Workshop Title: Using School-Level Data from the Stanford Education Data Archive

Primary Contact: Sean F. Reardon (sean.reardon@stanford.edu)

Presenters:

Sean F. Reardon, Stanford University (*sean.reardon@stanford.edu*)
Andrew D. Ho, Harvard University (*andrew_ho@gse.harvard.edu*)
Benjamin R. Shear, University of Colorado, Boulder (*benjamin.shear@colorado.edu*)
Erin M. Fahle, St. John's University (*fahlee@stjohns.edu*)

Proposed length: 4 hours

Session Description:

The Stanford Education Data Archive (SEDA), part of the Educational Opportunity Project at Stanford (https://edopportunity.org), is an initiative to help researchers, policymakers, and educational practitioners understand educational achievement and opportunity. SEDA includes a range of detailed data on educational conditions, contexts, and outcomes in public schools, school districts, counties and metropolitan statistical areas across the United States. It includes multiple measures of academic achievement and achievement gaps, as well as covariate data such as measures of racial and socioeconomic composition, racial and socioeconomic segregation patterns, and other features of the schooling system. SEDA is unique in providing educational achievement data on a nationally comparable scale using state tests and the National Assessment of Educational Progress (NAEP). The SEDA data are publicly available, so that anyone can obtain detailed information about American schools, communities, and student educational opportunity.

In September 2019, SEDA was expanded in a number of important ways. First, school-level achievement data summarizing the average academic performance and learning rates for nearly all U.S. public schools enrolling students in 3rd through 8th grade were included for the first time. Second, additional years of school district and county data, data for additional demographic subgroups, data for metropolitan statistical areas, and more detailed covariate data were added to the existing database. Finally, a new interactive website (https://edopportunity.org) was launched to facilitate exploration and use of the data by a general audience.

Our goals for the use of this data are twofold. First, we hope that this data will help researchers generate new evidence about what policies and contexts are most effective at increasing educational opportunity, and that such evidence will inform educational policy and practices. Prior research using the SEDA data has been presented in media such as The New York Times, as well as peer review outlets such as *AERA Open*, *Educational Researcher*, and *AERJ*, among others.[1] Second, we hope that, through making information on educational opportunity more widely accessible, educators and policymakers will be able to access, identify, and interpret relevant data and use that data to inform their practices in schools and districts.

Previous versions of this course based on the district-level SEDA data have been offered at the AERA Annual Meeting in 2017, NCME Annual Meeting in 2018, and SREE Annual Meetings in 2017 and 2018. The courses were well-attended and useful to participants. We held a training

---

[1] See examples in the NYTimes:
https://www.nytimes.com/interactive/2017/12/05/upshot/a-better-way-to-compare-public-schools.html
https://www.nytimes.com/interactive/2018/06/13/upshot/boys-girls-math-reading-tests.html
Scholarly research publications and working papers can be accessed here: https://edopportunity.org/research/
Special Topic in AERA Open: https://journals.sagepub.com/pb-assets/SEDA_ERO.pdf

session using preliminary school-level data at the SREE Annual Meeting in 2019, before the interactive website or complete school data were available.

Proposed Workshop Structure and Schedule

This course will be organized similarly to previous versions of the course, but with a greater focus on the school-level data and new website to support the goals outlined above. The course will include a mix of short presentations and small group activities. The presentations will describe the new website and data, as well as provide an overview the methodology used to construct SEDA. The small group activities will provide attendees opportunities to engage with the data, either using the raw data files or the website. Group discussions will be used to address any issues encountered while working with the data or website. Special emphasis will be placed on interpreting and communicating findings from the school-level data activities.

Tentative Schedule	
Introduction and Description of SEDA Contents • Introductory presentation and overview • Guided hands-on exploration of SEDA interactive website • What can these data tell us about educational opportunity at a large scale?	60 minutes
Break	*10 minutes*
Constructing SEDA • Description of how state proficiency cutscores are linked to the NAEP and scaled into a standardized, nationally-comparable metric • Description of how state test score distributions are recovered from the coarsened proficiency data in ED*Facts* using ordered probit models • Description of how measures of average performance, growth, and trend reported in SEDA are estimated in HLM	60 minutes
Break	*10 minutes*
Structured Small Group Explorations • Use the data files or website to explore a series of questions.	45 minutes
Break	*10 minutes*
Wrap-Up and Discussion • Group discussion on interpretation and data use • Guidelines for use and examples of prior research	45 minutes

Significance:

SEDA presents new information about educational opportunities in the roughly 70,000 public elementary and middle schools in the U.S. It provides insight not only into average test scores of each school, but also into school learning rates that are comparable across place and time. Our goal in this workshop is to help educators, policymakers, and researchers understand *what* these new data can tell us about students' educational opportunities and *how* they can use them in their

own work. We will accomplish this through providing detailed training on the data and website, as well as the appropriate uses and interpretations of the data, to a broad audience of educational researchers and stakeholders.

Target Audience:

This course is intended for a broad audience, including faculty, researchers and graduate students interested in using the data for research, as well as educators, policymakers and practitioners interested in learning more about the use of test score data in SEDA and its implications for their work. For those with expertise in quantitative methods and assessment, we will provide activities using the raw data that require the use of statistical software such as R or Stata. For those who want to use the data for other purposes, we will provide similar activities that can be completed through interacting with the website, and which do not require specialized software or statistical training. There are no prerequisites for attendance. Participants should bring a laptop; those interested in using the raw data should have a statistical program such as R or Stata installed.

We believe SREE is an ideal venue to engage our intended audience in a data-oriented discussion about educational opportunity. We also feel that this session fits well with the SREE Annual Meeting theme of "Practical Significance and Meaningful Effects: Learning and Communicating What Matters," by focusing on an initiative that aims to both use existing administrative test score data in a novel manner and to make the data more accessible to and useful for a wider audience.