

Computational Measures of Engagement Across Difference in Online Courses

**Submitted for consideration in the Measuring the Quality of Civic
and Political Engagement RFP in the Spencer New Civics Initiative**

Justin Reich, Dustin Tingley, Brandon Stewart, Andrew Ho

Theoretical Framework and Rationale

Political theorists have long argued that exposure to diverse perspectives is vital to a robust civil society and to the development of individuals within those societies (Kahne, Middaugh, Lee, & Feezell, 2012). Democratic discourse requires engaging with people who hold different perspectives on issues, and differing perspectives are often a function of different life circumstances. One serious threat to contemporary civic education, therefore, is the growing homogeneity of school populations. As residential neighborhoods are increasingly de facto segregated by race and class, students increasingly find themselves in schools with other students similar to themselves (Orfield, Kucsera, & Siegel-Hawley, 2012).

Online education offers one pathway for students to join communities of diverse learners beyond the bounds of geography (Reich, Romer, & Barr, Forthcoming). Consider, for instance, the JusticeX open online course taught by Michael Sandel through the edX platform (Reich et al., 2014). JusticeX examines political philosophy and moral reasoning, and in its first run in 2013 nearly 80,000 students from over 180 countries registered for the course. Students watched lectures, read cases, and participated in online discussion forums where they debated moral dilemmas with peers around the world.

A globally-diverse online learning community, however, does not guarantee that students encounter and consider different perspectives. Internet researchers have posed two competing theories for how people confront differences on the Web (Gardner & Davis, 2013). One theory holds that the Internet is a series of silos where individuals seek out media and communities that conform to their established beliefs (Parser, 2012). Another theory holds that the Internet contains many interest-driven spaces that serve as ideological bridges (Rheingold, 2000), where people attracted to these interest-driven spaces can be diverse in many dimensions. At present, we know little about which of these theories best characterizes open online courses like JusticeX. We know that many diverse students enrolled in JusticeX; we do not know whether those students directly engaged with others holding different beliefs.

In this study, we seek to develop new measures that would allow instructors to evaluate in real-time whether students in online course forums are building siloes or bridges. These measures will build upon previous work examining open classroom climate (Campbell, 2008; Kahne, Crow, & Lee, 2013), but differ in that our measures will draw upon event logs from online platforms rather than survey self-reports. We will use the extremely granular data collected by open online courses to closely examine two specific dimensions of engagement across difference: network shape and thematic convergence.

First, we will use social network analysis to describe the shape of interactions among people with diverse political beliefs in HarvardX courses. From analysis of the connections between posts and replies in forum threads, we will

evaluate the degree to which students in forums respond directly to students with differing opinions versus forming separate conversational siloes containing only like-minded students.

Second, we will use text analysis to evaluate whether the use of language in online forums converges or diverges among students with different political beliefs. Even if students with different beliefs form integrated networks, it may still be possible for students to use siloed language and to talk past one another. For instance, in a conversation about guns in American society, conservatives might address Second Amendment rights whereas liberals might discuss issues of public health. We hypothesize that in high-quality discussions where students directly address each other's ideas, computational text analysis will reveal more modest differences in the language used by people with opposing political beliefs, and we hypothesize that in lower-quality discussions, these gaps in language use will be wider.

From these two analyses, we propose a two-by-two matrix to summarize possible dimensions of engagement across difference in online discussion forums, shown in Figure 1. The bottom left quadrant includes forums where people with different political beliefs separate into different network clusters and use different language; these are the echo chambers of Internet discourse. The top left quadrant includes integrated threads where partisans use different language; these are spaces where students with different beliefs talk past one another. In the bottom right quadrant, students with different political beliefs discuss topics using a shared language, but they divide themselves into conversational siloes with like-minded

others. In the top right quadrant is our hypothesized ideal condition of deliberative discourse, where people with diverse beliefs engage in discussion using a common vocabulary.

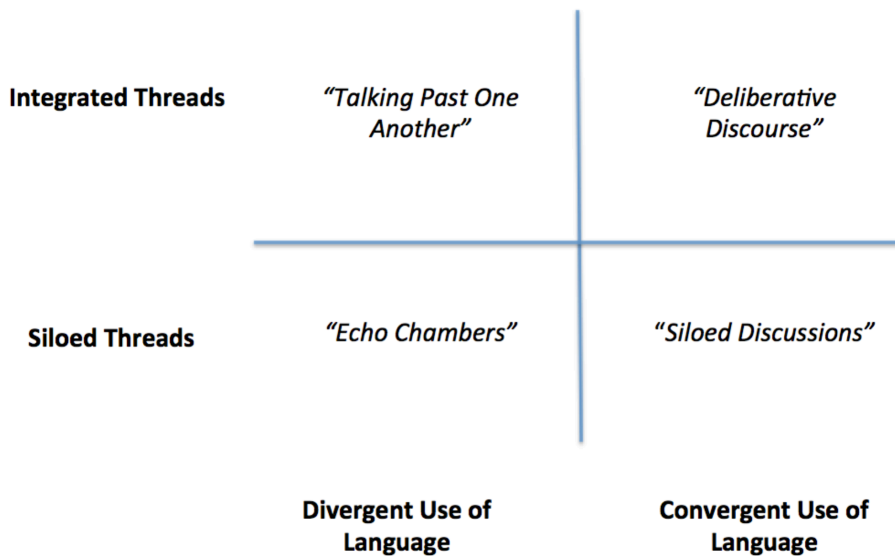


Figure 1: Two-by-two schematic of dimensions of engagement in online discussion forums based on whether thread participants include students with diverse political beliefs (integrated threads) or only those with like-minded beliefs (siloed threads) and whether language use in the thread differs substantially by students' political beliefs (divergent use of language) or if a shared language emerges among students with different beliefs. (convergent use of language).

We hypothesize that these two computational measures of discursive practices will correlate with established measures of open classroom climate as well as instructor perceptions of the quality of online discussions. We predict that in courses with high quality discussions, event log analysis will show that students in diverse networks communicate directly with each other using shared language, surveys will show that students report open classroom climates, and instructor

analyses will report positive perceptions of discussion quality. If our computational measures prove to be useful indicators of high quality discussion, then our research will point the way towards the development of instructor dashboards that can provide real-time indicators of the civic health of online discussions.

This research opportunity comes at a pivotal moment in the short history of massive open online courses (MOOCs) and the much longer history of online education. Although online courses have been available to students for decades, the scale of MOOC enrollment and their adoption by elite universities has dramatically raised both public and institutional attention to online learning. As occurs in all sectors of education, there is constant tension between the many possible purposes of MOOCs: as online job and workforce skill training, as a stimulating leisure activity for lifelong learners, as an entry point into advanced education for the underserved, or as a modern Agora where citizens from a global commonweal can learn and share together. Especially given the vast venture and philanthropic capital invested in MOOC platforms, there is tremendous pressure among MOOC providers to “pivot” (in the parlance of Silicon Valley) towards workforce training courses with a high likelihood of near-term financial return. A major research investment in the civic purposes of open online courses, and the development of new measures for the civic health of MOOCs, will support the advocacy of those who work to ensure that the civic mission of schools remains vital as higher education evolves in the digital age.

Methods

This research will be conducted using discussion forum data from several HarvardX courses where students engage in politically controversial material. We have a commitment from Thomas Patterson of the Harvard Kennedy School of Government to allow us to conduct research in his September 2015 course, Introduction to American Government. The 2015-2016 and 2016-2017 HarvardX slate of courses are still being developed, but other potential course partners include Michael Sandel's Justice class, David Laibson's Behavioral Economics, Paul Peterson's course on Education Policy ("Saving Schools"), and Meira Levinson's Foundations of Urban Education. We refer to the courses, like Intro to American Government, that we will target for investigation of our new measures as "target courses." In conducting these analyses, we will leverage the extraordinarily rich event logs collected by online learning platforms.

Our research will proceed in four steps. First, we will gather baseline data about students' political beliefs and course participation. Second, we will use social network analysis to characterize the degree to which people with different political beliefs directly interact in online course forums. Third, we will use text analysis to evaluate the thematic convergence of forum posts authored by students with differing beliefs. Finally, we will develop validity arguments for our new measures by correlating them with established measures and testing them against experimental interventions.

Baseline Course Participation Data

The starting point for our research is to better understand the political beliefs of HarvardX students. HarvardX currently deploys a common pre-course survey in all courses (currently with over 400,000 responses), soliciting information about student intentions, motivations, preparedness, and backgrounds. For our target courses, we will add scales from the General Social Survey, the World Values Survey, and American National Election Studies survey (Brace, Arceneaux, Johnson, & Ulbig, 2004) to measure the political beliefs of students. We will also include a smaller subset of these items in the surveys for all HarvardX courses in all subjects, so we can compare how student political ideologies in our target courses compare to the overall HarvardX student population.

We will then analyze how student political beliefs correlate with basic features of online course participation, such as initial enrollment, course participation (as measured by total clickstream events), forum participation, persistence, and performance (as measured by course grades) (Ho et al., 2014). From these measures, we can develop a baseline sense of how political beliefs shape patterns of student behavior.

These data will also allow us to better understand how our discussion measures may operate differently in courses with different initial compositions of students. We can anticipate that certain courses may be more attractive to people with certain political beliefs; for instance both Meira Levinson and Paul Peterson have proposed courses on issues of education policy, but we might expect a more conservative audience in Peterson's class and a more liberal one in Levinson's

course given the political leanings of their scholarly work and advocacy. (We anticipate a fairly balanced population in our initial target course: Thomas Patterson's Introduction to American Government.) By conducting this research in multiple courses over the two year period of the grant, we will be able to examine how our methods and measures are influenced by the initial composition of a student community.

Network Shape using Social Network Analysis

Building on these initial participation measures, we can use well-developed techniques from social network analysis to characterize the shape of participation networks in online discussions (Knoke & Yang, 2008). Online discussion forums naturally lend themselves to network analysis, where, at the simplest level, posts are nodes and direct replies in discussion threads form edges. Structural measures of the network such as *eigenvector centrality* (the connectedness of an individual node) and *network density* (the connectedness of the network) can provide a summary of interactions between students. These measures can be complemented by network visualizations that allow instructors and researchers to explore the full social graph created within an online forum.

In many kinds of social networks, people tend to associate with others who share common attributes, a property called assortative mixing or homophily (Newman, 2002). This observation of similarity has been the basis for several lines of research demonstrating the role of network structure in shaping behavior, health, and politics (Christakis & Fowler, 2009). We can use latent community structure (Goldenberg, Zheng, Fienberg, & Airolidi, 2010)—mixed-membership models of

network structure that model each node as belonging to a mixture of latent communities—to model homophily along multiple ideological dimensions. From these kinds of statistical analyses, we will develop measures of *network shape* that characterize the heterogeneity of political leanings within communities of participation. We hypothesize that in high-quality discussions, assortative mixing will be weak, and students will engage with diverse peers. Social network analysis will allow us to quantify and visualize the network shape in online forums to better understand student inclinations towards integrating and siloed behavior.

Thematic Convergence using Structural Topic Models

Our special methodological contribution will be to use advances in machine learning and text analysis to evaluate the degree to which online discussions converge around shared topics or diverge into distinct, non-overlapping discussions. Social network analysis can characterize the connections between people, but these analyses are blind to the content of student writing. To examine discussion forum content, we will use an emerging form of unsupervised machine learning called the Structural Topic Model (Roberts et al., 2014)¹. Topic models identify latent themes within a set of texts by leveraging patterns of syntactic variation in documents, such as frequencies of co-occurring words. *Structural Topic Models* uncover content themes found across many documents—such as forum posts—and then characterize how the distribution of themes differs by substantively interesting covariates, such as a student’s political beliefs.

¹ This article “Structural Topic Models for Open-Ended Survey Responses” just won the 2014 [Gosnell Prize for Excellence in Political Methodology](#).

As an example, our research group used the Structural Topic Model to examine threaded discussion forums in ChinaX, a course about ancient Chinese history from HarvardX (Reich, Tingley, Leder-Luis, Roberts, & Stewart, 2014). The course included tens of thousands of registered students and hundreds of thousands of posts across the forums, far too many for an instructor to read, motivating the need for computational analysis. The methodological approach both discovers the themes (also called topics) in the corpus as well as characterizes how the prevalence of topical content varies with attributes of the document or author. We examined one thread where students discuss their opinions on Chinese intellectuals from Confucius to the Hundred Schools Period. Among these 1,700 posts, the topic model uncovered several themes related to specific thinkers such as Confucius, Lao Tzu, and Han Fei. It also uncovered one common theme, which we denominated as “Vague Language,” where students expressed a general appreciation for Chinese philosophy without actually referencing particular philosophers. For each of these topics, we estimated the proportion of each document representative of each topic, and the overall proportion of the entire corpus representative of each topic.

We then examined how themes from this discussion varied by whether or not the post received an “up-vote” from a peer in the forums. We found that up-voted posts were more likely to reference Confucius and Lao Tzu, whereas posts without an up-vote were more likely to reference Han Fei and Legalism, giving some insight into students’ philosophical preferences. We also found that posts characterized by “Vague Language” were also less likely to receive an upvote. This example, detailed in the paper cited above, demonstrates how the Structural Topic

Model can uncover themes in large collection of texts, provide estimates of the distribution of those themes across the entire corpus, and then examine how the distribution of those themes varies by substantively interesting covariates.

For this project, we would conduct similar analyses, but rather examining how topics vary by a property of each document (number of up-votes), we will use a property of the document's author: his or her political beliefs. Using the Structural Topic Model we will uncover the range of themes discussed by students in a single thread, collection of related threads, or across the entire discussion forum for a course. These themes represent the language used by students in response to prompts in the forum. We will then evaluate how the distribution of themes differs by political beliefs. We hypothesize that sharp divergence in topic models along ideological lines will be indicative of forums where students talk past one another. If political belief is a strong predictor of the language used by students, then students may not be actively engaging with ideas that they disagree with. Convergence in topic models across students with different beliefs will be indicative of forums where students talk to each other. If the distribution of topics across a corpus is similar for students with different political beliefs, then we predict that will serve as evidence of discussion spaces where students are sharing a common conversation rather than forming ideological echo chambers.

Building Validity Arguments for High Quality Engagement Across Difference

We hope to use our new computational measures to draw inferences about the quality of engagement across difference in online discussions. We propose three steps to build the validity argument for these inferences (Kane 2013).

First, we will evaluate whether our measures correlate with each other in expected ways. Descriptive statistics can tell us whether people with certain political beliefs contribute more than others, social network analysis can tell us whether people with different beliefs reply to each other's posts, and the Structural Topic Model can examine whether people with different political beliefs are writing about similar or divergent themes. We hypothesize that in settings supportive of engagement across difference, students with different beliefs will participate in equitable rates, they will form inclusive and bridging networks, and students with different beliefs will write about topics using a shared language. By contrast, in courses with weak norms around open engagement, we expect inequitable participation, homophilic networks, and students talking past one another.

Second, we will explore two possible sources of convergent validity. First, we expect that our computational measures will correlate with established measures of open classroom climate (Campbell, 2008). We will conduct post-course surveys of students that include typical measures of course satisfaction and engagement, along with the political belief scales from the pre-survey and open climate measures. In addition, we will allocate funds to pay course developers, faculty, or outside experts (graduate students working with course faculty) to evaluate the quality of online discussions, in particular surfacing their appraisal of whether forums included

substantive discussion among people with diverse opinions and perspectives. We predict that our computational measures will correlate with these expert evaluations.

Finally, online courses are also excellent environments for experimental approaches to validity testing, since the edX platform allows us to easily assign students at random to different instructional conditions. We are currently conducting a small study in Michael Sandel's JusticeX course where we have piloted three experimental interventions to improve the *quantity* of student participation in discussion forums: email reminders, summaries from the previous year's discussions, and a self-test question asking students to self-report if they have completed the recommended weekly participation activities. With new measures of engagement across differences from this study, we will be able to examine how these and other interventions affect the *quality* of discussion. In particular, we plan to test a variety of instructional prompts—embedded in instructors forum posts, emails to students, replies to specific comments, and in the overall framing of particular discussion threads—that encourage students to respectfully engage with ideas that they disagree with. If these kinds of instructional prompts encourage more equitable participation, bridging conversations in networks, and the use of shared language, then that will provide further evidence that our measures capture important dimensions of the quality of online discussion.

One of the significant advantages of our research design is that it builds on the existing educational infrastructure of HarvardX. None of our resources need to go towards developing new curriculum, finding practitioner partners, incentivizing

participants, traveling to study sites, or other costs of typical educational research studies. Nearly all of our funds will go towards supporting our core research team, funding students and research assistants to aid in the analytic work, and disseminating our work.

Attention to Youth

MOOC participants are an extraordinarily diverse group. In the first year of HarvardX and MITx classes, over 840,000 students enrolled in 17 courses, including nearly 80,000 registrants in JusticeX (Ho et al., 2014). The median age of all HarvardX students is 28, so just under half of students are youth as defined in the RFP. We propose conducting all of the analyses proposed above with both the entire population of registrants in HarvardX courses and the subset of youth age 15-25. It will be useful and nearly costless to see how our measures of engagement across difference in forums vary by the age of our students.

Anticipated Products

The HarvardX Research Group is an active participant in multiple scholarly communities, and we are committed to sharing our findings, tools, and data in ways that advance research and practice. We release much of our work quickly and openly through the [HarvardX Working Papers](#), as well as in scholarly papers that can benefit from the process of peer review. We are committed to sharing well-documented code in open repositories that others can use, modify and improve. Our group has a full time research programmer, tasked with making sure that our

research code is made accessible. Analyses using the Structural Topic Model can be conducted with an open source **R** package, [stm](#), which we have designed to use simple syntax familiar to novice statistical programmers.

We expect that specific products for this grant would include the following:

- Detailed documentation of the survey instruments used in the HarvardX pre-course survey to measure political beliefs, including Qualtrics .qsf files for other researchers using Qualtrics.
- Software code, suitable for immediate use in other edX instances and extensible to other platforms for:
 - conducting initial analyses of participation metrics based on political beliefs
 - organizing and cleaning edX forum data for subsequent analysis
 - conducting social network analysis to produce measures of assortative mixing and visualizations of network shape
 - Structural Topic Model analyses of topical convergence across different political beliefs
- Working papers with initial analyses from our target courses
- Journal publications presenting methodological justifications of our measures illustrated by case studies from our target courses
- Conference presentations, with annual presentation in each year at both an online learning focused conference (such as the ACM Learning@Scale conference) and a civic education focused conference (such as CUFA or AERA).

Theory of Use

Enrollments in online learning environments—ranging from virtual schools in K-12 settings to credit-bearing courses in higher education to lifelong learning experiences such as MOOCs—continue to rise (Allen & Seaman, 2014). The promise of online learning environments is that they can be instrumented to provide teachers and students with real time feedback about student performance, so that instructional changes can be quickly implemented during courses to improve

student learning. Our hope is that the new measures developed within this study will be immediately applicable in other edX settings, and then easily generalized to other online learning platforms as well.

Our research will be primarily conducted using the open source statistical software **R**, and open source packages for **R** such as the **stm** package for conducting Structural Topic Model analysis. We will make all statistical code from our analyses publicly available. Since our research will be conducted on courses using the edX platform, our methods will be immediately extensible to other courses hosted through edX, including organizations using their own instances of the edX open source code, such as [Stanford's OpenEdX](#) or [Edraak](#) in the Kingdom of Jordan. Conducting analyses on discussion forums from other platforms, such as Coursera or Canvas, will require adapting our software code for other data structures, though not fundamentally different use cases.

Our longer-term hope is that new measures could be incorporated directly into online platforms in the form of instructor dashboards to allow for real-time analysis and pedagogical adjustment. Our analytic workflow will involve downloading data from the edX repository for our courses, and then conducting analyses, but our methods and measures could be embedded directly in the platform. For instance, the edX platform includes an instructor dashboard that offers course faculty access to data about student demographics, activity, and performance on quizzes and problems. If instructors in courses in the humanities, social sciences, and professions find our measures and analysis useful, then it would be possible for edX platform developers to extend our methods to include real-time

displays of social graphs in discussion forums, summary statistics of assortative mixing, themes from topic model analysis of forums, and the distribution of themes across substantively interesting subgroups.

We hope that the wider use of these measures of the quality of engagement across difference in online discussion forums will contribute to three important shifts or trends in online education. First, most of the early lines of research in open, online courses have focused on engagement rather than learning, on the quantity of interactions rather than the quality of learning. New measures can shift the priorities of instructors and course developers, to focus instructional improvement on supporting high quality discussion. As mentioned earlier, in the current run of JusticeX, course developers are experimenting with instructional design interventions to increase the quantity of participation. With new measures, these kinds of instructional design experiments could focus on improving the quality of participation rather than the quantity. Second, offering course faculty new measures of civic health and civic quality in their course can help individual instructors envision a wider range of educational purposes for their courses, placing civic outcomes alongside more narrowly academic outcomes. Finally, we hope that new measures of engagement across difference that are extensible to a wide variety of online learning settings can inform a public discussion about online learning that is too often narrowly focused on workforce training. New civic measures, data, and findings can remind stakeholders in online learning that even as higher education is transformed by new technologies, preparing young people for their responsibilities in civil society should be a bedrock purpose of the enterprise.

Staffing Plan

Our proposal requires expertise in civic education, online learning, statistics and machine learning, and psychometrics; and our research team includes four investigators with complementary experience in these domains.

The lead investigator for the project, Justin Reich, is an education technologist and learning scientist who studies how technology-mediated learning environments can empower students. He has extensive experience studying large-scale learning environments, as the project lead for a four-year study of wikis and deeper learning funded by the Hewlett Foundation and in his current role as the senior research fellow for HarvardX. Dr. Reich was also a high school history and middle school civics teacher, and he has taught pre-service history teachers at Harvard and through the Boston Teacher Residency. Dr. Reich will coordinate all aspects of the project, including developing partnerships with course teams, deploying pre-course surveys to measure political ideology, securing data for analysis, identifying research assistants for forum coding and computational analysis, and developing publications and other products to disseminate this research.

Dustin Tingley is a professor in Harvard's Government Department who is an expert in experimental methods and uses emerging techniques from computer science to study political discourses. He will be the Principal Investigator of record for the study, and supervise the text analysis using Structural Topic Models.

Brandon Stewart is a Ph.D. candidate in Harvard's Government Department, finishing in early 2015, with expertise in automated text analysis and applied statistics. Mr. Stewart will support the text analysis research and lead the social network analysis component of the research.

Andrew Ho is a psychometrician interested in the policy consequences of assessment design, a professor at the Harvard Graduate School of Education, and the co-chair of the HarvardX Research Committee. Dr. Ho will play a central role in developing key summary statistics from our computational measures, and shaping how we communicate interpretations of our new measures to diverse audiences. Because Dr. Ho has submitted a separate grant from the Spencer Foundation, we have not included him in the budget for this work nor listed him as an investigator. Per his attached letter of support, he is committed to actively serving as a central member of our research team and an advocate through the HarvardX Research Committee.

We will be supported by two HarvardX research staff. Glenn Lopez is the HarvardX Research Programmer, who will assist in accessing data from the edX platform, and cleaning and organizing those data for research. Rebecca Petersen, the HarvardX Senior Research Manager, will help coordinate efforts with the course teams.

The HarvardX Research Group has developed a thriving community of interested faculty and graduate students through regular meetings and public colloquia. We expect to be able to recruit several outstanding graduate students from computer science, government, or education to contribute to this effort.

References

- Allen, I. E., & Seaman, J. (2014). *Grade change: Tracking online education in the united states*. Working Paper. Babson Survey Research Group.
- Brace, P., Arceneaux, K., Johnson, M., & Ulbig, S. G. (2004). Does state political ideology change over time? *Political Research Quarterly*, *57*(4), 529-540.
doi:10.1177/106591290405700402
- Campbell, D. E. (2008). Voice in the classroom: How an open classroom climate fosters political engagement among adolescents. *Political Behavior*, *30*(4), 437-454.
- Christakis, N. A., & Fowler, J. H. (2009). *Connected: The surprising power of our social networks and how they shape our lives* New York: Hachette.
- Gardner, H., & Davis, K. (2013). *The app generation: How today's youth navigate identity, intimacy, and imagination in a digital world*. New Haven, CT: Yale University Press.
- Goldenberg, A., Zheng, A. X., Fienberg, S. E., & Airolidi, E. M. (2010). A survey of statistical network models. *Foundations and Trends in Machine Learning*, *2*(2), 129-233.
- Ho, A. D., Reich, J., Nesterko, S. O., Seaton, D. T., Mullaney, T., Waldo, J., & Chuang, I. (2014). *HarvardX and MITx: The first year of open online courses, Fall 2012-Summer 2013*. (HarvardX & MITx Working Paper No. 1).

- Kahne, J., Crow, D., & Lee, N. (2013). Different pedagogy, different politics: High school learning opportunities and youth political engagement. *Political Psychology*, 34(3), 419-441. doi:10.1111/j.1467-9221.2012.00936.x
- Kahne, J., Middaugh, E., Lee, N., & Feezell, J. T. (2012). Youth online activity and exposure to diverse perspectives. *New Media & Society*, 14(3), 492-512. doi:10.1177/1461444811420271
- Kane, T. (2013) Validating the interpretations and uses of test scores. *Journal of Educational Measurement*, 50(1), 1-73.
- Knoke, D., & Yang, S. (2008). *Social network analysis* Thousand Oaks, CA: Sage.
- Newman, M. E. (2002). Assortative mixing in networks. *Physical Review Letters*, 89(20), 208701.
- Orfield, G., Kucsera, J., & Siegel-Hawley, G. (2012). *E pluribus... separation: Deepening double segregation for more students*. Working Paper. Los Angeles: Civil Rights Project/Proyecto Derechos Humanos.
- Parser, E. (2012). *The filter bubble: How the new personalized Web is changing what we read and how we think*. New York: Penguin.
- Reich, J., Nesterko, S. O., Seaton, D. T., Mullaney, T., Waldo, J., Chuang, I., & Ho, A. D. (2014). *JusticeX: Spring 2013 course report*. (HarvardX Working Paper No. 4). Retrieved from http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2382248

Reich, J., Romer, A., & Barr, D. J. (Forthcoming). Dialogue across difference: A case study of Facing History and Ourselves' Digital Media Innovation Network. In E. Middaugh, & B. Kirshner (Eds.), *Becoming political in a digital age*. Charlotte, NC: Information Age.

Reich, J., Tingley, D., Leder-Luis, J., Roberts, M. E., & Stewart, B. M. (2014). *Computer-assisted reading and discovery for student generated texts in MOOCs*. Working Paper. Retrieved from <http://scholar.harvard.edu/files/dtingley/files/educationwriteup.pdf>

Rheingold, H. (2000). *The virtual community: Homesteading on the electronic frontier*. Cambridge, MA: MIT.

Roberts, M. E., Stewart, B. M., Tingley, D., Lucas, C., Leder-Luis, J., Gadarian, S. K., . . . Rand, D. G. (2014). Structural topic models for Open-Ended survey responses. *American Journal of Political Science*,