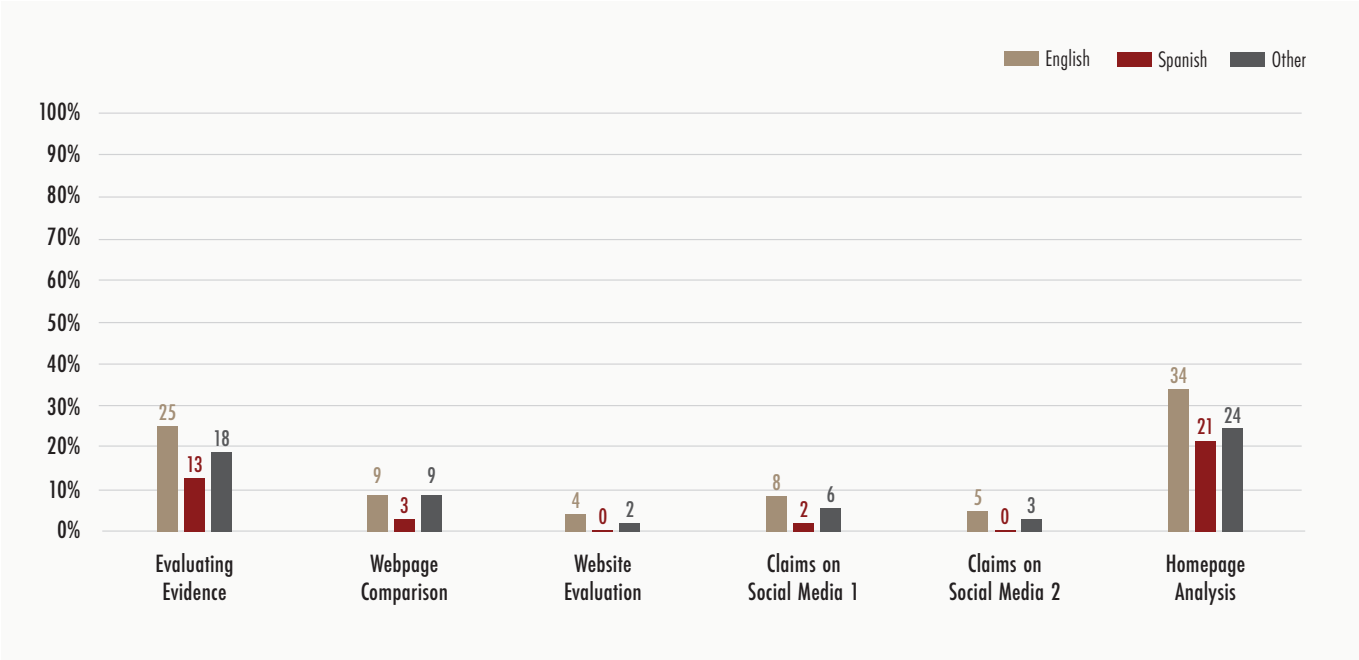
Percent Scoring Emerging or Mastery by Gender and Task



Note: This figure includes 2,898 students who selected a gender category though the number of responding students varied by task. Excluded categories include "Prefer Not to Say" and "Missing." See Appendix C for student counts by task for all available response options for this question.

Figure 9

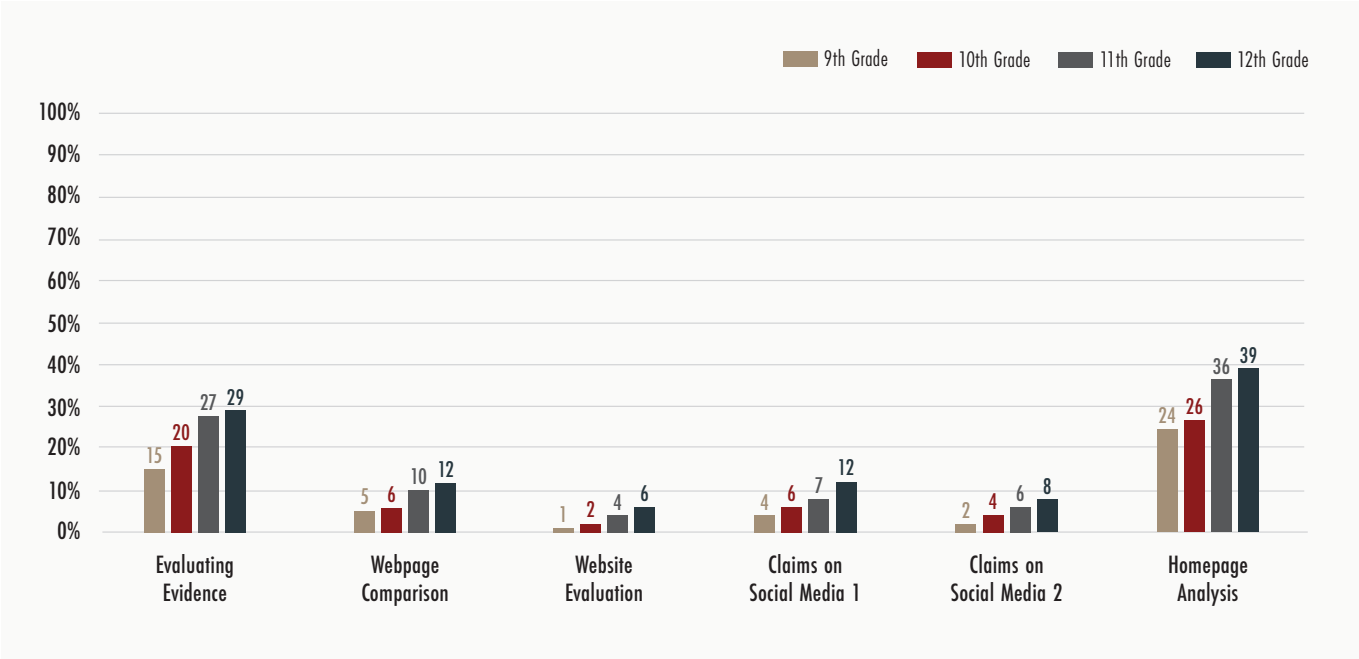
Percent Scoring Emerging or Mastery by Home Language and Task



Note: This figure includes 2,946 students who selected a language category though the number of responding students varied by task. Responses from students who did not answer the Home Language question were excluded. See Appendix C for student counts by task for all available response options for this question.

Figure 10

Percent Scoring Emerging or Mastery by Grade and Task



Note: This figure includes 2,939 students who selected a grade category though the number of responding students varied by task. Excluded categories include "8th grade" and "Missing." See Appendix C for student counts by task for all available response options for this question.

better on the tasks than students who reported lower grades. Students who reported earning "mostly A's" outperformed their peers who reported receiving lower grades on all six tasks.

Multivariate Regression Analysis. The descriptive portrait reveals patterns in the data by student characteristics. At the same time, the descriptive statistics above did not control for the interrelated nature of many student characteristics. We used multivariate regression to disentangle these relationships and to explore whether student characteristics were related to student performance on the tasks. To estimate student performance across all six tasks, we first created an overall composite score. We then examined the relationship between student characteristics and composite score using multivariate regression. Below, we outline how we constructed the composite score and specify the regression model used for the analysis.

Composite Score. The composite score was an average of students' scores on the items they completed.[6] So, if a student completed all six items, then their composite score was the mean of their six scores; if they completed three items, their score was the mean of those three scores, and so on.[7] The average composite score ranged between 0 (a student received a score of 0 on all completed items) to 2 (a student received a score of 2 on all completed items). Table 10 shows the distribution composite scores. The mean was 0.17. Fifty-eight percent of students had an average composite score of 0, indicating that they scored at the Beginning level on all of the tasks they completed. Of the more than 3,400 students included in the analytic sample, only 13 students scored at the Mastery level for all six items (0.38%).

Table 10
Distribution of the Composite Scores

Composite Score	*N*	%
0.0	1,974	58.0%
0.17 to 0.49	966	28.4%
0.5 to 0.99	351	10.3%
1.0 to 1.49	88	2.6%
1.5 to 1.99	10	0.29%
2.0	13	0.38%
Total	**3,402***	**100%**

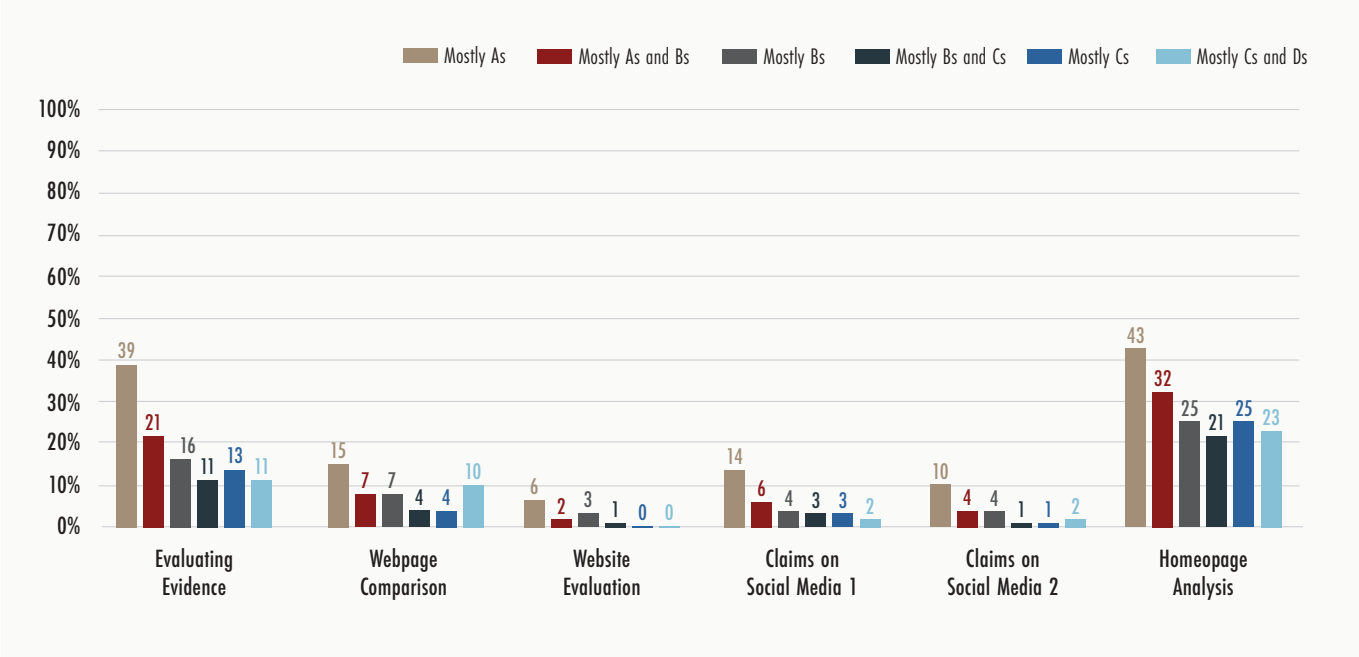
**Forty-two students in the analytic sample dropped out of the composite score analysis because the students indicated technical problems or provided nonsensical answers.*

Multivariate Models. To analyze the relationship between student characteristics and COR performance, we fit a single two-level hierarchical linear model (HLM), with random intercepts at the school level. We included students' self-reported grades as a measure of prior academic achievement. The "student grades" ranged from 1 (student reported they earned "mostly D's") to 7 (student reported "mostly A's"). We also anticipated that students in upper grades might

[6]To ensure that combining the items into a single score was psychometrically sound, we conducted an exploratory factor analysis and found that the correlation patterns between items supported constructing a single scale. The eigenvalue for a single factor scale was 3.04, and a single factor model accounted for nearly 60% of the total variance in scores. We also estimated internal consistency and found that Cronbach's alpha was .58. Tasks 4 and 5 share the same stimulus, so we checked to ensure that scores for these two items were sufficiently independent to treat as separate items. Scores for these two items were only weakly correlated (r=.19), and crosstabs showed that there was little overlap in scores across the two questions.

[7]We used this approach because we observed a non-trivial amount of missing data in the completed assessments. If we had summed the raw scores and required students to have scores on all six items to be included in the analysis, we would have lost a large (and non-randomly discarded) proportion of the sample. Alternatively, if we had used a sum approach and included all students, we would have inadvertently penalized those who had fewer than six scored tasks (even if they had attained mastery scores on the ones they completed). The mean approach allowed us to retain the maximum number of students while not penalizing students who did not complete all six scored items.

Figure 11
Percent Scoring Emerging or Mastery by Average Grades and Task



Note: This figure includes 2,808 students who selected an average grades category though the number of responding students varied by task. Excluded categories include "Mostly D's," "Prefer Not to Say," and "Missing." See Appendix C for student counts by task for all available response options for this question.

perform better than students in lower grades, so we included grade level in our model.

Additionally, we included students' geographic region and locale to adjust for any systematic differences by student location. Finally, we included demographic variables that our descriptive portrait suggested might be relevant to student performance on COR tasks, including race/ethnicity, gender, free/reduced lunch status, maternal education, and first language spoken at home. Appendix D describes each of these variables and provides details about how each was measured.

Multivariate Results. Our model revealed significant differences in students' composite scores for a variety of demographic factors. Table 11 shows the regression coefficients and standard errors for each of the variables in our model. Adjusting for all other included variables, students who reported earning higher grades performed significantly better than students who reported earning lower grades. A one-step increase in self-reported grades (e.g., from "mostly B's" to "mostly A's and B's") was associated with a .03 improvement in average composite score (a noticeable improvement in a scale whose maximum is 2). Similarly, grade level was a significant predictor of student performance. Each increase in grade level (e.g., from sophomore to junior) was associated with a .03 increase in students' composite scores for students who are similar for all other reported characteristics.

We also observed significant results for proxy measures of socioeconomic status. Maternal education was a significant predictor of average composite score. For every one-step increase in a mother's reported education level (e.g., from less than high school to high school, or some college to a bachelor's degree), performance improved by about .01 points on average. Free/reduced price lunch status was also a significant predictor of performance, with students eligible to receive free/reduced lunch performing -.04 points lower,

Table 11
Results from Mixed Effects Regression Model

	Composite Score
Student Grades	0.029 (.00)***
National School Lunch Program Eligible	-0.044 (.02)**
Mother's Education	0.014 (0.01)*
Grade Level	0.029 (.07)**
Region (compared with Midwest)	
Northeast	-0.036 (.04)
South	0.009 (.03)
West	0.001 (.04)
Locality (compared with Rural)	
Suburban	0.043 (.03)
Urban	0.107 (.03)***
Gender (compared with Female)	
Male	0.017 (.01)
Non-Binary/Prefer to Self-Describe	-0.020 (.03)
Home Language (compared with English)	
Other	-0.036 (.03)
Spanish	-0.012 (.02)
Race/Ethnicity (compared with White)	
Asian	0.019 (.03)
Black or African-American	-0.120 (.03)***
Other	-0.083 (.04)*
Two or more races	-0.039 (.02)
Hispanic	-0.014 (.18)
_cons	-0.157 (.05)**
Variance: School	0.003 (.001)*
Schools: *n*	31: 1,690
Wald X^2	87.59

Note: *$p < .05$, **$p < .01$, *** $p < .001$; Standard errors in parentheses.

Table 12
Contrasts of Marginal Effects for Race/Ethnicity and Locality

	Contrast
Race/Ethnicity	
Black vs. Asian	-0.139 (.04)**
Black vs. White	-0.120 (.03) ***
Black vs. Other	-0.037 (.04)
Black vs. Two or more races	-0.081 (.03)**
Black vs. Hispanic	-0.105 (.03)**
Locality	
Urban vs. Rural	0.107*(0.03)
Urban vs. Suburban	0.064*(0.03)

*Note: *p < .05, **p <.01, *** p<.001; Standard errors in parentheses.*

on average, than students not eligible, adjusting for all other variables in the model.

Our model also revealed significant differences by locale. In our sample, urban students significantly outperformed their rural and suburban peers. Adjusting for all other variables, urban students scored .11 points higher than rural students and .06 points higher than suburban students. Table 12 includes results for each locale comparison.

Race/ethnicity was also a significant factor in our model. Even after adjusting for all other variables, Black/African American students in our sample had lower composite scores than students who were Asian/Pacific Islander (-.14 points), Hispanic (-.11 points), multiracial (-.08 points), or White (-.12 points). Table 12 includes results for each race/ethnicity comparison.

Other variables were not significant predictors of student scores. Gender was not a significant predictor, and the differences by geographic region and home language we observed in the descriptive portrait were also not statistically significant when adjusting for all other variables.

DISCUSSION

Our results are sobering. Despite intense interest in digital literacy in the wake of the 2016 election, students remain unprepared to navigate the digital landscape. When presented with an anonymously posted Facebook video claiming to show voter fraud in the United States, less than nine percent of students formulated a coherent explanation for why it constituted a problem. Fifty-two percent believed that the video, which actually showed footage from Russia, provided *strong* evidence of voter fraud in the U.S. Such credulity makes it too easy for bad actors to undermine faith in the democratic process. Students displayed a troubling tendency to accept websites at face value. Over 96% did not consider why ties between a climate change website and the fossil fuel

industry might lessen that website's credibility. Instead of investigating who was behind the site, students focused on weak markers of credibility: the site's aesthetics, its top-level domain, its About page or non-profit status.

The vast majority of students we surveyed would benefit from carefully designed, thoughtfully delivered, and rigorously evaluated curriculum. At the same time, our findings suggest that, when it comes to evaluating the quality of digital sources, those most affected are students who have been underserved by our nation's schools. Students' socioeconomic status and their ethnicity/race were significant predictors of performance. Equitable access to civic life depends on providing these students with the support they need to develop the skills of digital evaluation.

The results of our study should in no way be interpreted as an indictment of students. The American educational system lags behind the dizzying changes wrought by the information revolution. The "checklist approach" remains the dominant means of teaching students to evaluate online content. Students are provided with lengthy lists of questions that focus attention on a single website. The CRAAP Test (Meriam Library, California State University, Chico, 2010), a ubiquitous tool that appears on scores of college and university websites, is the most widely available example of digital curriculum. This approach instructs students to answer a variety of questions about a site's currency, relevance, authority, accuracy and purpose (hence, CRAAP). By focusing students' attention on a single site, rather than teaching them how to consult the broader web to establish a site's trustworthiness, the CRAAP test is inconsistent with how expert evaluators reached sound judgments (Wineburg & McGrew, 2019). Despite its widespread use, the CRAAP Test has no basis in research and can actually lead students dangerously astray (Breakstone, McGrew, Smith, Ortega, & Wineburg, 2018).

We desperately need research-based approaches to digital literacy instruction. Small-scale but promising studies conducted by our research group, as well as by others, suggest that it's possible to "move the needle" and improve students' digital skills—sometimes even after only a brief intervention (Brodsky et al., 2019; Kahne & Bowyer, 2017; McGrew, 2019; McGrew, Smith, Breakstone, Ortega, & Wineburg, 2019). Clearly, a much greater investment in research and development is needed to prepare students to make wise decisions about issues that affect them, their families, their communities, and their world.

CONCLUSION

Reliable information is to civic health what proper sanitation and potable water are to public health. High-quality educational materials, validated by research, and distributed freely are essential to sustaining the vitality of American democracy.

Educational systems move slowly. Technology doesn't. If we don't act with urgency, our students' ability to engage in civic life will be the casualty.

This research was supported by the William and Flora Hewlett Foundation, Grant 2017-5939, Sam Wineburg, Principal Investigator. The content is solely the responsibility of the authors and does not necessarily represent the views of the William and Flora Hewlett Foundation.

The authors acknowledge Rex Long and Sarah Narvaiz from Gibson Consulting for their roles in data collection and scoring.

References

Bray, H. (2016, November 16). Are you ready for a war on fake news? *Boston Globe*. Retrieved from https://www.bostonglobe.com/business/2016/11/16/fake-news-first-casualty-trump-election/PcPpY0tL7t-6BqoUWGf5yJL/story.html

Breakstone, J., McGrew, S., Smith, M., Ortega, T., & Wineburg, S. (2018). Why we need a new approach to teaching digital literacy. *Phi Delta Kappan*, *99*, 27-32. doi:10.1177/0031721718762419

Brodsky, J.E., Brooks, P.J., Scimeca, D., Todorova, R., Galati, P., Batson, M., . . . Caulfield, M. (2019). *Teaching college students the four moves of expert fact-checkers*. Paper presented at the Association for Psychological Science Annual Convention, Washington, D.C.

Hobbs, R. (2010). *Digital and media literacy: A plan of action*. Washington, D.C.: The Aspen Institute.

Kahne, J, & Bowyer, B. (2017). Educating for democracy in a partisan age: Confronting the challenges of motivated reasoning and misinformation. *American Educational Research Journal*, *54* (1), 3-34.

Kahne, J., Lee, N., & Feezell, J. (2012). Digital media literacy education and online civic and political participation. *International Journal of Communication*, *6*, 1–24.

Kavanagh, J., & Rich, M.D. (2018). *Truth decay: An initial exploration of the diminishing role of facts and analysis in American public life*. Santa Monica, CA: Rand Corporation. Retrieved from https://www.rand.org/pubs/research_reports/RR2314.html

Knight Commission on the Information Needs of Communities in a Democracy. (2009). *Informing communities: Sustaining democracy in the digital age*. Retrieved from https://knightfoundation.org/reports/informing-communities-sustaining-democracy-digital

McGrew, S. (2019). Learning to evaluate: An intervention in civic online reasoning. *Computers & Education*. Advance online publication. doi:10.1016/j.compedu.2019.103711

McGrew, S., Breakstone, J., Ortega, T., Smith, M., & Wineburg, S. (2018). Can students evaluate online sources? Learning from assessments of civic online reasoning. *Theory & Research in Social Education*, *46*, 165–193. doi:10.1080/00933104.2017.1416320

McGrew, S., Ortega, T., Breakstone, J., & Wineburg, S. (2017). The challenge that's bigger than fake news: Civic reasoning in a social media environment. *American Educator*, *41* (3), 4-11.

McGrew, S., Smith, M., Breakstone, J., Ortega, T., & Wineburg, S. (2019). Improving university students' web savvy: An intervention study. *British Journal of Educational Psychology*, *89*, 485-500. doi:10.1111/bjep.12279

Meriam Library, California State University, Chico. (2010). *Evaluating information — applying the CRAAP test*. Retrieved from www.csuchico.edu/lins/handouts/eval_websites.pdf

Metzger, M. J. (2007). Making sense of credibility on the web: Models for evaluating online information and recommendations for future research. *Journal of the American Society for Information Science and Technology*, *58*(13), 2078–2091.

Metzger, M. J., Flanagin, A. J., & Medders, R. B. (2010). Social and heuristic approaches to credibility evaluation online. *Journal of Communication*, *60*, 413–439. doi:10.1111/j.1460-2466.2010.01488.x

Mihailidis, P., & Thevenin, B. (2013). Media literacy as a core competency for engaged citizenship in participatory democracy. *American Behavioral Scientist*, *57*, 1611–1622. doi:10.1177/0002764213489015

New York Times. (2016, November 19). Facebook and the digital virus called fake news. *New York Times*. Retrieved from https://www.nytimes.com/2016/11/20/opinion/sunday/facebook-and-the-digital-virus-called-fake-news.html

Reardon, S. F., & Townsend, J. B. (2018). SEG: Stata module to compute multiple-group diversity and segregation indices. Retrieved from https://econpapers.repec.org/software/bocbocode/s375001.htm

Reich, R., Dorn, L. & Sutton, S. (2009). *Anything goes: Approval of nonprofit status by the IRS*. Stanford, CA: Stanford Center on Philanthropy and Civil Society. Retrieved from https://pacscenter.stanford.edu/publication/anything-goes-approval-of-nonprofit-status-by-the-irs/

Shellenbarger, S. (2016, November 21). Most students don't know when news is fake, Stanford study finds. *Wall Street Journal.* Retrieved from https://www.wsj.com/articles/most-students-dont-know-when-news-is-fake-stanford-study-finds-1479752576

Stanford History Education Group. (2016). *Evaluating information: The cornerstone of civic online reasoning*. Stanford, CA: Stanford University. Retrieved from https://purl.stanford.edu/fv751yt5934

Tapper, J. (2016, November 16). Fake news stories thriving on social media. CNN. Retrieved from https://www.cnn.com/videos/politics/2016/11/17/fake-news-social-media-tapper-dnt-lead.cnn

Todd, C. (2016, November 17). The fight against fake news. NBC News. Retrieved from https://www.nbcnews.com/dateline/video/the-fight-against-fake-news-811761219833

Wineburg, S. (2018a). *Why learn history (when it is already on your phone)*. Chicago: University of Chicago Press.

Wineburg, S. (2018b, October 1). Why California's new media literacy law for schools could backfire. *Washington Post.* Retrieved from https://washingtonpost.com/education/2018/10/01/why-californias-new-media-literacy-law-schools-could-backfire/

Wineburg S., & McGrew, S. (2019). Lateral reading and the nature of expertise: Reading less and learning more when evaluating digital information. *Teachers College Record, 121*(11). Advance online publication.

APPENDICES

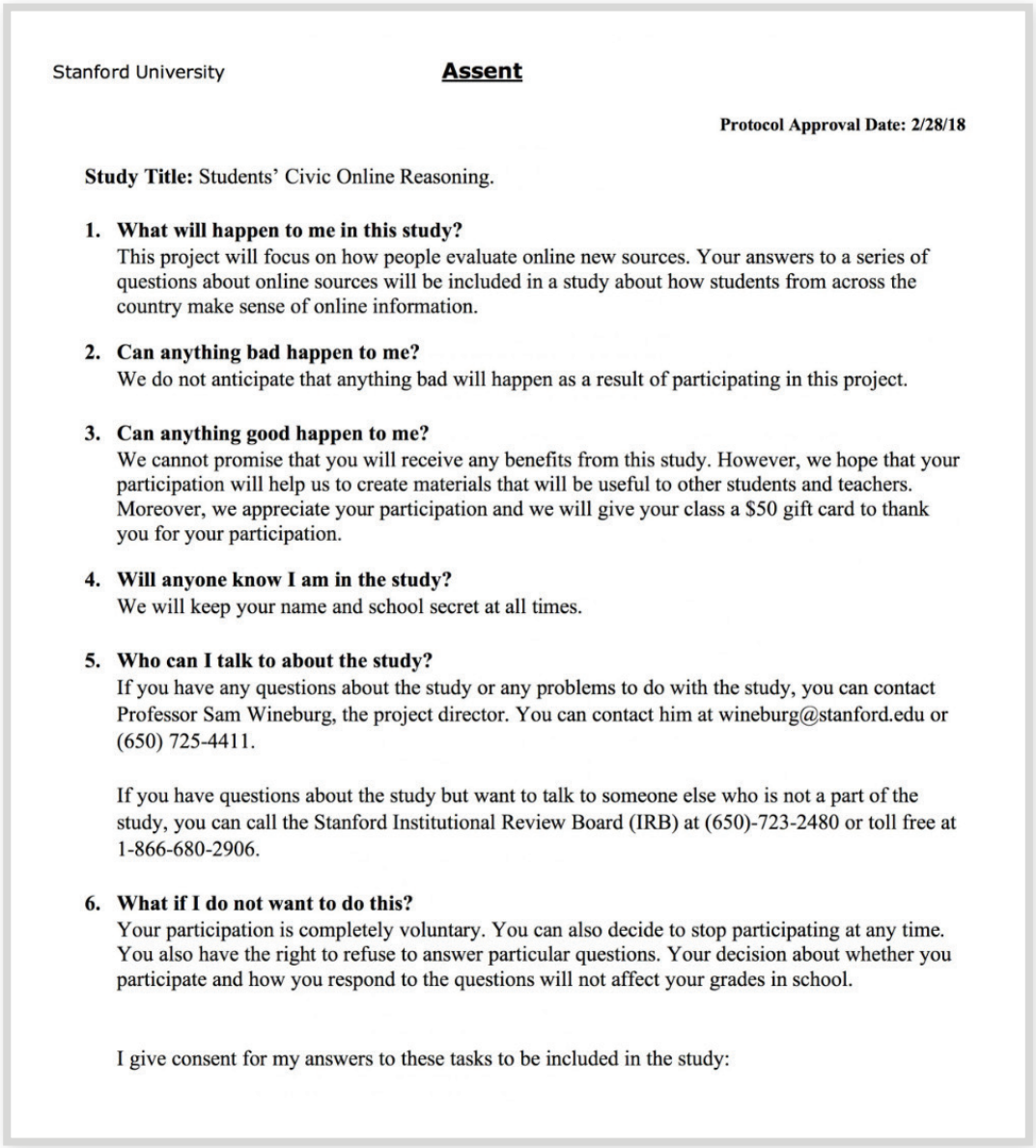
Appendix A: Online Assessment

Stanford History Education Group: Students' Online Reasoning Study

You are invited to participate in a research project that looks at how people evaluate online sources. Findings from this project will help us create resources to teach students how to navigate and evaluate information on the Internet.

Completing the tasks will take about 30-40 minutes. Your answers are completely anonymous. Your answers will not affect your grade in this course in any way.

Please read the assent form below.



Stanford University **Assent**

Protocol Approval Date: 2/28/18

Study Title: Students' Civic Online Reasoning.

1. **What will happen to me in this study?**
   This project will focus on how people evaluate online new sources. Your answers to a series of questions about online sources will be included in a study about how students from across the country make sense of online information.

2. **Can anything bad happen to me?**
   We do not anticipate that anything bad will happen as a result of participating in this project.

3. **Can anything good happen to me?**
   We cannot promise that you will receive any benefits from this study. However, we hope that your participation will help us to create materials that will be useful to other students and teachers. Moreover, we appreciate your participation and we will give your class a $50 gift card to thank you for your participation.

4. **Will anyone know I am in the study?**
   We will keep your name and school secret at all times.

5. **Who can I talk to about the study?**
   If you have any questions about the study or any problems to do with the study, you can contact Professor Sam Wineburg, the project director. You can contact him at wineburg@stanford.edu or (650) 725-4411.

   If you have questions about the study but want to talk to someone else who is not a part of the study, you can call the Stanford Institutional Review Board (IRB) at (650)-723-2480 or toll free at 1-866-680-2906.

6. **What if I do not want to do this?**
   Your participation is completely voluntary. You can also decide to stop participating at any time. You also have the right to refuse to answer particular questions. Your decision about whether you participate and how you respond to the questions will not affect your grades in school.

   I give consent for my answers to these tasks to be included in the study:

Click here to download a copy of the assent form for you to keep.

I consent to have my responses included in this study

- Yes
- No – Selecting “No” will end this activity

Please sign here to consent to participate in this study. Use your mouse or track pad to sign.

What is today’s date? Please use the following format: MM/DD/YYYY

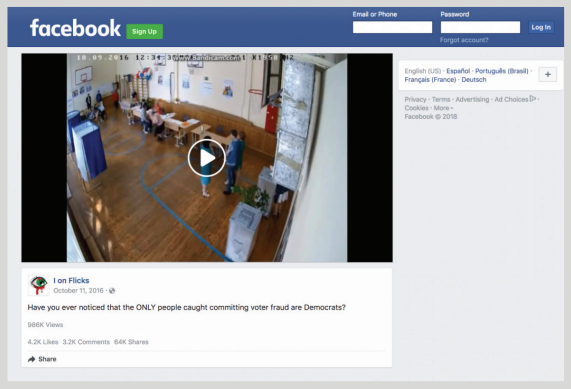
On the following pages you will be presented with five tasks. Please keep the following in mind as you complete each one:

- You may open a new tab and complete an Internet search to help you answer any of the questions.
- Once you advance, you will not be able to return to previous pages. Your browser's back button won't work either.

Thank you for participating!

Task 1 of 5

Below is a screenshot of a Facebook post from October 2016. Please watch the video posted below the screenshot.



You can watch the video from the post here: (Note, there is no audio.)

Question: Does the Facebook post provide strong evidence of voter fraud during the 2016 Democratic primary election? (The primary was the election to decide who would be the Democratic candidate for president.)

- Yes
- No

Explain your reasoning.

Were you able to view the video above? (Note: The video does not have sound.)

- Yes
- No

Task 2 of 5

Imagine you are doing research on gun control and you find the following webpages. Because webpages can change frequently, these links will take you to archived versions of the articles. Please do not consider that in your evaluations.

a) http://web.archive.org/web/20161205141632/http://people.duke.edu/~gnsmith/articles/myths.htm
b) http://web.archive.org/web/20170222133426/https://en.wikipedia.org/wiki/Gun_politics_in_the_United_States

Question: Which of these two webpages do you think is a better place to start your research?

- http://people.duke.edu/~gnsmith/articles/myths.htm
- https://en.wikipedia.org/wiki/Gun_politics_in_the_United_States

Explain your choice. Refer to both webpages in your answer.

Were you able to access both links about gun control?

- Yes
- No

Task 3 of 5

Visit the following webpage: http://www.co2science.org/

Question: Is this a reliable source of information about global warming? You may use any information on this website, or you can open a new tab and do an Internet search if you want.

- Yes
- No

Explain your answer, citing evidence from the webpages you used. Be sure to provide the URLs to the webpages you cite.

Were you able to access the above link about climate science?

- Yes
- No

Task 4 of 5

The following tweet appears in your Twitter feed (please click on the link to view):
https://twitter.com/MoveOn/status/666772893846675456?lang=en

Question: Why might this tweet be a useful source about NRA members' opinions on background checks? List any sources you used to make your decision.

Question: Why might this tweet not be a useful source about NRA members' opinions on background checks? List any sources you used to make your decision.

Were you able to view the tweet above?

- Yes
- No

Task 5 of 5

Here is the home page of Slate.com. Some of the things that appear on Slate.com are news stories, and others are advertisements.

Question: Is #1 an advertisement?

- Yes
- No

Explain your reasoning.

Question: Is #2 an advertisement?

- Yes
- No

Explain your reasoning.

Question: Is #3 an advertisement?

- Yes
- No

Explain your reasoning.

Additional Questions

On a scale of 0-10, how seriously did you take these tasks? "0" indicates that you did not take the tasks seriously at all. "5" indicates that you took the tasks fairly seriously. "10" indicates that you took the tasks very seriously.

1	2	3	4	5	6	7	8	9	10

Thank you for completing this exercise! Please take a few minutes to answer the questions below. Your answers to these questions help us learn about the students who completed the exercises. All of your responses are still anonymous.

Please select your school district.

Please select your school.

In which class you are completing this survey?	**Is this class an Honors or IB class?**	**What grade are you in?**
• US/American History I • US/American History II • World History • AP US History • AP US Government & Politics • Other (please enter class name	• Yes, Honors • Yes, IB • No	• 8th grade or below • 9th • 10th • 11th • 12th

How old are you?		**What is your gender?**	**Are you Hispanic or Latino?**
• 12 or younger • 13 • 14 • 15 • 16	• 17 • 18 • 19 • 20 or older	• Female • Male • Non-binary • Prefer to self-describe • Prefer not to say	• Yes • No • Prefer not to say

What is your race? **Mark one of more boxes**	**What is the highest degree or level of school that your mother completed?**
• American Indian or Alaska Native • Asian • Black or African American • Native Hawaiian or Other Pacific Islander • White • Other • Prefer not to say	• Less than high school • High school • Some college • Bachelor's degree • Advanced degree (e.g., MA, JD, MD, PhD). • I don't know • Prefer not to say

What is the language(s) most used in your home? Drag and drop up to three languages from the left column to the right box. Then rank them in order of how much they're used in your home.

- English
- Arabic
- Burmese
- Cambodian
- Cantonese
- Cape Verdean Creole
- French
- Greek
- Hmong
- Haitian-Creole
- Italian
- Korean
- Mandarin
- Portuguese
- Russian
- Somali
- Spanish
- Toishanese
- Vietnamese
- Other (Please specify):
- Prefer not to say

Most used languages(s) in your home (select and rank up to 3)

What is the zip code where you live? (If you live in multiple zip codes, choose the one in which you spend more time.)

Are you eligible for free or reduced-price lunch at school?

- Yes
- No
- I don't know.
- Prefer not to say

Which best describes the grades you usually earn on your report card?

- Mostly A's
- Mostly A's and B's
- Mostly B's
- Mostly B's and C's
- Mostly C's
- Mostly C's and D's
- Mostly D's
- Prefer not to say

How many hours a day do you typically spend online? Include the time you spend on applications like Instagram, Snapchat, etc.

- 0
- 1 to 3
- 4 to 6
- 7 to 9
- 10 to 12
- 13 or more

Select the three online activities you do most on an average day and drag them to the box on the right.

- Checking and sending email
- Keeping up what with friends and family are doing
- Visiting accounts by people you don't personally know (meme accounts, celebrity account, etc.)
- Research topics you are interested in
- Keeping up with current events
- Checking the weather, traffic, or public transportation
- Shopping and researching products
- Finding information about events, movies, restaurants, etc.
- Playing Games
- Other(s) (Please specify)

Top 3 Online Activities

In the last week, where did you get information about the news or current events? Check all that apply.

- Television
- Print newspaper (e.g., New York Times, Wall Street Journal)
- News website (e.g., Huffington Post, Wall Street Journal website, CNN.com)
- Social media (e.g., Snapchat, Instagram, Twitter, YouTube)
- Radio or podcasts
- Talk with family or friends
- Teachers
- I don't keep up with the news.
- Other (Please specify)

From which social media websites or apps did you get the most information about the news or current events this past week (select up to 3).

- Facebook
- Twitter
- Instagram
- Snapchat
- YouTube
- Tumblr
- Reddit
- WhatsApp, GroupMe, Kik, or other group messaging app
- Other (Please specify)

On a scale of 0-10, indicate how skilled or knowledgeable you think you are in comparison to other Internet users in the following areas:

- **"0" indicates that you think you are much less skilled than other Internet users**
- **"10" indicates that you think you are much more skilled than other Internet users**
- **"5" indicates that you think you are about as skilled as other Internet users**

Much less skilled than other users			About as skilled as other users				Much more skilled than other users		
1	2	3	4	5	6	7	8	9	10

Technical skills (e.g., fixing connection problems, changing computer settings)

1	2	3	4	5	6	7	8	9	10

Search skills (ability to find what you are looking for online)

1	2	3	4	5	6	7	8	9	10

Evaluation skills (ability to effective evaluate the trustworthiness of online information)

1	2	3	4	5	6	7	8	9	10

I check whether the information I find online is trustworthy:

- Always
- Often
- Sometimes
- Not often
- Never

Do you have your own smartphone?

- Yes
- No

Appendix B: Sample Rubric for Task 1 (Evaluating Evidence)

Task 1: Evaluating Evidence Rubric

This assessment asks students to determine whether a video posted on Facebook provides strong evidence of voter fraud in the 2016 election. Successful students will conclude that the video is not strong evidence and provide a clear explanation why it isn't. Successful explanations can focus on a variety factors, including: (1) there is no evidence that the video was actually shot where the poster claims it was; (2) nothing is known about the identity of the person who posted the video (I on Flicks) and they are likely not a reliable source of information; (3) after doing research on the video, it is clear that the video is not what the poster claims it is; or (4) after researching the identity of the person who posted the video, it is clear that "I on Flicks" may not be a reliable source of information.

(2) MASTERY	Student rejects the video as strong evidence and does at least one of the following: • Questions who posted the video • Questions the source/location of the video • Investigates who posted the video and concludes they are unreliable • Investigates the video and determines the recordings are not from the 2016 Democratic primary
(1) EMERGING	Student questions the content of the video rather than the source, implying or stating that an unedited video would be automatically reliable **OR** Student questions the source of the post/video, but also does one of the following: • Waffles/straddles the fence, both accepting the video as good evidence but also raising relevant concerns • States incorrect reasoning for why the post/video is unreliable
(0) BEGINNING	Student rejects the video but uses completely incorrect or incoherent reasoning. **OR** Student concludes the post <u>does</u> provide strong evidence.

Appendix C: Descriptive Results Tables

Table C1

Percent Scoring Emerging or Mastery by Locale for Tasks 1 to 5

Locale	*N*	Evaluating Evidence %	Webpage Comparison %	Website Evaluation %	Claims on Social Media 1 %	Claims on Social Media 2 %	Home Page Analysis %
Rural	820	20.9%	4.2%	2.6%	5.0%	2.3%	23.1%
Suburban	1,659	20.9%	7.9%	2.7%	6.8%	3.7%	31.0%
Urban	967	27.3%	11.2%	4.6%	10.6%	9.3%	39.4%
Total	3,446	3,119	2,678	2,994	2,431	2,431	2,969

Table C2

Percent Scoring Emerging or Mastery by Region for Tasks 1 to 5

Region	*N*	Evaluating Evidence %	Webpage Comparison %	Website Evaluation %	Claims on Social Media 1 %	Claims on Social Media 2 %	Home Page Analysis %
Midwest	852	16.4%	4.5%	1.7%	4.0%	2.5%	26.9%
Northeast	732	15.3%	5.2%	2.3%	4.2%	1.1%	22.4%
South	1,084	27.6%	10.4%	4.2%	10.0%	7.2%	37.1%
West	778	29.2%	10.1%	3.9%	5.4%	3.7%	36.4%
Total	3,446	3,119	2,678	2,994	2,431	2,431	2,969

Table C3

Percent Scoring Emerging or Mastery by Grade for Tasks 1 to 5

Grade	*N*	Evaluating Evidence %	Webpage Comparison %	Website Evaluation %	Claims on Social Media 1 %	Claims on Social Media 2 %	Home Page Analysis %
8th Grade	17	14.3%	0.0%	0.0%	0.0%	0.0%	13.3%
9th Grade	820	15.1%	5.2%	1.3%	4.2%	1.5%	24.4%
10th Grade	491	19.8%	5.8%	1.5%	6.0%	4.3%	25.7%
11th Grade	867	26.7%	9.6%	3.6%	7.2%	5.5%	36.1%
12th Grade	761	29.2%	11.8%	5.6%	12.1%	8.0%	38.8%
Missing	490	20.0%	4.9%	3.8%	5.5%	1.1%	10.9%
Total (*N*)	3,446	3,119	2,678	2,994	2,431	2,431	2,969

Table C4

Percent Scoring Emerging or Mastery by Race/Ethnicity and Task

Race/Ethnicity	*N*	Eval. Evidence %	Webpage Comparison %	Website Evaluation %	Claims on Social Media 1 %	Claims on Social Media 2 %	Home Page Analysis %
Asian/Pacific Islander American	112	33.6%	22.8%	5.4%	16.7%	7.2%	43.8%
Indian/Alaska Native	29	17.9%	3.8%	0.0%	11.5%	3.8%	31.0%
Black or African American	306	11.0%	2.8%	0.0%	2.2%	1.3%	17.6%
Hispanic	698	17.6%	3.9%	0.9%	3.9%	3.2%	24.7%
Other	42	13.5%	6.7%	2.7%	6.5%	9.4%	31.0%
Two or more races	224	29.5%	9.0%	3.2%	6.5%	6.0%	32.6%
White	1,373	27.8%	11.5%	5.0%	9.7%	6.1%	38.3%
Prefer not to say	28	19.2%	10.0%	7.4%	8.7%	4.3%	37.0%
Missing	634	18.0%	3.6%	2.2%	3.6%	0.5%	17.0%
Total	3,446	3,119	2,678	2,994	2,431	2,431	2,969

Table C5

Percent Scoring Emerging or Mastery by Average Grades for Tasks 1 to 5

Grades	*N*	Evaluating Evidence %	Webpage Comparison %	Website Evaluation %	Claims on Social Media 1 %	Claims on Social Media 2 %	Home Page Analysis %
Mostly As	794	38.9%	14.7%	6.5%	13.8%	10.0%	43.1%
Mostly As and Bs	1,116	21.4%	6.6%	2.4%	6.4%	3.5%	32.1%
Mostly Bs	268	16.4%	6.9%	2.7%	4.4%	3.8%	24.8%
Mostly Bs and Cs	445	11.1%	3.7%	1.0%	2.8%	1.1%	21.1%
Mostly Cs	108	12.9%	3.5%	0.0%	2.8%	1.4%	24.5%
Mostly Cs and Ds	77	10.8%	10.2%	0.0%	1.8%	1.8%	23.4%
Mostly Ds	33	7.7%	0.0%	0.0%	0.0%	0.0%	26.7%
Prefer not to say	101	7.8%	3.6%	1.1%	1.4%	0.0%	20.0%
Missing	504	19.5%	5.0%	3.5%	4.9%	1.0%	10.7%
Total	3,446	3,119	2,678	2,994	2,431	2,431	2,969

Table C6

Percent Scoring Emerging or Mastery by Gender for Tasks 1 to 5

Gender	*N*	Evaluating Evidence %	Webpage Comparison %	Website Evaluation %	Claims on Social Media 1 %	Claims on Social Media 2 %	Home Page Analysis %
Female	1,443	24.5%	7.1%	2.7%	7.1%	4.3%	31.8%
Male	1,283	22.2%	10.2%	4.0%	7.8%	5.7%	32.0%
Non-binary	28	20.0%	5.0%	9.1%	0.0%	5.3%	46.2%
Prefer to self-describe	144	15.8%	1.8%	0.0%	5.3%	2.2%	25.4%
Prefer not to say	58	15.8%	1.8%	0.0%	5.3%	2.2%	25.4%
Missing	490	20.0%	4.9%	3.8%	5.5%	1.1%	10.9%
Total	3,446	3,119	2,678	2,994	2,431	2,431	2,969

Table C7

Percent Scoring Emerging or Mastery by National School Lunch Program Eligibility for Tasks 1 to 5

Eligible for School Lunch Program	*N*	Evaluating Evidence %	Webpage Comparison %	Website Evaluation %	Claims on Social Media 1 %	Claims on Social Media 2 %	Home Page Analysis %
Yes	850	16.0%	4.4%	0.5%	3.7%	0.9%	21.1%
No	1,475	29.7%	11.5%	5.3%	10.8%	8.6%	39.9%
I don't know	533	16.8%	6.1%	1.6%	4.2%	1.2%	28.9%
Prefer not to say	84	13.9%	2.7%	0.0%	3.3%	0.0%	16.0%
Missing	504	19.5%	5.0%	3.5%	4.9%	1.0%	10.7%
Total	3,446	3,119	2,678	2,994	2,431	2,431	2,969

Table C8

Percent Scoring Emerging or Mastery by Mother's Education for Tasks 1 to 5

Highest degree/level of education mother completed?	*N*	Evaluating Evidence %	Webpage Comparison %	Website Evaluation %	Claims on Social Media 1 %	Claims on Social Media 2 %	Home Page Analysis %
Less than high school	222	20.3%	5.4%	1.4%	5.9%	3.5%	25.6%
High school	506	14.5%	3.7%	0.4%	3.6%	1.3%	21.9%
Some college	460	22.8%	9.2%	2.3%	4.9%	3.3%	29.5%
Bachelors degree	736	28.1%	11.4%	4.5%	9.5%	7.2%	39.5%
Advanced degree	473	34.6%	13.4%	7.7%	15.0%	11.7%	43.6%
I don't know	411	17.4%	5.1%	1.3%	3.6%	0.3%	25.7%
Prefer not to say	148	8.1%	3.2%	0.7%	1.9%	0.0%	21.5%
Missing	490	20.0%	4.9%	3.8%	5.5%	1.1%	10.9%
Total	3,446	3,119	2,678	2,994	2,431	2,431	2,969

Table C9

Percent Scoring Emerging or Mastery by Home Language for Tasks 1 to 5

Home Language	*N*	Evaluating Evidence %	Webpage Comparison %	Website Evaluation %	Claims on Social Media 1 %	Claims on Social Media 2 %	Home Page Analysis %
English	2,425	24.8%	8.9%	3.6%	8.0%	5.4%	33.7%
Spanish	302	13.1%	2.6%	0.3%	2.1%	0.4%	21.4%
Other	219	17.8%	9.1%	2.0%	6.3%	3.2%	24.0%
Missing	500	19.5%	4.7%	3.6%	5.2%	1.1%	11.1%
Total	3,446	3,119	2,678	2,994	2,431	2,431	2,969

Appendix D: Regression Model Variables and Results

Demographic Variables

Race/Ethnicity: We included the composite race/ethnicity variable in the multivariate models. Students who did not answer the question or who answered "prefer not to say" were not included.

Gender: We included a categorical measure of gender with three categories: female, male, and a third category which combined nonbinary and "prefer to self-describe" response options. Individuals who did not answer the question or answered 'prefer not to say' are not included in the multivariate models.

Free/reduced lunch status: FRL status was a binary indicator (eligible/participating versus not eligible/not-participating). Students who did not answer the question or who responded "I don't know" or "prefer not to say" were not included in the multivariate models.

Mother's education: Student reports on their mother's highest level of education were coded 1 (less than a high school education) to 5 (advanced degree). Students who did not answer the question or who responded "I don't know" or "prefer not to say" are not included in the multivariate models.

Home language: The assessments were in English, so students for whom English was not a first language may have had more difficulty interpreting the sources and conveying their thinking in English. As a result, we included a categorical measure of home language (English versus not English) based on the students' response about the language(s) spoken at home. Students who did not answer the question were not included in the multivariate models.

Missing Data

There were a number of circumstances that reduced the total number of student responses included in each model. Missingness on tasks may be due to running out of time, technical issues, providing a nonsense response or because a student chose to stop the assessment (as participation was voluntary within the classroom setting). On the demographic and academic variables, missing data may be due to not progressing through to the end of the activity, or they could have chosen a non-substantive response, as described in the above section. Each of the six models ultimately included between 1,456 and 1,813 students.

Full Model Results

Table D1 includes coefficients, standard errors, and associated p-values for each of the six multivariate models.

Table D1

Results from Multivariate Mixed Effects Logistic Regression Models

	Evaluating Evidence	Webpage Comparison	Website Evaluation	Claims on Social Media 1	Claims on Social Media 2	Home Page Analysis
Student course marks	0.459 (.06)***	0.165 (.09)	0.499 (.16)**	0.441 (.11)***	0.387 (.13)**	0.138 (.05)**
National School Lunch Program Eligible	-0.150 (.17)	-0.095 (.28)	-0.804 (.55)	-0.281 (.29)	-1.459 (.47)**	-0.450 (.15)**
Mother's Education	0.002 (0.06)	0.159 (.09)	0.318 (.15)*	0.098 (.10)	0.154 (.12)	0.126 (.05)*
Grade level	0.193 (.06)**	0.157 (.10)	0.505 (.14)**	0.300 (.10)**	0.200 (.11)	0.156 (.05)**
Region (compared with Midwest)						
Northeast	0.307 (.36)	-0.054 (.50)	0.077 (.55)	-0.422 (.41)	-1.155 (.42)*	-0.349 (.30)
South	0.245 (.33)	0.226 (.47)	-0.049 (.46)	0.135 (.32)	0.046 (.44)	0.203 (.28)
West	0.431 (.33)	-0.019 (.48)	-0.269 (.50)	0.020 (.36)	0.079 (.52)	-0.086 (.28)
Locality (compared with Rural)						
Suburban	0.079 (.30)	1.060 (.52)*	0.220 (.42)	0.461 (.30)	0.313 (.34)	0.334 (.25)
Urban	0.220 (.32)	1.254 (.53)**	0.435 (.45)	0.477 (.36)	1.198 (.35)***	0.650 (.27)*
Gender (compared with Female)						
Male	-0.184 (.12)	0.395 (.19)*	0.640 (.25)**	0.304 (.19)	0.618 (.22)**	-0.012 (.11)
Non-Binary/Prefer to Self-Describe	-0.153 (.33)	-0.267 (.57)	-0.666 (1.04)	-0.034 (.56)	0.038 (.65)	-0.138 (.28)
Home Language (compared with English)						
Other	-0.202 (.29)	0.460 (.40)	-0.605 (.79)	-0.341 (.46)	-0.733 (.64)	-0.001 (.25)
Spanish	-0.006 (.29)	-1.30 (.78)	-0.458 (1.12)	-0.955 (.66)	-2.087 (1.06)*	0.302 (.25)
Race/Ethnicity (compared with White)						
Asian	-0.069 (.29)	0.635 (.39)	-0.337 (.58)	0.749 (.40)	-0.191 (.58)	-0.083 (.27)
Black or African-American	-0.768 (.30)**	-1.354 (.53)**	No estimate	-1.308 (.64)*	-0.413 (.65)	-0.857 (.25)**
Other	-0.429 (.43)	-1.127 (.76)	-0.890 (1.05)	-0.283 (.64)	0.641 (.60)	-0.526 (.36)
Two or more races	0.051 (.22)	-0.476 (.35)	-1.038 (.61)	-0.396 (.36)	0.131 (.38)	-0.413 (.20)*
Hispanic	-0.004 (.20)	-0.304 (.33)	-0.696 (.53)	-0.017 (.33)	1.058 (.34)**	-0.215 (.18)
_cons	-4.599 (.58)***	-5.382 (.88)***	-9.493 (1.41)***	-6.778 (.93)***	-6.858 (1.1)***	-2.547 (.45)***
Variance: School	0.210 (.09)*	0.295 (.18)*	0.071 (.17)	0.024 (.06)	0.000 (.00)	0.139 (.07)**
Schools: n	31: 1,690	31: 1,507	31: 1,604	31: 1,456	31: 1,458	31: 1,813
Wald X^2	92.76	54.13	54.59	66.84	83.31	92.55

Note: *$p < .05$, **$p < .01$, *** $p < .001$; Standard errors in parentheses.