Online Appendix for:
The Long-Term Spillover Effects of Changes in the Return to Schooling*

Ran Abramitzky
Stanford University
and NBER

Victor Lavy
Hebrew University,
University of Warwick
and NBER

Santiago Pérez
University of
California, Davis
and NBER

October 2020

a. Robustness Checks, High-School Results

First, in the third row of Panel A of Table 1, we show that the results are similar when we estimate a cross-sectional regression using only the post early reforms cohorts. This finding implies that the DID estimates are driven by improvements in the treatment group rather than by a decline in performance of the control group. This panel also shows that the pre-reform outcomes were very close in both groups: none of the pre-treatment differences in outcomes are statistically significant. Consistent with this pattern, Online Appendix Table A.10 shows that the results are similar when not including school fixed effects in our baseline specification.

Second, in Table A.11 in we show that the results are similar when we add a student's family average earnings in 2000-2002 as an additional control in the DID estimation. We prefer a multi-year average because it is more likely to be correlated with the permanent level of family resources. Note that performing this exercise was not possible in Abramitzky and Lavy (2014), since family income cannot be properly measured among families who live in the kibbutz.

Third, in Online Appendix Table A.12 we show that the results are similar when we implement an instrumental variables strategy in which we instrument a class-level treatment indicator with the treatment indicator defined at the grade level. The validity of this instrument rests on the assumption that cohort-to-cohort changes in the exposure to students from reformed kibbutzim is random conditional on school fixed effect that account for any confounding factors. This is a reasonable assumption because within a short period of time it is safe to assume that students from adjacent cohorts in a given school have similar

* We thank Jaime Arellano-Bover, Alvaro Calderón, Arun Chandrasekhar, Raj Chetty, Giacomo De Giorgi, Nathaniel Hendren, Matt Jackson, Magne Mogstad, Karthik Muralidharan, John Pencavel, Emmanuel Saez, Tom Zohar, Gabriel Zucman, seminar participants at Pontificia Universidad Católica de Chile, Hebrew University, Warwick, and Stanford, and participants at the CEPR Labor Economics Conference at LSE 2016, the All California Labor Economics Conference at Stanford 2017, and GRIPS 2018 conference in Tokyo for useful discussions and suggestions. We thank Hadar Avivi, Elior Cohen, and Nadav Kunievsky for excellent research assistance. Lavy acknowledges financial support from the European Research Council through ERC Advance Grant 323439 and from CAGE.

characteristics and face the same school environment, except for the fact that one cohort has more students from reformed kibbutzim due to purely random factors. We note that the reduced form effect of this instrument is exactly the grade level treatment effect that we presented above. Secondly, note that within a school the proportion of students from reformed kibbutzim in a grade is highly correlated with the students from reformed kibbutzim in a class, which forms the first stage regression in this 2SLS set up.

We next explore the sensitivity of our results to the sample restrictions and to different definitions of the treatment and control groups. In Table A.14 we present results from two alternative samples. In the first, we restrict the analysis to schools and grades that have at least three students from either early (1998-2000) or late (2003-2004) reformed kibbutzim, both before (1995-1996) and after the early reforms (1999-2000). In the second sample we require at least 6 students. We jump from 3 to 6 students because there are no schools with 4 or 5 such students. Remarkably, the estimates we obtain from these two smaller samples are very similar to the estimates obtained when the restriction is at least 2 students. For example, the effect on high school completion is 0.018 in the 2+ and 6+ samples. The effect on matriculation certification is 0.088 and 0.079, respectively. These similarities are obtained even though the sample size declines by 18 percent.

We next report a specification in which we keep the same students in the control group but drop from the treatment group all the grades with students from both early and late reformed kibbutzim. That is, we compare grades with early reformers but no late reformers to grades with late reformers but no early reformers, before and after the implementation of the early reforms. We report the results of this exercise, as well as the corresponding balancing and sample size Tables in Online Appendix Tables A.15, A.16. and A.17. The results are similar to those in our main specification both qualitatively and quantitatively. Note, however, that the sample size goes down reflecting the more stringent definition of the treatment group.

Appendix Table A.X shows that our results are also similar when we include students from non- and late-reformed kibbutzim in our sample. Moreover, we find similar results when we restrict the sample to *only* include such students (i.e. we exclude non-kibbutz students).

Finally, we assess the robustness of our results to time and school-specific shocks correlated with the presence of early reformers in the grade. There are a number of reasons why such shocks are unlikely to explain our results. First, our findings (discussed below) that some of the effects are larger for males and for students whose parents are less educated largely rule out grade level factors that affect all students equally, such as improvements in schooling infrastructure, changes in teaching practices or in the composition of teachers.[1]

---

[1] We noted earlier in the paper that the high schools that are attended by students from kibbutzim are regional schools where over 80 percent of the students are from non-collective localities in the area. Kibbutzim do not own these schools and they have no influence the composition of the schools' staff. Some of these schools belong to the regional

Second, since we define treatment at the school-year level, our sample includes schools that have grades in both the treatment and control groups, both in the pre and in the post-reform cohorts. In a robustness check, we exploit this feature of the data and estimate our model with a restricted sample that includes *only* schools that have at least one grade in the control group. To have the largest possible sample for this robustness check, we use in the control group peers of students from any kibbutzim. More explicitly, the treatment group in this robustness check includes peers of students from kibbutzim that reformed early and the control group includes peers of students from kibbutzim that reformed late or never reformed. The estimates from this model are presented in Online Appendix Table A.18. The estimates are remarkably similar to our main high school results presented in Table 1. For example, the effect on the summary index in Table A.18 is 0.118 (se=0.076) and in Table 4 it is 0.153 (se=0.058). These results practically rule out the possibility that a school-specific shock is driving our results.

b. Robustness Checks, Long-Term Results

In Tables A.19 and A.20, we show that, similar to the results on high school outcomes, the results for university and college schooling and labor market outcomes are similar when controlling for average family income in the regressions. In Table A.21, we show that the results also hold when we estimate aggregate treatment effects using a summary index for post-secondary schooling and labor market outcomes**.**

A natural question about the above estimated effect on earnings is whether it captures the permanent long-term effects. First, note that we measure the effect on earnings when individuals already completed their post-secondary schooling. Second, based on a sample of older cohorts, we find that earnings at age 30-35 is a strong predictor of earnings at an older age. Yet, it is important to note that earnings have larger variation over time than other personal outcomes. To get a better indication about the permanency of the effect on earnings, we estimated the effect on the percentile rank of individuals in the respective distribution of their cohort (at the national level). There is no direct evidence that suggests that rank forecast is more stable than earnings or log earnings. However, recent papers in the intergenerational mobility literature provide some indirect evidence that is relevant to this issue. These studies have shown that movements across ranks in the income distribution are uncorrelated with parental income conditional on rank at age 30;

---

or district authority and some belong to private or non-profit organizations. Note also that laying off teachers, even if ineffective, is almost impossible in Israel education system because of the tenure system and the strong influence of the teachers' union.

in contrast, movement in log earnings are correlated with parental income conditional on log income at age 30.[2]

Table A.22 in the Online Appendix presents estimates of the effect of the program on percentile rank of earnings, where the rank is computed separately for each cohort based on their percentile in the national income distribution. The estimates are fully consistent with the estimated effects on earnings that are presented in Table 5. After 12 years from high school graduation, the spillover effects moved treated individuals by about 4 percentile ranks in the national income distribution.

[2] For example, Nybom and Stuhler (2016) show with data from Sweden that the relationship between a child's income rank and their parental income rank stabilizes by around age 30; in contrast, the relationship in log earnings is less stable. Chetty et al (2016) find a similar pattern in the US tax data, reporting that percentile ranks predict well where children of different economic backgrounds will fall in the income distribution later in life. Using instead log earnings leads to inferior predictions because of the growth path expansions at the top of the income distribution.