Questions to answer in preparing for February 4 class and DARE #2

**The Holden paper**

1. Focus of the paper
   1. What is/are the author’s central research question(s)? (again, think about the big, underlying question and the specific one the paper is able to answer)
      * Big: How does additional funding for textbooks affect student achievement?
      * Specific: How did the exogenous provisions of an additional $96.90 of instructional material funding affect school-level achievement levels in CA in the mid-2000s?
   2. What is/are his answer(s)?
      * Increased average achievement (math/ELA combined) at the ES level by 0.15-0.20 SD units
2. Data and Sample
   1. What is the source of Holden’s data?
   2. What are the characteristics of this sample?
   3. Describe the structure of the data (what does a row consist of, at what level are the variables measured, etc.?). How is the structure of Holden’s data similar to or different from that for DARE #2?
      * Rows are school-year observations. Outcomes are measured at the school-year level and school characteristics are primarily computed at the same level, except for measures of district expenditure such as pupil-student ratio and teacher experience. These are measured at the district level
      * These are identical to the structure of the data for the DARE
   4. What is the forcing variable in Holden’s analysis? To which variable does this correspond in your DARE?
      * The API score of schools in 2003
      * Corresponds to api\_rank directly, but the version centered at 0 is norm
   5. For Holden’s main outcome, he takes the mean scale score for each school-year in math and reading and subtracts the sample mean and standard deviation for that year. This gives him a standardized score in math and reading. He then constructs the primary outcome variable **average\_score** by taking the weighted-average of the standardized scores in math and ELA. What do you make of this approach? Why is he not concerned that he has abstracted away from the original scale of the assessment and combined substantially different domains?
      * His primary interest is in the effect of receiving additional funding *relative to other schools that did not receive the funding*. This is, in essence, a relative rank measure of overall performance where the distance to other schools depends on the standard deviation of the sample.
   6. Holden creates an index of school characteristics and “evaluate[s] the smoothness of this index” to test for sorting across the discontinuity. What is this index and in what ways are its properties similar to/different from the characteristics you will use to conduct similar tests in DARE #2?
   7. Holden interprets the magnitude of his outcomes using both school-level and student-level standard deviations of test score outcomes. What is the difference and why is this important?
3. Methodology
   1. On what source of exogenous variation in textbook (instructional materials) funding does Holden’s identification strategy rely?
   2. What is the mechanism through which schools are identified for Instructional Materials-Williams Case (IMWC) funding?
   3. What assumptions are embedded in this identification strategy (*use the specific context of this study to describe these assumptions*)?
   4. Which of these assumptions can you test given the data you have for DARE #2? Which of these assumptions can you argue are met from a contextual or logical standpoint? Which assumptions will you need to state are ultimately untestable?
   5. Examine Equation 2. Can you describe what each variable and coefficient represent? Which is the causal parameter of interest? How is Eq. 2 different from the equation in FN #16?
   6. For Holden’s main formal analyses, in what bandwidth around the discontinuity in funding does he estimate his models?[[1]](#footnote-1) What does he find when he changes the bandwidth? If you conduct sensitivity tests such as these, how are they similar/different from what you found?
   7. Are Holden’s estimates Intent-to-Treat or Treatment-on-the-Treated estimates? Why? How important is the distinction in this case?
      * Intent-to-Treat as defines all schools that should have qualified for Williams funding based on their 2003 API as being recipients of the Williams. There was one school that did not receive it, and that school’s test scores are likely endogenous to the reason it did not receive funding
      * However, the distinction is not that important as this affects a single school
4. Results
   1. Describe the magnitude of Holden’s preferred estimates. How big of an effect did increased textbook funding have? How much did it cost? Do the magnitude of the effects seem plausible to you?
      * Between 0.15 and 0.20 SD units for $96.90 per student
      * The magnitude of these effects are quite large
   2. Compare your results from DARE #2 to Holden’s results in Figure 5 and Table 5. How are they similar? Different? From where do these similarities/differences come?
      * This will vary at the individual level
   3. Holden finds effects only in elementary school. What reasons does he offer for finding textbook funding effects exclusively in elementary school? Which of these reasons (if any) are convincing to you?
5. Threats to validity & robustness checks
   1. What is “fiscal substitution” in this context and why does Holden spend considerable time attempting to show that it has not occurred?
   2. Holden notes that, “the Williams settlement did not use this threshold to allocate other types of funding” (p. 104). Why is this important to his identification strategy?
      * API score (the accountability index in CA) determines textbook funding receipt. If it determined other sources of funding, it would be impossible to distinguish the effects of increased textbook funding from other benefits of being close to this accountability score
   3. What functional form does Holden use in his preferred specifications to describe the secular relationship between the forcing variable and the outcome variable?
      1. How does he justify this functional form? Does this argument seem reasonable to you?
      2. How sensitive are his results to using other functional forms?
      3. Which functional forms are you most interested in exploring as robustness checks in the optional task in your DARE?
         * Individual responses will vary

1. He does not describe how he selects this bandwidth other than to reference the authors (Calonico, Cattaneo & Titiunik, 2014) of the Stata package **rdrobust**. For now, it is sufficient to know that their approach selects a bandwidth that minimizes mean-square error in the RD estimates. [↑](#footnote-ref-1)