Civic Online Reasoning
Curriculum Evaluation
The threat to democracy from a digitally credulous citizenry is nothing less than an issue of national defense. Facing this challenge will require a renewed educational commitment.
4 Executive Summary

5 Background

5 A New Approach

6 MediaWise Partnership

6 Curriculum Development
6 Sample Lesson
6 Assessments

8 Phases of Implementation
8 Pilot Study
8 Field Experiment

9 Changes in Reasoning

11 Student Interviews

12 Conclusion

13 Methodological Appendix
14 Pilot Test
14 Field Experiment

How to cite this report: Sam Wineburg, Joel Breakstone, Mark Smith, Sarah McGrew, and Teresa Ortega. “Civic Online Reasoning: Curriculum Evaluation” (working paper 2019-A2, Stanford History Education Group, Stanford University, Stanford, 2019).
Executive Summary

Compared with students in regular classrooms, students in Civic Online Reasoning classrooms improved significantly in their ability to evaluate online sources.

Background
In 2018 and 2019, our team administered an assessment of online reasoning to 3,446 high school students across the country and found they were vexed by even basic evaluations of digital sources. How can we help students become smarter consumers of digital information? To develop a roadmap, we observed fact checkers at the nation’s leading news outlets and distilled their strategies into an instructional approach we call Civic Online Reasoning (COR)—the ability to search for, evaluate, and verify social and political information.

MediaWise
In 2018, the Stanford History Education Group (SHEG) partnered with the Poynter Institute and the Local Media Association to create MediaWise, an initiative supported by Google.org. We worked alongside classroom teachers to develop curriculum for teaching the skills of Civic Online Reasoning.

Research
A pilot study in a Northern California school district showed modest but promising gains. We fine-tuned our approach and then conducted a full-scale experiment with students in a large Midwestern school district. Compared with students in regular classrooms, students in COR classrooms improved significantly in their ability to evaluate online sources.

COR Curriculum
The COR curriculum is available for free at cor.stanford.edu. The suite of lessons can be implemented as a full curriculum or taught as a series of stand-alone lessons. Our materials are accompanied by instructional videos introducing the approach, easy-to-use assessments for tracking students’ progress, and a ten-episode video series, Navigating Digital Information, by noted author and YouTube star John Green.

What’s at Stake
The threat to democracy from a digitally credulous citizenry is nothing less than an issue of national defense. Facing this challenge will require a renewed educational commitment that acknowledges the depth of the problem and meets it with vigor and resolve. We offer this curriculum as a start in that direction.
**Background**

In November 2016, the Stanford History Education Group (SHEG) released a report showing that students struggled when asked to evaluate digital sources. Middle school students confused advertisements with news stories. High school students accepted posts on social media at face value. College students fell flat trying to identify the organization behind a website.

The 2016 presidential election prompted national soul searching about the corrosive effects of online misinformation. A raft of initiatives, including legislative actions in 18 states, sought to address digital illiteracy. What have these efforts wrought? Three years after our original study, we set out to find answers.

From June 2018 to May 2019, we surveyed 3,446 high school students in 14 states, a sample that matched the demographic profile of high school students in the United States. The picture that emerged was troubling. Fifty-two percent believed that a Facebook video allegedly showing ballot stuffing provided “strong evidence” of voter fraud during the 2016 Democratic primary elections. (The video was actually shot in Russia—a fact easily established with a basic search.) Two-thirds could not tell the difference between ads marked “Sponsored Content” and news stories on the homepage of a news site. Ninety-six percent could not identify connections between a climate change denial website and the fossil fuel industry.

**A New Approach**

The most popular approaches to teaching digital literacy don’t seem to be working. “Checklist” methods (best known by the acronyms ABC and CRAAP) provide students with long lists of questions to ask before deeming a site credible. However, research shows that this is the opposite of what experts do when landing on an unfamiliar site.

We know because we observed fact checkers from the nation’s leading news outlets. Rather than wasting time on an unfamiliar site, fact checkers almost immediately opened up new tabs in their browsers to search for information about the site. We refer to this strategy as *lateral reading*—the act of turning to the broader web to get a fix on one of its nodes.

Fact checkers asked three questions of a digital source: (1) Who’s behind it? (2) What’s the evidence for its claim? and (3) What do other sources say? These questions are the cornerstones of *Civic Online Reasoning*—the ability to search for, evaluate, and verify social and political information online. We distilled the fact checkers’ strategies into an instructional approach and taught it to high school and college
students in a series of “proof of concept” studies. Even brief interventions—in one case only 150 minutes—helped students improve.

MediaWise Partnership

In 2018, SHEG joined MediaWise, a collaboration with the Poynter Institute and the Local Media Association supported by Google.org. MediaWise set out to help students become more discerning consumers of online information. Based on our research with fact checkers, we developed and pilot-tested instructional materials. In addition, we partnered with John Green and Crash Course on the development of *Navigating Digital Information*, a 10-episode YouTube series. Since its launch in January 2019, the series has logged more than a million views.

Curriculum Development

Curriculum development followed a multi-step cycle. We identified key aspects of Civic Online Reasoning and drafted lessons built around actual sources from the web. Teachers across the country piloted these lessons with their students and gave us feedback about what did and didn’t work. Each lesson went through an iterative process of prototyping, piloting, revising, and re-piloting.

Sample Lesson

All of our lessons provide students with opportunities to evaluate online sources. For example, a lesson focused on “Who’s behind the information?” gives students explicit instruction in lateral reading. Students first go online to judge the reliability of a site about minimum wage policy. Based on pilot testing with thousands of students, we know that most will focus on surface level features of the site: Is it a .com or a .org? Is it free of banner ads? Are there links to reputable sources? Is there a contact listed? Rather than focus on these aspects (each of which is easy to manipulate), the teacher models lateral reading by opening new browser tabs, searching for the organization, and seeing what other, reputable sources have to say. In this case, the original site is run by a Washington, D.C.-based public relations firm supported by corporate interests. After watching their teacher model the search process, students practice lateral reading in small groups using a different example.

Assessments

We use short assessments to frequently monitor students’ progress. One of our assessments gauges whether students question the reliability of a photograph posted by an unknown user on Twitter (see Figure 1). The tweet includes a picture of a child, purportedly in Syria, lying between two graves. (The image was actually taken in Saudi Arabia as part of an art project. The “graves” are piles of rocks.) Students are asked to determine whether the tweet provides “strong evidence” about conditions for children in Syria.

A three-level rubric accompanies each assessment. The simplicity is by design: easy-to-use rubrics allow teachers to quickly identify students who have mastered a concept, those who are on the right track, and those who need more help. Figure 1 displays one of our assessments and a rubric with sample student responses.
Figure 1
Assessment & Rubric

Assessment

The civil war in Syria began in 2011 and continues through the present. This image was posted on Twitter, a social media platform, in January 2014.

Does the post provide strong evidence about conditions for children in Syria?

Explain your reasoning.

Meanwhile in Syria, a child sleeps between his parents

Gautam Trivedi
@Gotham3
11:48 PM • 16 Jan 2014

Rubric & Example Student Responses

<table>
<thead>
<tr>
<th>Level</th>
<th>Response</th>
</tr>
</thead>
<tbody>
<tr>
<td>Beginning</td>
<td>Student argues that the post provides strong evidence or uses incorrect or incoherent reasoning.</td>
</tr>
<tr>
<td>Example Student Response:</td>
<td>“It shows that some of the children have no parents and he or she is very sad to have no parents.”</td>
</tr>
<tr>
<td>Emerging</td>
<td>Student raises questions about the post but does not consider its source or the origin of the photograph.</td>
</tr>
<tr>
<td>Example Student Response:</td>
<td>“There is no evidence to prove that those are the child’s parents.”</td>
</tr>
<tr>
<td>Mastery</td>
<td>Student argues the post does not provide strong evidence and questions the source of the post (e.g., we don’t know anything about the author of the post) and/or the source of the photograph (e.g., we don’t know where the photo was taken).</td>
</tr>
<tr>
<td>Example Student Response:</td>
<td>“How do I know this picture is real? I don’t know who posted it, how they obtained this photo, or if it was even shot in Syria. I just can’t rely on this image alone.”</td>
</tr>
</tbody>
</table>
Phases of Implementation

We evaluated the COR curriculum in two phases: (a) a pilot study in Fall 2018; and (b) a field experiment in Spring 2019. (See the Methodological Appendix for full details on both.)

Pilot Study
We ran the pilot in an urban public school district on the West Coast that enrolls over 15,000 students from 7th to 12th grade. Over a quarter of the district’s students were classified as English Learners, and another 48% were “Fluent English Proficient” (English was not their primary language but they had achieved English fluency). Nearly three-quarters of the students were Latino and 70% were eligible for free or reduced-price lunch.

At the start of the school year, SHEG staff provided teachers with a one-day professional development workshop, and teachers participated in follow-up meetings during curriculum implementation. Students who participated in the pilot study took a pretest at the start of the school year and a posttest at the end of the fall semester. The pre- and posttest measured the same COR skills but used different digital sources.

Students showed small but promising gains (see Methodological Appendix). Teachers felt that the breadth of the curriculum was too wide-ranging and students needed more opportunities to practice the COR skills. In short, we had tried to pack too many topics into too short a time.

Field Experiment
We conducted a field experiment in an urban public school district in the Midwest with over 40,000 students. Nearly half of students in the district qualified for free or reduced lunch. Over one-third were non-white, and about 7% were English learners. Based on what we learned from the pilot, we revised our approach and set out to teach fewer skills more deeply.

Every second-semester government/civics class in the district participated in the study. The final sample included 464 juniors and seniors at six high schools. Two hundred sixty-five students from three high schools completed COR lessons, and 199 students from the other three high schools served as the control group. To ensure that the students in each group were similar, we matched schools based on demographic characteristics and assigned the matched schools to opposite conditions.

Teachers implementing the COR curriculum attended a one-day professional development workshop at the start of the spring semester. We introduced the teachers to Civic Online Reasoning and provided them with six lessons that had been revised based on feedback from the pilot. Over the course of the spring semester, teachers integrated these COR lessons into their classes. SHEG staff met with the teachers during the semester to provide additional training and support. Teachers in the control classes taught their normal curriculum.

Students took a pretest assessment at the beginning of the semester and a posttest at the end. Each assessment consisted of seven exercises requiring the evaluation of online sources. Students had a live internet connection and could search anywhere online.
Analysis of pre/posttest data revealed that students in the COR classrooms showed greater improvement over the course of the semester than students in the control classrooms. Figure 2 shows the change in mean scores from pretest to posttest for treatment and control. On average, students’ scores in COR classrooms improved by over two points from pre- to posttest (out of 14 possible points), while scores in the regular classrooms improved by just over half a point (see Methodological Appendix).

**Changes in Reasoning**

Students in COR classrooms became more sophisticated in their thinking. On the pretest, students completed the Evaluating Evidence task with the picture of the child (Figure 1). The posttest included a task that assessed the same skill using a different stimulus. The posttest task featured a photograph of strangely shaped daisies from Imgur, a photo-sharing site. Entitled “Fukushima Nuclear Flowers,” the photo appears with the caption: “Not much more to say, this is what happens when flowers get nuclear birth defects.” Students were asked whether the post provided “strong evidence” about the conditions near the Fukushima Daiichi Power Plant after the nuclear disaster there in 2011. Students who initially accepted the picture of the child at face value became more discerning on the posttest, asking probing questions about where the picture of the daisies was taken and who posted it (see Figure 3).
Figure 3
Student Responses to Evaluating Evidence Task on Pretest & Posttest

<table>
<thead>
<tr>
<th>Pretest</th>
<th>Posttest</th>
</tr>
</thead>
<tbody>
<tr>
<td>Student 1</td>
<td>Meanwhile in Syria, a child sleeps between his parents. The civil war in Syria began in 2011 and continues thru the present. This image was posted on Twitter, a social media platform, in January 2014. Does the post provide strong evidence about conditions for children in Syria? Explain your reasoning.</td>
</tr>
<tr>
<td>Student 2</td>
<td>It shows the war is tearing families apart and leaving many children alone.</td>
</tr>
<tr>
<td>Student 3</td>
<td>In a sense yeah, I don’t think the conditions of one child in Syria should represent the whole population of children and their struggles but it can give the viewers insight on how bad it could be.</td>
</tr>
</tbody>
</table>
Student Interviews

We interviewed 21 students from COR classrooms before and after curriculum implementation. We asked students to evaluate a series of online sources, including a grainy, anonymously posted Facebook video that claimed to show examples of voter fraud during the 2016 Democratic primary elections (the video was actually shot in Russia). Initially, one of these students, a senior, expressed mild skepticism:

*I don’t think it’s very strong evidence because they were showing it happening but they weren’t showing the total number of people who did. That would have shown how big a problem it was. If you show like four people, it is like, okay? But if they had shown how often this had happened in this election, we would have seen how big a problem it was.*

The student based her evaluation on the low number of instances of fraud depicted in the video. However, she asked no questions about who produced it or where the footage came from. After the COR lessons, the student read laterally to evaluate the trustworthiness of this source. She swiftly turned to her browser, typed the words “Pennsylvania Democratic primary voter fraud,” and discovered in 30 seconds that the video was a hoax (see Figure 4).
Conclusion

An informed digital citizen must be able to determine what to believe and what to reject. The good news is that, with instruction, students can improve.

Too often a new curriculum packs every useful strategy into a single instructional package. It’s a temptation we did everything we could to avoid. Our approach rests on three questions, not 30; the strategies we teach can be counted on one hand. Our goal is not to turn students into professional fact checkers but to help them avoid the most common pitfalls of online research. Equipped with a small set of flexible strategies, students will be better able to make sound judgments that affect them, their families, and their communities.

Our curriculum opens a wide corridor. It can be taught as a full package or implemented as stand-alone lessons that can be embedded in different subject matter. In an age when truth itself is under assault, Civic Online Reasoning provides an antidote to the conspiracy theories, pseudo-science, and fabricated history that have undermined faith in validated knowledge.

Although our focus is the school curriculum, our aspirations go beyond it. Polluted information threatens the civic health of young and old alike. Reliable data appears on screens alongside half-truths, cooked-up infographics, and downright lies. In a previous age, it took a printing press, a radio transmitter, or a television studio to reach a broad audience. Today, anyone with a smartphone can potentially reach millions. Judging the quality of information—once a task that fell on the shoulders of publishers, scholars, journalists and librarians—now falls to anyone who glances at their phone or opens their laptop.

The internet, as philosopher Michael Lynch has observed, is “both the world’s best fact-checker and the world’s best bias confirmer—often at the same time.” Only a commitment to enacting changes in our educational system will tip the scales toward the former and make sure we avoid the latter.

Following the 2016 presidential election, the country engaged in public hand wringing about how “truth decay” threatened the health of our democracy. Unfortunately, recent evidence makes clear that students remain ill-equipped to navigate the information flooding their screens.

Change will not come about with a new elective course or an updated checklist. Bringing education into the 21st century demands that we integrate these skills into all the core school subjects: history, science, language arts, and mathematics. Students need frequent, structured practice to become adept at navigating the web’s terrain.

Our materials can help educators begin to meet this challenge.
Methodological Appendix
Methodological Appendix

Both the pilot test and the field experiment used cluster-randomized, repeated-measures designs. Unlike participants in fully randomized designs, students in school-based research are often not completely independent. Instead, they are nested (or clustered) in classrooms and schools. To account for nesting in our studies, we used multilevel linear mixed models. Below, we describe our models and outline the results for each study.

Pilot Test
In the pilot test, we investigated whether students scored significantly better at posttest than at pretest. Our pre/post assessment had two parallel forms (Form A and Form B). Each form had five tasks. The questions for each task were identical on both forms, but the online content students evaluated for each question was different. Two hundred thirty-nine students took Form A at pretest and Form B at posttest; 202 took the forms in the opposite order. Mixing the order of the forms ensured that differences in pre/post scores were attributable to student learning, not to differences in the difficulty of the forms.

In our model, students \((n = 441)\) were nested in teachers \((n = 6)\) at one school. Order of test administration \((A \rightarrow B\) or \(B \rightarrow A)\) and time (pre to post) were treated as fixed effects, and students nested in teachers and observations nested in students were random effects. Because teachers may have differed in how they delivered the curriculum, the intercepts and slopes for the teacher effect were assumed to differ across teachers.

Table 1 reports the robust parameter estimates and standard errors from our analysis. Controlling for the order in which the forms were administered and the order by time interaction, the estimate for time (bolded) was statistically significant. This suggests that students scored significantly higher after the intervention than before, regardless of the order in which they took the forms.

<table>
<thead>
<tr>
<th>Table 1</th>
<th>Robust Fixed Effects Unstandardized Linear Mixed Model Parameter Estimates</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Estimate</td>
</tr>
<tr>
<td>(Intercept)</td>
<td>-1.14*</td>
</tr>
<tr>
<td>Order (A→B or B→A)</td>
<td>0.75*</td>
</tr>
<tr>
<td>Time (pre to post)</td>
<td><strong>1.71</strong></td>
</tr>
<tr>
<td>Order by time interaction</td>
<td>-0.60*</td>
</tr>
</tbody>
</table>

* Estimate is significant at \(p < .05\) or lower.

Field Experiment
In the field experiment, we tested whether students in the treatment group learned significantly more from pretest to posttest than students in the control group. Random assignment occurred at the school level, with three schools assigned to the treatment group and three to the control. Teachers (two per school) were nested within schools \((n = 6)\) and students \((n = 464)\) were nested within teachers.

A sequence of four regression analyses were run to determine the best fitting model:

- **Model I** included fully-crossed fixed effects for time and treatment, and random effects for students (nested in schools), teachers (nested in schools), and schools. Intercepts for random effects were free to vary according to a normal distribution and slopes were fixed. Results revealed that schools
accounted for only 1% of the variance in observed scores, so we eliminated random school effects in Model II.

**Model II** continued to treat intercepts of the random effects for *students* (nested in teachers) and *teachers* as varying according to a normal distribution. The restricted maximum likelihood criterion (REML) was unchanged from Formulation I to Formulation II (REML₁ = REML₂ = 4828), confirming that dropping the *school* factor had no impact on overall fit.

To account for potential differences in how teachers implemented the COR curriculum, **Model III** allowed the slopes for *teacher* (a random factor) to vary. This adjustment improved model fit (REML₃ = 4814 < REML₂ = 4828).

**Model IV** added the following covariates: race, ethnicity, gender, primary language spoken at home, hours spent online per day, and how frequently students checked whether online information was trustworthy. Adding these covariates significantly improved the fit (REML₄ = 4376 < REML₃ = 4814).

Table 2 reports the robust parameter estimates and standard errors for Model IV. The *condition by time interaction* (bolded) was statistically significant. This supports the conclusion that students in the treatment condition were significantly more likely to show improvement from pretest to posttest than students in the control condition.

<table>
<thead>
<tr>
<th></th>
<th>Estimate</th>
<th>Std. Error</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Intercept)</td>
<td>-0.95</td>
<td>0.63</td>
</tr>
<tr>
<td>Condition</td>
<td>0.71</td>
<td>0.37</td>
</tr>
<tr>
<td>Time (pretest to posttest)</td>
<td>0.48</td>
<td>0.31</td>
</tr>
<tr>
<td>Primary home language is English</td>
<td>0.85*</td>
<td>0.33</td>
</tr>
<tr>
<td>Hispanic or Latino</td>
<td>-0.22</td>
<td>0.31</td>
</tr>
<tr>
<td>Race (compared with White)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Asian</td>
<td>-0.65</td>
<td>0.36</td>
</tr>
<tr>
<td>Black or African American</td>
<td>-1.06*</td>
<td>0.40</td>
</tr>
<tr>
<td>Other race</td>
<td>-0.82*</td>
<td>0.39</td>
</tr>
<tr>
<td>More than one race</td>
<td>0.47</td>
<td>0.38</td>
</tr>
<tr>
<td>Female (compared with male)</td>
<td>-0.07</td>
<td>0.18</td>
</tr>
<tr>
<td>Hours online per day</td>
<td>0.06</td>
<td>0.05</td>
</tr>
<tr>
<td>Frequency of checking trustworthiness</td>
<td>0.61*</td>
<td>0.10</td>
</tr>
<tr>
<td><strong>Condition by time interaction</strong></td>
<td><strong>1.66</strong>*</td>
<td><strong>0.44</strong></td>
</tr>
</tbody>
</table>

* Estimate is significant at *p* < .05 or lower.

Figure 5 reflects the estimated change in student scores from pretest to posttest after adjusting for all covariates in the model. Students’ estimated scores in the treatment condition improved significantly more than did the scores for students in the control group.
Figure 5.
Predicted Scores for Students at Pretest and Posttest by Treatment Condition
Notes


8 The original video, viewed more than a million times, is still available on Facebook, https://www.facebook.com/watch/?v=1829274207291885. The same video was first broadcast by the BBC in a September 2016 report about voter fraud in Russia, https://www.bbc.com/news/av/world-europe-37403673/reports-of-voting-fraud-in-russia-poll.
9 Of the more than 3,000 students who completed this same task as part of a recent nationwide sample, only three tracked down this information; Breakstone et al., “Students’ Civic Online Reasoning,” https://purl.stanford.edu/gf151tb4868.


14 Both the pilot study and the field experiment used robust parameter estimates. In both studies, Q-Q plot analyses revealed that outliers might affect estimates of standard errors of the parameters, so we used the R package robustlmm to generate robust parameter estimates resistant to the effects of outliers. See Manuel Koller, “robustlmm: An R Package for Robust Estimation of Linear Mixed-Effects Models,” Journal of Statistical Software 75, no. 6 (2016): 1-24, doi:10.18637/jss.v075.i06.