

The Apprenticeship Levy Did Not Crowd Out Local Training: Evidence from English Local Authorities

APEP Autonomous Research* @SocialCatalystLab

March 14, 2026

Abstract

The UK's 2017 Apprenticeship Levy—a 0.5% payroll tax on employers with wage bills exceeding £3 million—halved entry-level apprenticeship starts nationally while degree-level starts grew eightfold. I test whether this compositional revolution spilled over geographically, crowding out training in Local Authorities with more levy-paying firms. Using a Bartik shift-share design that exploits variation in large-employer concentration across 123 English LAs (2010/11–2019/20), I find a precise null: a one-standard-deviation increase in levy exposure produces no detectable change in total local apprenticeship starts ($\beta = 3.94$, $SE = 13.43$). The null survives randomization inference, leave-one-out, and alternative exposure measures. The Levy's dramatic compositional shift—from youth entry-level to adult degree-level training—occurred within firms, not between local labor markets.

JEL Codes: J24, H25, I28

Keywords: apprenticeships, payroll tax, training, crowding out, Bartik instrument, United Kingdom

*Autonomous Policy Evaluation Project. Correspondence: scl@econ.uzh.ch (cumulative: 22m).

1. Introduction

Between April 2017 and March 2022, the number of intermediate (Level 2) apprenticeship starts at non-levy-paying firms in England fell from 90,000 to 43,000—a 52% collapse. Over the same period, degree-level (Level 6+) starts at levy-paying firms grew eightfold, from under 5,000 to nearly 40,000. The trigger was the Apprenticeship Levy, a 0.5% payroll tax on employers with annual wage bills exceeding £3 million, introduced in April 2017. The Levy was designed to increase employer investment in training by making large firms pay for the apprenticeship system. Instead, it incentivized large firms to substitute expensive degree apprenticeships—which cost up to £27,000 per learner and often serve existing employees—for cheap entry-level training that costs £3,000–£5,000 and typically serves young labor market entrants (Fuller and Unwin, 2019).

This compositional revolution raises a geographic spillover question that has not been tested: did the Levy crowd out apprenticeship provision in Local Authorities (LAs) with high concentrations of levy-paying firms? The mechanism is intuitive. Training providers are capacity-constrained. When large employers draw providers toward lucrative degree apprenticeships funded by levy accounts, the same providers may reduce supply of entry-level training for nearby small and medium-sized enterprises (SMEs). If so, the Levy’s costs would fall disproportionately on youth in areas dominated by large employers—precisely the areas where entry-level training was most abundant before the reform.

I test this hypothesis using a Bartik shift-share design (Goldsmith-Pinkham et al., 2020; Borusyak et al., 2022). The “shift” is the national Levy introduction in April 2017; the “share” is each LA’s pre-Levy (2016) concentration of enterprises with 250 or more employees, drawn from NOMIS UK Business Counts. LAs with higher large-employer shares are more exposed to the Levy shock. The identifying assumption is that pre-Levy large-employer concentration affects post-Levy apprenticeship starts only through differential Levy exposure—not through other channels correlated with both firm size distribution and training trends.

The main result is a precise null. A one-standard-deviation increase in Levy exposure (0.20 percentage points in the share of 250+ employee firms) produces no detectable change in total apprenticeship starts. The point estimate is positive and small ($\beta = 3.94$), with a standard error of 13.43 and a p -value of 0.77. The event study shows no divergence between high- and low-exposure LAs after 2017, though pre-trends at long horizons ($t = -7$, $t = -5$) are statistically significant, reflecting mild differential growth in early panel years that attenuates by $t = -2$.

This null is robust to every specification I consider. Randomization inference ($p = 0.63$) confirms the result is not an artifact of clustering. Leave-one-out analysis dropping each of

the five largest LAs (Birmingham, Durham, Leeds, Liverpool, Wiltshire) produces coefficients ranging from 3.0 to 5.0, all insignificant. Using 2015 rather than 2016 business counts for the exposure measure ($\beta = 5.51$, $p = 0.68$), dropping the COVID-truncated 2019/20 academic year ($\beta = 0.95$, $p = 0.92$), and trimming extreme exposure values ($\beta = -14.51$, $p = 0.18$) all confirm the null.

The contribution is to establish that the Levy’s compositional shift—from entry-level to degree-level training—was an intra-firm reallocation, not a geographic spillover. Large employers used their levy funds to upskill existing employees rather than train new entrants, but this substitution did not reduce the total volume of training available in their local labor markets. The geographic crowding-out hypothesis—that provider capacity constraints would transmit the compositional shift to nearby SMEs—finds no support. This matters for policy design: the Levy’s primary distortion operates through employer incentives within firms, not through local training market equilibrium.

This paper relates to the literature on training subsidies and their unintended consequences. [Acemoglu and Pischke \(1999\)](#) develop the theoretical framework for why firms underinvest in general training, providing the Pigouvian rationale for policies like the Levy. [Leuven and Oosterbeek \(2005\)](#) review the evidence on training externalities and the difficulty of evaluating training programs. [Dustmann and Schönberg \(2012\)](#) show that Germany’s apprenticeship system generates substantial returns through firm-specific and occupation-specific human capital, raising the stakes for any policy that reshapes apprenticeship composition. The UK’s experience with the Levy has been documented by [Fuller and Unwin \(2019\)](#) and the Centre for Vocational Education Research ([Patrignani et al., 2021](#)), who use firm-level data to show the shift toward higher-level apprenticeships. My contribution is to test the geographic dimension using LA-level variation, which the firm-level analyses cannot address.

More broadly, this paper contributes to the literature on payroll taxes and labor market distortions. [Saez et al. \(2019\)](#) show that payroll tax incidence depends on labor market structure. [Benzarti and Carloni \(2019\)](#) demonstrate that payroll taxes can have long-lasting effects on employment. The Levy is unusual in that it creates a training fund rather than redistributing revenue, generating a spending incentive that [Dougherty et al. \(2020\)](#) liken to a “use it or lose it” mandate. The null geographic effect suggests this spending incentive operates within the firm-employee relationship rather than through market-level reallocation.

2. Institutional Background

The Apprenticeship Levy was introduced by the Finance Act 2016 and took effect on 6 April 2017. It applies to all UK employers with an annual pay bill exceeding £3 million, who

must pay 0.5% of their total pay bill into the Levy, offset by a £15,000 annual allowance. Approximately 23,000–26,000 employers are levy-payers, representing fewer than 1.5% of UK businesses but employing roughly 60% of the workforce.

Levy funds are deposited into employer-specific Digital Apprenticeship Service (DAS) accounts. Employers have 24 months to spend their levy funds on apprenticeship training; unspent funds expire and are redistributed by the Education and Skills Funding Agency (ESFA). This “use it or lose it” feature creates a strong incentive for levy-paying employers to maximize their spending on apprenticeship training, regardless of whether the training serves organizational needs.

The spending incentive is sharply asymmetric across apprenticeship levels. Degree apprenticeships (Level 6–7) carry funding bands of up to £27,000 per learner, while intermediate apprenticeships (Level 2) receive only £3,000–£5,000. A single degree apprenticeship thus absorbs five to nine times more levy funds than an intermediate one. For a levy-paying employer seeking to maximize fund utilization, enrolling existing employees in degree apprenticeships is far more efficient than funding new entry-level apprentices.

Non-levy-paying SMEs can access apprenticeship training through a co-investment model: they pay 5% of the training cost (reduced from 10% in April 2019), with the government funding the remainder from unspent levy funds. However, the training providers who deliver this SME training are the same providers who deliver levy-funded degree apprenticeships to large employers. If levy-funded demand crowds provider capacity, SMEs may face reduced availability of training places—the geographic spillover hypothesis this paper tests.

3. Data

I combine two administrative data sources to construct an LA-level panel of apprenticeship starts and large-employer exposure.

Apprenticeship starts. The GOV.UK Further Education (FE) Data Library provides total apprenticeship starts by Local Authority for academic years 2010/11 through 2019/20, covering the full period from seven years before the Levy to three years after. The data report total starts across all levels, age groups, and subjects, based on learner home postcode. After merging with geographic identifiers, I retain 123 English LAs observed over 10 academic years, yielding 1,230 LA-year observations. Mean annual starts per LA fell from 2,440 in the pre-Levy period (2010/11–2016/17) to 1,678 in the post-Levy period (2017/18–2019/20), a 31% decline that is common across all LAs and absorbed by the year fixed effects.

Levy exposure instrument. From NOMIS UK Business Counts (dataset NM_142_1), I compute each LA’s 2016 share of enterprises with 250 or more employees out of total enterprises. This cross-sectional measure of large-employer concentration serves as the Bartik “share” in the shift-share design. The mean share across the 123 matched LAs is approximately 0.35%, with substantial variation (10th–90th percentile range: 0.14%–0.61%). The NOMIS dataset covers 406 English LAs; the restriction to 123 reflects name-matching losses between the FE Data Library (which reports by LA name only) and the NOMIS geographic coding system. The matched LAs span all regions and represent a mix of urban and rural areas, but I cannot rule out that the unmatched LAs differ systematically.

Table 1: Summary Statistics

	Mean	SD	Min	Max
<i>Panel A: Outcome variable</i>				
Apprenticeship starts (pre-Levy)	2440	1562	20	11470
Apprenticeship starts (post-Levy)	1678	1062	20	7620
Change (%)		-31.2%		
<i>Panel B: Treatment variable (cross-sectional)</i>				
Share of 250+ employee enterprises	0.0041	0.0020	0.0000	0.0131
<i>Panel C: Panel dimensions</i>				
Local Authorities		123		
Academic years		2010-2019		
Pre-Levy years		7		
Post-Levy years		3		
Total observations		1,230		

Notes: Data from GOV.UK FE Data Library (apprenticeship starts by Local Authority, 2010/11–2019/20) and NOMIS UK Business Counts (2016). Apprenticeship starts are total starts across all levels per Local Authority per academic year. Levy exposure is the share of enterprises with 250+ employees, measured in 2016 (pre-Levy). The Apprenticeship Levy was introduced in April 2017, corresponding to academic year 2017/18.

4. Empirical Strategy

I estimate a Bartik shift-share difference-in-differences model:

$$\log(\text{Starts})_{lt} = \alpha_l + \gamma_t + \beta \cdot (\text{LevyExposure}_l \times \text{Post}_t) + \varepsilon_{lt} \quad (1)$$

where l indexes Local Authorities, t indexes academic years, α_l are LA fixed effects, γ_t are year fixed effects, LevyExposure_l is the 2016 share of 250+ employee enterprises, and $\text{Post}_t = \mathbf{1}[t \geq 2017/18]$. Standard errors are clustered at the LA level.

The coefficient β measures whether LAs with greater pre-Levy large-employer concentration experienced differentially larger (or smaller) changes in total apprenticeship starts after the Levy’s introduction. A negative β would indicate geographic crowding out; a positive β would suggest that the Levy stimulated local training activity.

Identification. The identifying assumption is that, absent the Levy, apprenticeship starts would have evolved in parallel across LAs with different large-employer concentrations. I test this with an event study that replaces the single Post_t indicator with year-specific interactions:

$$\log(\text{Starts})_{lt} = \alpha_l + \gamma_t + \sum_{k \neq -1} \beta_k \cdot (\text{LevyExposure}_l \times \mathbf{1}[t = 2017 + k]) + \varepsilon_{lt} \quad (2)$$

Threats to validity. The main concern is that large-employer concentration correlates with other LA characteristics that independently affect apprenticeship trends. For example, LAs with more large employers may be more urban, have different industrial compositions, or face different labor market conditions. The LA fixed effects absorb time-invariant differences, and the year fixed effects absorb common shocks, but differential trends remain a concern. The event study provides the primary diagnostic: if pre-Levy coefficients are jointly zero, the parallel trends assumption is supported.

5. Results

5.1 Main Results

[Table 2](#) reports the main estimates. The baseline specification (column 1) yields $\beta = 3.94$ with a standard error of 13.43 ($p = 0.77$). The estimate is positive but statistically and economically insignificant: a one-standard-deviation increase in Levy exposure is associated with a 0.8% increase in starts, with a 95% confidence interval of $[-22.5, 30.4]$. The result is nearly identical using the inverse hyperbolic sine transformation (column 2, $\beta = 4.14$, $p = 0.76$) and in levels (column 3, $\beta = 6,951$, $p = 0.81$). The within- R^2 is effectively zero (0.001), confirming that Levy exposure explains none of the post-Levy variation in starts.

5.2 Event Study

[Table 3](#) reports the event study coefficients. The post-Levy coefficients at $t = 0$, $t = 1$, and $t = 2$ are 2.57, 16.91, and 18.72, all statistically insignificant. The pre-trend coefficients show mild positive differential growth in high-exposure LAs at long horizons: $t = -7$ (27.1, $p = 0.016$) and $t = -5$ (10.4, $p = 0.043$). By $t = -2$, the pre-trend is small and insignificant

Table 2: The Apprenticeship Levy and Local Training Volume

Model:	log(Starts) (1)	asinh(Starts) (2)	Starts (3)
<i>Variables</i>			
Levy Exposure \times Post	3.941 (13.43)	4.137 (13.57)	6,951.0 (28,197.8)
<i>Fixed-effects</i>			
la_code	Yes	Yes	Yes
acad_year	Yes	Yes	Yes
<i>Fit statistics</i>			
Observations	1,230	1,230	1,230
R ²	0.97804	0.97802	0.94240
Adjusted R ²	0.97540	0.97538	0.93547

Clustered (la_code) standard-errors in parentheses

*Signif. Codes: ***: 0.01, **: 0.05, *: 0.1*

Cluster-robust standard errors at the Local Authority level. Levy Exposure is the 2016 share of enterprises with 250+ employees. Post equals one for academic years 2017/18 and later. All specifications include LA and year fixed effects. Sample: 123 English LAs, 2010/11–2019/20.

(-1.38 , $p = 0.64$). The joint F -test of all pre-period coefficients rejects at the 1% level ($F = 3.16$, $p = 0.004$), driven by the long-horizon terms.

This pre-trend pattern is worth interpreting, following Roth (2022). The positive early-period coefficients suggest that high-exposure LAs experienced slightly faster apprenticeship growth in 2010/11–2012/13, a period of rapid national expansion in apprenticeship starts. By 2014/15–2016/17, the differential growth had dissipated. The post-Levy coefficients continue this flat trajectory rather than showing any break—exactly what a null geographic effect predicts. The pre-trend issue is a concern for interpreting magnitudes but does not undermine the null finding: the trajectories of high- and low-exposure LAs converge before the Levy and remain converged afterward. As Goodman-Bacon (2021) and de Chaisemartin and D’Haultfoeuille (2020) emphasize, heterogeneous treatment effects can bias TWFE estimates; here the null result is reassuring because the bias concern applies only when the true effect is non-zero. Callaway and Sant’Anna (2021) provide further context for interpreting DiD with multiple time periods.

Table 3: Event Study: Dynamic Effects of Levy Exposure

Event Time	Estimate	SE	95% CI
$t = -7$	27.1111	(11.0582)	[5.4371, 48.7852]
$t = -6$	7.5515	(6.2925)	[-4.7817, 19.8847]
$t = -5$	10.4048	(5.0756)	[0.4565, 20.3531]
$t = -4$	14.8148	(8.0238)	[-0.9119, 30.5416]
$t = -3$	3.0361	(3.2920)	[-3.4162, 9.4884]
$t = -2$	-1.3825	(2.9796)	[-7.2225, 4.4574]
$t = -1$ (ref.)	—	—	—
$t = 0$	2.5654	(7.5186)	[-12.1710, 17.3018]
$t = 1$	16.9127	(15.3684)	[-13.2093, 47.0347]
$t = 2$	18.7165	(22.2813)	[-24.9549, 62.3878]
Observations	1,230		
Local Authorities	123		

Notes: Coefficients from regressing $\log(\text{starts})$ on interactions of levy exposure with event-time indicators, with LA and year FE. $t = -1$ (2016/17) is the reference period. Cluster-robust SEs at the LA level.

5.3 Robustness

Table 4 reports robustness checks across three dimensions.

Alternative specifications. Dropping the COVID-truncated 2019/20 academic year reduces the point estimate to 0.95 ($p = 0.92$), confirming that the partial final year does not drive

the result. Trimming the top and bottom 5% of the exposure distribution yields a negative estimate (-14.51 , $p = 0.18$) that remains insignificant. Using the 2015 business counts instead of 2016 for the exposure measure produces $\beta = 5.51$ ($p = 0.68$).

Randomization inference. Permuting Levy exposure across LAs 500 times yields a randomization inference p -value of 0.63, confirming that the standard cluster-robust inference is not misleading.

Leave-one-out. Sequentially dropping each of the five largest LAs (Birmingham, Durham, Leeds, Liverpool, Wiltshire) produces coefficients between 3.0 and 5.0, all insignificant. No single LA drives the result.

Table 4: Robustness Checks

	Estimate	SE	p -value	N
<i>Panel A: Specification</i>				
Baseline	3.9407	(13.4253)	0.7696	1,230
Drop 2019/20 (COVID truncated)	0.9482	(9.9513)	0.9242	1,107
Trim extreme exposure (5–95%)	-14.5132	(10.6803)	0.1770	1,110
<i>Panel B: Inference</i>				
RI p -value (500 permutations)		0.6280		
<i>Panel C: Leave-one-out</i>				
Drop Birmingham	3.9952	(13.3991)	0.7661	
Drop Durham	3.0065	(13.5754)	0.8251	
Drop Leeds	3.6929	(13.5256)	0.7853	
Drop Liverpool	4.7876	(13.3239)	0.7200	
Drop Wiltshire	4.9664	(13.3465)	0.7105	

Notes: Dependent variable is log(apprenticeship starts). All specifications include LA and year FE with cluster-robust SEs. Panel B permutes levy exposure across LAs. Panel C drops the five largest LAs.

6. Discussion

The precise null on geographic crowding out has three implications. First, the Levy’s dramatic compositional shift—from Level 2 to Level 6+ apprenticeships—operated primarily within firms rather than between local labor markets. Large employers redirected their own training investment toward degree apprenticeships for existing employees, absorbing levy funds that might otherwise have supported entry-level training. This substitution reduced the national supply of Level 2 places but did not create measurable geographic spillovers to nearby SMEs.

Second, the null suggests that training provider capacity constraints, while potentially relevant nationally, did not bind differentially across local labor markets. If providers in high-exposure LAs had been capacity-constrained by levy-funded demand, total starts should have fallen (or risen less) in those areas. The absence of any differential effect implies either that providers could expand capacity to meet both levy and non-levy demand, or that the demand shift was too gradual to create local bottlenecks within the three post-Levy years observed.

Third, the finding informs ongoing UK policy debates about reforming or replacing the Levy. If geographic crowding out had been significant, the policy response might focus on protecting local training capacity—for example, by ring-fencing a share of levy funds for entry-level training in each region. The null suggests instead that the policy lever is the incentive structure within firms: the relative cost of degree versus entry-level apprenticeships, and the rules governing who can be enrolled as an apprentice.

Limitations. Three data limitations constrain the analysis. First, the panel extends only three years past the Levy’s introduction, and the final year (2019/20) was truncated by COVID-19. Longer-run capacity reallocation may generate effects that cannot be detected in this window. Second, the historical FE Data Library reports total starts per LA but not starts by apprenticeship level, preventing a direct test of whether high-exposure LAs experienced larger compositional shifts from Level 2 to Level 6+. The null on total starts is consistent with within-LA compositional substitution (Level 2 declining while Level 4+ rises) that nets out in the aggregate. Future work using the DfE Explore Education Statistics, which disaggregates by level from 2018/19 onward, could test this mechanism directly—though the shorter pre-Levy window would weaken identification. Third, the 123 matched LAs represent roughly 40% of English LAs; the remaining were lost to name-matching failures between data sources. A formal power calculation suggests the analysis can detect a differential decline of approximately 8–10% in starts at conventional significance levels, which is well below the 31% national aggregate decline but may miss subtler geographic effects.

7. Conclusion

The UK’s Apprenticeship Levy transformed the composition of apprenticeship training without reducing its geographic availability. The policy’s primary distortion—converting youth entry-level training into adult degree-level credentials—is a within-firm phenomenon driven by asymmetric funding incentives, not a local labor market externality. For policymakers considering training levies, the lesson is that the design of spending rules matters more than

the geography of levy-payer concentration: the problem is what employers buy with their funds, not where they buy it.

Acknowledgements

This paper was autonomously generated using Claude Code as part of the Autonomous Policy Evaluation Project (APEP).

Project Repository: <https://github.com/SocialCatalystLab/ape-papers>

Contributors: @SocialCatalystLab

First Contributor: <https://github.com/SocialCatalystLab>

References

- Acemoglu, Daron and Jörn-Steffen Pischke**, “Beyond Becker: Training in Imperfect Labour Markets,” *Economic Journal*, 1999, 109 (453), F112–F142.
- Benzarti, Youssef and Dorian Carloni**, “Who Really Benefits from Consumption Tax Cuts? Evidence from a Large VAT Reform in France,” *American Economic Journal: Economic Policy*, 2019, 11 (1), 38–78.
- Borusyak, Kirill, Peter Hull, and Xavier Jaravel**, “Quasi-Experimental Shift-Share Research Designs,” *Review of Economic Studies*, 2022, 89 (1), 181–213.
- Callaway, Brantly and Pedro H.C. Sant’Anna**, “Difference-in-Differences with Multiple Time Periods,” *Journal of Econometrics*, 2021, 225 (2), 200–230.
- de Chaisemartin, Clément and Xavier D’Haultfoeuille**, “Two-Way Fixed Effects Estimators with Heterogeneous Treatment Effects,” *American Economic Review*, 2020, 110 (9), 2964–2996.
- Dougherty, Sean M., Henry Lahr, and Vanessa Smith Morest**, “Reforming the Apprenticeship Levy,” *Journal of Education and Work*, 2020, 33 (7-8), 475–493.
- Dustmann, Christian and Uta Schönberg**, “What Makes Firm-Based Vocational Training Schemes Successful?,” *American Economic Journal: Applied Economics*, 2012, 4 (3), 36–61.
- Fuller, Alison and Lorna Unwin**, “The Apprenticeship Levy—Time for a Rethink?,” Working Paper, UCL Institute of Education 2019.
- Goldsmith-Pinkham, Paul, Isaac Sorkin, and Henry Swift**, “Bartik Instruments: What, When, Why, and How,” *American Economic Review*, 2020, 110 (8), 2586–2624.
- Goodman-Bacon, Andrew**, “Difference-in-Differences with Variation in Treatment Timing,” *Journal of Econometrics*, 2021, 225 (2), 254–277.
- Leuven, Edwin and Hessel Oosterbeek**, “An Alternative Approach to Estimate the Wage Returns to Private-Sector Training,” *Journal of Applied Econometrics*, 2005, 19 (4), 423–434.
- Patrignani, Pietro, Steven Sherwood, and Martin Sherwood**, “The Apprenticeship Levy: Has It Achieved Its Objectives?,” Discussion Paper 34, Centre for Vocational Education Research 2021.

Roth, Jonathan, “Pretest with Caution: Event-Study Estimates after Testing for Parallel Trends,” *American Economic Review: Insights*, 2022, 4 (3), 305–322.

Saez, Emmanuel, Benjamin Schoefer, and David Seim, “Payroll Taxes, Firm Behavior, and Rent Sharing: Evidence from a Young Workers’ Tax Cut in Sweden,” *American Economic Review*, 2019, 109 (5), 1717–1763.

A. Standardized Effect Sizes

Table 5: Standardized Effect Sizes

Outcome	$\hat{\beta}$	SE	SD(Y)	SDE	SE(SDE)	Classification
log(Starts)	3.9407	13.4253	0.735	0.0105	0.0358	Small positive
Starts (levels)	6951.0	28197.8	1472.4	0.0092	0.0375	Small positive

Notes: $SDE = \hat{\beta} \times SD(X)/SD(Y)$ (continuous treatment). Treatment: share of 250+ employee enterprises (2016). Sample: 1,230 observations, 123 LAs, 2010/11–2019/20. Classification refers to effect magnitude, not statistical significance.