In the Pipeline Poster Session for Works in Progress

Building Statistical Thinking with Social Justice Investigations and Social Science Data

Background/Context:
There is an increasing demand in science, technology, engineering, and mathematics (STEM) fields and beyond for people with robust data literacy skills, which require strong understandings of the process and practices of statistical thinking and inquiry. Economic forecasts suggest that schools are neither preparing students adequately nor drawing enough students to the study of statistics and data analysis to meet this demand (Manyika et al., 2011). Statistics educators also argue that K-12 mathematics curricula currently do not help students build the types of multivariable reasoning practices that are critical for analyzing data in a big-data world (Engel, 2016; Engel, Gal, & Ridgway, 2016). In addition, current reports suggest that groups that have been traditionally underrepresented in STEM fields (e.g., women, Blacks, and Latinos) are entering data science fields at disproportionately low rates (Priceonomics, 2017).

Purpose/Objective/Hypotheses/Research Questions:
This project is a three-year exploratory/early stage development effort that is working to build knowledge of promising strategies to support statistical thinking and interest in quantitative data analysis among high school students, particularly underrepresented minorities. The project is creating and studying a set of prototype curriculum modules that will provide high school students with hands-on opportunities to investigate patterns of social and economic inequality and historical changes in U.S. society using person-level microdata from the U.S. Census Bureau.

Through this effort, the project is examining a set of conjectures about the types of social issues, curriculum design approaches, pedagogical strategies, and data visualization tools that will engage traditionally underserved students and build their interest in and understanding of quantitative data analysis practices with large-scale data. A primary conjecture of the project is that many high school students – particularly those from historically marginalized groups– may find statistics and data analysis compelling when they can use authentic population data to examine patterns of social and economic inequality and questions that are related to social justice (Gutstein, 2003; Lesser, 2007; Voss & Rickards, 2016). A summary of the project's theory of change is shown in Figure 1.

The project's conjectures motivate the project's guiding research questions:

RQ1. What is the feasibility of implementing project modules in participating classrooms?
RQ2. In what ways do the modules and their components appear to support students' interests in and learning of statistical concepts and practices?
RQ3. To what extent do students who use project modules show improved understandings of important statistical concepts and greater interest in statistics and quantitative data analysis?

Population/Participants/Subjects/Setting:
The project is collaborating with 10 public high school teachers of non-AP statistics classes in the Boston, MA region to develop and test the new curriculum modules with approximately 200

students during each of two years. Teachers have been recruited from high schools with high proportions of students from traditionally underserved groups, such as Blacks, Latinos, those who are economically disadvantaged, and English learners.

Intervention/Program/Practice:
The project is developing two sets of two-week curriculum modules that are intended to replace existing lesson materials, projects, and application exercises in non-AP statistics classes. Using microdata from the U.S. Census Bureau and online data visualization tools, students analyze quantitative distributions and associations among categorical variables to investigate questions such as: What does income inequality in the U.S. look like, and how has it changed over time? How do patterns of immigration to the U.S. today compare to a century ago?

Research Design:
The project is conducting two years of design-based research (Cobb, Jackson, & Sharpe, 2017; Cobb, Confrey, diSessa, Lehrer, & Schauble, 2003), followed by a year of data analysis, synthesis, and dissemination. Across cycles of iterative design and testing, collaborating teachers implement project modules with their students as part of their regular classroom instruction, and the project team is collecting data on teachers' and students' experiences and responses. The first and second years involve the development and testing of "alpha" and "beta" versions of the two modules, respectively. The project is collecting a mix of qualitative and quantitative data to inform module revisions and to provide evidence (or counter-evidence) for the project's conjectures.

Data Collection and Analysis:
Data collected during the alpha phase of module development and testing were primarily qualitative and came from teacher and student focus groups, teacher individual interviews, classroom observations, teacher implementation logs, and student work samples. *A priori* codes and an open coding process were employed to capture emergent themes (Strauss & Corbin, 1998). Team members reviewed and discussed data by code to identify unit implementation successes, challenges, and suggested implications for unit improvements or refinements. The team is looking at data patterns to support, counter, or shed new light on the original conjectures about how unit design features may be associated with student learning and interest outcomes.

Data collected during the beta phase are the same as for the alpha phase, with the addition of quantitative surveys and assessments to measure pre- vs. post-module growth in students' interests in and knowledge of statistical concepts and data practices. Students' interests in data analysis will be measured using a set of academic interest scales (Linnenbrink-Garcia, et al, 2010). Their statistical knowledge will be measured with items from the Levels of Conceptual Understanding in Statistics (LOCUS) instrument (Jacobbe, Case, Whitaker, & Foti, 2014). Matched-pair dependent-sample *t*-tests will examine whether student scores on each measure are higher after students complete each unit. OLS regression models will examine whether post-test scores on each measure are associated with levels of unit implementation, controlling for pre-test scores. Because students within each test will be clustered within teachers, a design effect will be calculated to adjust for standard errors. Concurrent collection of quantitative and qualitative data will provide the study with a mixed methods, convergent parallel design (Creswell & Plano Clark, 2011). This design will allow the team to compare qualitative and quantitative data results using tables and matrices to highlight areas where results converge or diverge. Areas of convergence will lend strength to project findings, whereas areas of divergence will identify areas for future research and exploration.

References

Cobb, P, Confrey, J, diSessa,, C. Lehrer , R., & Schauble, L. (2003). *Educational Researcher*, *32*(1), 9-13.

Cobb, P., Jackson, K., & Sharpe, C. D. (2017). Conducing design studies to investigate and support mathematics students' and teachers' learning. In J. Cai (Ed.), *Compendium for research in mathematics education*. Reston, VA: National Council of Teachers of Mathematics.

Creswell, J. W., & Plano Clark, V. L. (2011). *Designing and conducting mixed methods research* (2nd ed.). Thousand Oaks, CA: SAGE Publications.

Engel, J. (2016). Open data at the interface of mathematics and civics education: Challenges of the data revolution for the statistics curriculum. *Journal of Mathematics and Statistical Science*, *2016*, 264–273.

Engel, J., Gal, I., & Ridgway, J. (2016). Mathematical literacy and citizen engagement: The role of civic statistics. Presented at the 13th International Conference on Mathematical Education, Hamburg, Germany.

Gutstein, E. (2003). Teaching and learning mathematics for social justice in an urban, Latino school. *Journal for Research in Mathematics Education*, 37–73.

Jacobbe, T., Case, C., Whitaker, D., & Foti, S. (2014). Establishing the validity of the LOCUS assessments through an evidenced-centered design approach. In *Sustainability in statistics education. Proceedings of the Ninth International Conference on Teaching Statistics (ICOTS9)*. Retrieved from https://icots.info/9/proceedings/home.html

Lesser, L. M. (2007). Critical values and transforming data: Teaching statistics with social justice. *Journal of Statistics Education*, *15*(1), 1–21.

Linnenbrink-Garcia, L., Durik, A. M., Conley, A. M., Barron, K. E., Tauer, J. M., Karabenick, S. A., & Harackiewicz, J. M. (2010). Measuring situational interest in academic domains. *Educational and Psychological Measurement*, *70*(4), 647–671. https://doi.org/10.1177/0013164409355699

Manyika, J., Chui, M., Brown, B., Bughin, J., Dobbs, R., Roxburgh, C., & Hung Byers, A. (2011). *Big data: The next frontier for innovation, competition, and productivity* (pp. 1–143). McKinsey Global Institute. Retrieved from http://www.mckinsey.com/business-functions/digital-mckinsey/our-insights/big-data-the-next-frontier-for-innovation

Priceonomics. (2017, September 28). The data science diversity gap. *Forbes*. Retrieved from https://www.forbes.com/sites/priceonomics/2017/09/28/the-data-science-diversity-gap/#3d4d8ee5f58b

Voss, R., & Rickards, T. (2016). Using Social Justice Pedagogies to Improve Student Numeracy in Secondary School Education. *Journal of Education and Practice*, *7*(15), 40–47.

Figure 1. Theory of Change

A. Key components of the SDLC modules

Statistical concepts and practices

1. Four-step data inquiry cycle
2. Variability, associations, and comparisons among categorical and quantitative data variables and distributions
3. Multivariable data and interaction effects

Activities

4. Applied, hands-on experiences with existing data sets
5. Work through full data inquiry cycle
   a. Pose/formulate questions
      (includes learning about the structure and types of questions that can be answered with data)
   b. Gather/assemble data
      (includes learning about how secondary data sets are collected and structured, how variables are defined)
   c. Explore/analyze data
      (emphasis on making sense of data visually and understanding concepts, not calculations)
   d. Make arguments/summarize conclusions
      (includes facilitated opportunities for group discussion, debate, and feedback)
6. Use technology to visualize data, with

Structure and organization

7. Extended, 2-week modules
8. Students complete data inquiry cycle twice in each module

Data sets and topics

10. U.S. Census data sets from iPUMS
11. Social and economic trends among U.S. population groups, with a focus on patterns of inequality and issues of social justice, particularly among marginalized groups
12. Questions that students choose involving social and economic data available from U.S. Census data extracts
13. Supplementary background resources on relevant U.S. social and economic trends, the Social Studies C3 Inquiry Arc, and making strong scientific arguments

B. Intermediate student outcomes

1. Students see statistics less as disconnected theorems and recipes, and more as a body of knowledge and practices that can be applied to meaningful endeavors, integral to the work of professional scientists
2. Students see statistics and data analysis as relevant to their own lives, particularly for those from groups underrepresented in data fields
3. Students see connections between the data inquiry cycle and inquiry cycles in other subjects, and therefore the relevance of statistics to other fields

C. Target student outcomes

1. Improved understanding of the data inquiry cycle and key statistical concepts and practices within it:
   a. Asking questions
   b. Assembling data
   c. Analyzing data
   d. Interpreting results and making arguments
2. Improved understanding of variability, associations, and comparisons among data types
3. Improved awareness of multivariable relationships and their potential interactions

4. Greater interest in statistics and quantitative data analysis, particularly among students from historically marginalized groups/ underrepresented in data fields
   a. Greater interest in statistics assignments
   b. Greater interest in statistics/data analysis courses
   c. Greater interest in statistics and quantitative data analysis as fields