Does Paid Parental Leave Affect Children's Schooling Outcomes? Replicating Danzer and Lavy (2018)

Claudia Troccoli

31 August

Danzer & Lavy (EJ 2018) I

Paid Parental Leave and Children's Schooling Outcomes

- Research question: Does increased time at home when the child is young affect children's cognitive development?
- Cognitive development measure: PISA scores at age 15
- Quasi-experiment: Reform increasing paid parental leave from 1 to 2 years in Austria
  - parents giving birth after 1st July 1990 are eligible for longer leave
  - compare children born before and after the cutoff
  - use information on cohorts not affected by the reform to subtract "month-of-birth effects"
- Main sample: PISA participants born between May and August 1990 and 1987

Danzer & Lavy (EJ 2018) II

Paid Parental Leave and Children's Schooling Outcomes

DiD Equation:

$$y_i = \alpha + \beta_1 Post\_June_i + \beta_2 bc1990_i + \beta_3 Post\_June_i \times bc1990_i + \text{Birth\_month}_i\theta_m + X_i\mu + \epsilon_i \quad (1)$$

where:

- y_i is the child i's score in mathematics, reading or science
- $Post\_June_i$ is a dummy variable indicating that the child was born between 1st of July and 31st of August, and zero if the child was born between May 1st and June 30th
- $bc1990_i$ is a dummy indicator for the birth cohort 1990
- $Birth\_month_i$ is a set of dummy variables indicating the month of birth
- X_i includes the set of background controls: mother's and father's educational attainment, school location and migration background
- β_3 is the coefficient of interest measuring the intention-to-treat effect of the reform
- Standard errors clustered by school programme, school location and gender

Table 3

RD-DID Estimation Results (Boys and Girls)

	Full sample		High education mothers		Low education mothers	
	(1)	(2)	(3)	(4)	(5)	(6)
Mathematics						
Treatment effect	−0.4	2.0	17.2*	16.1*	−7.1	−5.0
	(6.9)	(6.7)	(9.7)	(9.2)	(9.6)	(8.8)
Post-June	0.2	−2.5	−21.7*	−19.5*	8.3	6.0
	(7.0)	(6.8)	(10.9)	(9.9)	(8.6)	(8.0)
Born May	−2.3	−4.2	−11.7	−14.1	1.0	0.2
	(5.9)	(5.6)	(8.9)	(8.6)	(6.0)	(5.7)
Born July	−3.8	−2.9	8.4	5.1	−10.0	−8.4
	(5.2)	(4.8)	(8.8)	(8.6)	(6.7)	(5.2)
bc1990	5.0	3.9	−6.7	−6.2	8.1	8.3
	(6.4)	(5.8)	(10.4)	(8.9)	(6.8)	(6.4)
Constant	503.8***	418.5***	528.2***	457.1***	494.8***	407.6***
	(12.4)	(19.0)	(14.2)	(23.3)	(12.4)	(20.8)
Reading						
Treatment effect	−7.4	−4.1	22.7*	21.2*	−19.8*	−17.1*
	(8.6)	(8.1)	(12.4)	(11.7)	(10.4)	(9.5)
Post-June	4.6	1.3	−24.6*	−22.4**	15.4*	13.5*
	(7.6)	(7.3)	(12.4)	(11.1)	(7.9)	(7.5)
Born May	6.9	4.7	0.6	−2.7	8.4	8.2
	(4.9)	(4.7)	(10.1)	(8.9)	(5.7)	(5.8)
Born July	−1.8	−0.9	17.0	13.3	−11.3	−9.9
	(5.7)	(5.3)	(10.2)	(10.0)	(7.9)	(6.0)
bc1990	6.5	4.8	−13.5	−12.2	13.0*	12.8*
	(7.0)	(6.4)	(11.8)	(11.2)	(7.5)	(7.1)
Constant	515.9***	408.7***	545.7***	447.5***	505.3***	399.3***
	(13.0)	(22.2)	(15.7)	(27.5)	(13.6)	(24.8)
Science						
Treatment effect	−1.0	2.1	23.7**	23.0**	−11.1	−8.5
	(7.6)	(7.4)	(10.3)	(10.0)	(9.9)	(9.0)
Post-June	0.7	−2.7	−22.0**	−20.7**	9.0	6.4
	(7.0)	(6.8)	(10.6)	(9.7)	(8.6)	(8.0)
Born May	3.9	1.6	0.5	−2.7	4.2	3.4
	(5.9)	(5.7)	(7.8)	(6.9)	(6.9)	(6.8)
Born July	−1.1	−0.1	13.1	9.6	−8.2	−6.6
	(5.3)	(5.0)	(8.7)	(8.7)	(7.5)	(5.8)
bc1990	22.1***	20.5***	5.2	5.4	27.6***	27.2***
	(6.8)	(6.2)	(10.2)	(8.8)	(7.3)	(6.9)
Constant	495.0***	388.4***	520.5***	419.7***	485.9***	380.3***
	(12.7)	(18.1)	(14.6)	(24.0)	(12.9)	(18.7)
Observations	2,840	2,840	943	943	1,897	1,897
Background controls	–	✓	–	✓	–	✓

Notes. All regressions control for gender. The control variables on parental background include dummy variables for mother's and father's educational attainment, school location and migration background. Robust standard errors in parentheses (clustered by school programme, school location and gender). Estimation weighted by individual inverse probability weights provided in the PISA data set. ***p < 0.01, **p < 0.05, *p < 0.1. *Source.* PISA data set (OECD), own calculations.

Danzer and Lavy (2018) Results I

- No overall statistically significant effect (cols 1-2)
- Heterogeneous effects:
  - High-education mothers: Positive and weakly significant effect (cols 3-4)
  - Low-education mothers: Negative but mostly insignificant effect (cols 5-6)

Danzer and Lavy (2018) Results II

Table 4

RD-DID Estimation Results by Gender

	Full sample		High education mothers		Low education mothers	
	(1)	(2)	(3)	(4)	(5)	(6)
Boys						
Mathematics	0.1	0.3	17.9	15.8	−8.4	−9.0
	(7.9)	(8.1)	(12.4)	(12.3)	(13.2)	(11.8)
Reading	−7.3	−6.8	34.0**	33.1**	−27.7*	−26.6**
	(9.8)	(9.7)	(15.2)	(15.0)	(14.5)	(12.9)
Science	−2.1	−1.2	41.0***	40.4***	−23.7	−23.3*
	(9.2)	(9.2)	(10.9)	(11.4)	(15.4)	(13.4)
Observations	1,426	1,426	482	482	944	944
Girls						
Mathematics	−0.5	4.1	18.2	16.0	−5.6	−2.0
	(11.1)	(10.7)	(15.5)	(15.2)	(14.3)	(13.3)
Reading	−7.4	−1.9	13.4	13.9	−12.8	−8.9
	(13.7)	(12.6)	(19.7)	(19.1)	(15.5)	(14.0)
Science	0.6	5.8	7.8	6.3	1.7	5.8
	(12.1)	(11.9)	(16.7)	(15.8)	(14.1)	(13.1)
Observations	1,414	1,414	461	461	953	953
Background controls	–	✓	–	✓	–	✓

Notes. Each cell reports the estimated treatment effect from a separate regression. The upper panel includes only male, the lower panel only female students. All regressions include dummy variable controls for survey year, birth months and for all children born post-June. The control variables on parental background include dummy variables for mother's and father's educational attainment, school location and migration background. Robust standard errors in parentheses (clustered by school programme, school location and gender). Estimation weighted by individual inverse probability weights provided in the PISA data set. ***p < 0.01, **p < 0.05, *p < 0.1. *Source.* PISA data set (OECD), own calculations.

- Boys of High-education mothers: Positive and significant effects (cols 3-4)
- Boys of Low-education mothers: Negative and significant effects (cols 5-6)
- Girls: statistically insignificant effects (cols 1-6)

There are at least two methodological concerns with these estimates

1. The authors do not account for the stochastic nature of PISA performance data (i.e. they use only one of the 5 plausible values)
2. They do not take into account the sampling design (e.g. do not use appropriate weights)

Stochastic nature of PISA scores I

- International Large Scale Assessments: PISA, TIMSS, PIAAC, PIRLS
- Hundreds of questions on each test subject (Math, Science, Reading)
- Students do not answer the full set of questions, but only 1 out of 13 booklets (which might not contain questions on all subjects)
- Random assignment of booklets $\rightarrow$ unanswered questions are Missing Completely At Random
- Multiple imputation is applied to obtain a distribution of potential scores
  - Imputation takes into account information on student's characteristics, their performance in the answered questions, as well as the scores obtained by pupils in the same school

Stochastic nature of PISA scores II

	STIDSTD	PV1MATH	PV2MATH	PV3MATH	PV4MATH	PV5MATH	PV1READ	PV2READ	PV3READ	PV4READ	PV5READ	PV1SCIE	PV2SCIE	PV3SCIE	PV4SCIE	PV5SCIE
1	1	417.308	398.6135	436.0025	407.1818	429.771	450.8965	392.3439	474.9593	466.1363	441.2715	379.3813	315.8351	392.9984	388.4594	363.0409
2	2	360.2898	329.1323	350.9425	304.9852	339.2585	318.7922	332.4277	295.5315	208.1036	317.188	451.5517	413.424	360.7714	307.211	336.2607
3	3	210.6559	209.098	237.9187	270.6341	236.3608	326.106	228.3881	304.6557	324.5171	313.3947	331.177	318.4677	351.1487	340.255	290.3258
4	4	271.413	297.8969	241.8134	296.339	262.0657	256.2291	256.2291	301.9483	318.7922	254.6249	332.0848	329.3614	368.3969	360.2267	281.2478
5	5	562.5019	490.8397	496.2923	507.9763	469.0294	550.7569	497.0168	519.4753	543.538	541.9339	544.5108	460.0851	468.2553	529.0781	500.9362
6	6	384.5926	400.1714	400.1714	435.2236	373.6875	416.5149	439.5541	429.2262	439.5541	409.3648	384.7374	396.5388	348.4253	385.6452	303.9429

Figure: Plausible Values

- Five Plausible Values are drawn from the posterior distribution of scores
- D&L only used the 1st plausible value in the analysis

Stochastic nature of PISA scores III

- Modified version of Rubin's rule for multiple imputation
- The equation of interest is estimated five times, with each PV as outcome variable.
- The averages of the five parameters and the five sampling error estimates are, respectively, the final parameter (β_*) and the final sampling error (σ_*)
- Calculate the imputation error: $\delta_* = \frac{\sum_{pv=1}^{5}(\beta_{pv}-\beta_*)^2}{n_{pv}-1}$
- Calculate the standard error: $se_* = \sqrt{\sigma_*^2 + (1+\frac{1}{PV})\cdot\delta_*^2}$
- If authors consider only one PV and additionally ignore the recommended procedure to adjust standard errors, these are underestimated, which may artificially inflate the statistical significance of the estimates.
- Stata packages available

PISA sampling design

- Within each country and each explicit stratum (schools in the same regions and of the same "type"), schools are chosen randomly, with probability proportional to each school's size
- Each school is assigned two substitute schools based on characteristics correlated with PISA scores (implicit stratification), to limit bias caused by non-response
- 30 students are randomly picked in each selected school
- Individual Student Weights rescale the sample to the size of the population within each country and generalize to the overall population of potential test takers
- Balanced Repeated Replication Weights adjust for uncertainty with regards to sampling by taking into account the two-stage stratification design, whereby schools are selected and students are randomly drawn from each school.

Troccoli (2020) Replication I

- I replicate D&L study, using the same specification and the official PISA dataset publicly available, and following the appropriate procedure
- Three main differences:
  1. My sample includes 20 additional people (2860 vs. 2840) - irrelevant for results
  2. 5 plausible values (D&L use only the first one)
  3. all relevants weights (D&L only account for student weights)
- Theoretically, accounting for (ii) and (iii) should lead to larger (estimated) standard errors
- This point has been made earlier by Jerrim et al. (Economics of Education Review 2017) who use as an example of bad practice another paper published in the Economic Journal by one of the authors of this paper.

Troccoli (2020) Replication Results I

	MATHEMATICS	READING	SCIENCE
	(1)	(2)	(3)
Panel 1, full sample, N=2860			
PV1 (DL, N=2840)	2.00	-4.14	2.10
	(6.70)	(8.09)	(7.41)
PV1	1.85	-3.81	2.01
	(6.62)	(7.99)	(7.31)
PV2	1.82	-5.85	0.52
	(6.83)	(8.11)	(7.66)
PV3	1.26	-4.90	1.26
	(7.73)	(8.66)	(7.82)
PV4	-1.30	-6.36	-0.95
	(7.14)	(7.06)	(6.62)
PV5	-3.08	-9.40	-1.15
	(6.63)	(7.71)	(7.03)
PV1-PV5 (Rubin)	0.11	-6.06	0.34
	(7.13)	(7.36)	(6.80)
PV1-PV5 & BRR	0.11	-6.06	0.34
	(9.30)	(9.23)	(8.91)

Troccoli (2020) Replication Results II

	MATHEMATICS	READING	SCIENCE
	(1)	(2)	(3)
Panel 2, sons of high education mothers, N=484			
PV1 (DL, N=482)	15.83	33.12**	40.40***
	(12.28)	(14.99)	(11.45)
PV1	15.78	33.10**	40.38***
	(12.31)	(15.00)	(11.45)
PV2	15.46	5.67	12.47
	(13.25)	(14.69)	(13.61)
PV3	16.11	10.30	21.71
	(14.35)	(16.03)	(12.96)
PV4	6.35	13.24	19.99
	(13.26)	(16.83)	(15.92)
PV5	14.49	0.57	12.45
	(13.72)	(14.75)	(13.54)
PV1-PV5 (Rubin)	13.64	12.58	21.40
	(18.02)	(22.98)	(21.38)
PV1-PV5 & BRR	13.64	12.58	21.40
	(18.73)	(23.27)	(21.48)

Troccoli (2020) Replication Results III

	MATHEMATICS	READING	SCIENCE
	(1)	(2)	(3)
Panel 3, sons of low education mothers: N=944			
PV1 (DL, N=944)	-9.03	-26.63**	-23.25*
	(11.77)	(12.87)	(13.38)
PV1	-9.03	-26.63**	-23.25*
	(11.77)	(12.87)	(13.38)
PV2	-2.30	-24.73*	-20.49
	(10.48)	(12.32)	(12.99)
PV3	-12.28	-18.72	-18.97
	(12.09)	(13.00)	(13.12)
PV4	-13.77	-15.31	-18.10
	(10.86)	(11.00)	(11.94)
PV5	-16.51*	-21.85	-25.30*
	(9.65)	(13.20)	(13.14)
PV1-PV5 (Rubin)	-10.78	-21.45	-21.22*
	(13.00)	(13.31)	(12.02)
PV1-PV5 & BRR	-10.78	-21.45	-21.22
	(16.41)	(16.45)	(15.61)

Troccoli (2020) Replication Results IV

	MATHEMATICS	READING	SCIENCE
	(1)	(2)	(3)
Panel 4, daughters of high education mothers, N=468			
PV1 (DL, N=461)	16.00	13.91	6.33
	(15.18)	(19.08)	(15.82)
PV1	16.03	14.38	6.16
	(14.90)	(18.76)	(15.60)
PV2	4.68	22.38	18.11
	(14.65)	(18.41)	(16.57)
PV3	6.83	-2.80	-8.28
	(16.39)	(18.96)	(15.53)
PV4	11.88	-4.98	-2.22
	(13.74)	(14.59)	(13.28)
PV5	5.54	-1.00	-1.60
	(18.04)	(13.80)	(16.21)
PV1-PV5 (Rubin)	8.99	5.60	2.43
	(18.30)	(22.04)	(20.36)
PV1-PV5 & BRR	8.99	5.60	2.43
	(23.83)	(25.25)	(23.01)

Troccoli (2020) Replication Results V

	MATHEMATICS	READING	SCIENCE
	(1)	(2)	(3)
Panel 5, daughters of low education mothers, N=964			
PV1 (DL, N=953)	-2.02	-8.91	5.82
	(13.28)	(13.98)	(13.13)
PV1	-2.91	-9.39	5.27
	(13.17)	(13.82)	(12.92)
PV2	-4.18	-8.90	5.37
	(14.45)	(15.67)	(16.68)
PV3	3.40	-1.40	13.46
	(14.83)	(15.63)	(15.84)
PV4	-0.94	-9.72	5.06
	(15.48)	(14.20)	(13.79)
PV5	-4.46	-9.48	12.98
	(13.06)	(14.86)	(13.00)
PV1-PV5 (Rubin)	-1.82	-7.78	8.43
	(11.68)	(11.99)	(11.89)
PV1-PV5 & BRR	-1.82	-7.78	8.43
	(13.60)	(13.96)	(13.05)

Troccoli (2020) Replication Results VI

Notes: The table reports estimates for β_3 of Equation **??**. The outcome variable standardised such that the standard deviation is 100, hence the coefficients can be interpreted as "percent of a standard deviation" (e.g. 2.00 is 2% of a st.dev. or 0.02 st.dev.). Each cell corresponds to a separate regression. The first column reports results using PISA scores in Mathematics, column 2 in Reading and column 3 in Science. Each panel presents information for a different sample of children. The first row of each panel, titled *PV1 (DL)*, presents the results reported by Danzer and Lavy (2018) in Table 3 (p. 101) and 4 (p. 104), where they estimate equation **??** using as outcome variable the first Plausible Value and accounting for individual inverse probability weights. In the second row, titled *PV1*, I report the results that I obtain when I replicate their analysis using the same outcome variable and specification, but with a slightly different sample size. Next, in rows 3-6 I report similar estimates using as outcome variables Plausible Values 2-5. Row 7, titled *PV1-PV5 (Rubin)*, provides the coefficients for the estimation that uses all five Plausible Values and the Student Weights. Finally, row 8, titled *PV1-PV5 & BRR*, reports the results using the procedure recommended by the survey organisers, e.g. using all five Plausible Values and the Student Weights, as well as the eighty Balance Replication Weights. Following DL 2018, all regressions include controls for gender (only in panel 1), month of birth, mother's and father's educational attainment, school location (urban or rural) and migration background (whether the student's home language is not German). Robust standard errors in parentheses (following DL 2018: clustered by school programme (academic or vocational), school location and gender). *** $p<0.01$, ** $p<0.05$, * $p<0.1$

THANK YOU