

EDLD 650: Data Analysis and Replication Exercise (DARE) 4

David D. Liebowitz

Due: March 3, 2024

In this project, you will replicate and extend analysis from Umansky and Dumont (2021). In the dataset `umansky_dumont_ECLSK.dta`, you will find the variables listed in [Table 1](#). Note that—different from the original article—these data include just a single year’s observation (Kindergarten) for each student.

Table 1: Variable definitions

<code>race</code>	student race with five categories
<code>hisp</code>	student is Hispanic/Latinx
<code>ses</code>	continuous measure of student’s socioeconomic status
<code>chrabsk</code>	student chronically absent in kinder
<code>prelas</code>	student’s total PreLAS score
<code>ebrs</code>	student’s total EBRS score
<code>kread</code>	student’s kinder reading test score
<code>kmath</code>	student’s kinder math test score
<code>kexecfunc1</code>	student’s kinder executive functioning score
<code>tchrexp</code>	teacher’s number of years of teaching experience
<code>elprgm</code>	student participates in an English learner program
<code>tlangk</code>	standardized measure of teacher perceptions of student’s kinder language skills
<code>tmathk</code>	standardized measure of teacher perceptions of student’s kinder math skills
<code>female</code>	student is female
<code>rural</code>	student attends school in a rural area

A. Baseline differences (3 points)

For the following tasks, give your best attempt at completing the analysis. If you are unable to conduct the programming or analysis, describe what

you are attempting to do and what your results would mean.

- A1. Present graphical and/or numerical evidence on whether teachers perceive students who are classified as ELs as having weaker language and mathematics skills in this data set. Describe the results of your assessment in 3-4 sentences. Should the evidence you have shared here be interpreted as a plausibly causal estimate of the effect of being classified as an EL student on teachers' perceptions of students' ability? In 1-2 sentences, answer why or why not.
- A2. Are there other ways in which students classified as EL learners are different from other students who live in homes where a language other than English is predominantly spoken, but are *not* classified as ELs? Present a table with quantitative information summarizing this fact and describe how this motivates your analytic strategy in [section B](#). in 3-4 sentences.

B. Replication and Extension (7 points)

For the following tasks, give your best attempt at completing the analysis. If you are unable to conduct the programming or analysis, describe what you are attempting to do and what your results would mean.

- B1. Develop a formal model (an equation) that describes the probability that a student who lives in a home where a language other than English is predominantly spoken will be identified as an EL. Start with a basic model that defines EL-classification as a function of PreLAS score, EBRs score, SES, Rurality, Gender and Ethnicity (Hispanic/Latinx or not). From this starting probability, present a visual describing the region of common support for EL- and non-EL-classified students. Describe the substantive implications to your analytic strategy of this figure in 2-3 sentences.
- B2. Construct a Coarsened Exact Matching (CEM) algorithm similar to Umansky and Dumont (see Class 8 Lecture) that relies on the following matching variables: (1) **prelas**, (2) **ebrs**, (3) **ses**, (4) **rural**, (5) **female** and (6) **hisp**. Variables 1-3 are continuous. You should decide whether you will follow Umansky and Dumont's choices for cutpoints or select other reasonable cutpoint values. Variables 4-6 are dichotomous and you should insist on exact matches for these categories. Write 1-2 paragraphs describing the identification strategy (remember this is different from your estimation strategy), its accompanying assumptions, your matching procedures and the resulting number of excluded sample members.
- B3. Assess the quality of your matches by looking at the region of common

support in your newly matched sample. Assess the quality of your matches by comparing baseline variable values in the treated and non-treated conditions. Provide a summary assessment of the quality of your matches, the extent to which you have accomplished balance, and the impact your matching has had on your sample as it relates to both variance and generalizability. Do you think you should try different matching criteria to achieve a better result? Why or why not (*it is not necessary at this point to actually conduct multiple re-matching procedures, just assess whether they would be valuable*)?

- B4. Using your newly matched sample, estimate the average treatment effect of EL classification on teachers' perceptions of students' math and language ability in your newly matched sample. ***If you decide to do B5, present these results and associated discussion along with the rest of your results in B5. If you do not, answer the rest of the prompt with just the CEM results.*** Present your CEM results and compare them to your results in A1 in a table and an accompanying write-up as you would report these in an academic paper in 1-2 paragraphs.
- B5. (**OPTIONAL**) Conduct a robustness check by estimating the causal effect of EL classification on teacher perceptions of student skills by using a propensity score matching approach (or another approach from the matching family if you choose). Share information on the quality of these matches and any additional assumptions associated with this approach. Present these results alongside your results in B4 in an accompanying table(s) and write-up as you would report these in an academic paper in 2-3 paragraphs.
- B6. Write a discussion paragraph in which you present the substantive conclusions (and limitations) of your results about the effects of EL classification on teacher perception of student skills in Kindergarten.