Learning by Doing (Together): Collaboration and Teacher Skill Formation

John Fallon

June 25, 2025

Abstract

How do workplace collaborations affect individual worker productivity once the partnership ends? I study co-teaching in schools to measure whether teachers genuinely learn from collaboration or simply benefit from immediate assistance. Using administrative data on all teachers in a state from 2012-2019, I exploit plausibly exogenous variation in co-teaching assignment driven by special education enrollment and scheduling constraints rather than teacher quality. Teachers experience persistent improvement in student achievement after returning to solo teaching, with gains averaging 0.04σ and reaching 0.10σ when paired with highly experienced partners (16+ years). These lasting productivity gains demonstrate that strategic workplace collaboration can accelerate skill development, with important implications for team formation and mentorship design across industries.

1 Introduction

I examine how workplace collaboration creates lasting peer effects on individual worker productivity. Using co-teaching partnerships in education, I can observe worker productivity both during and after collaborative relationships end, allowing me to distinguish match-specific returns from persistent peer effects.

Co-teaching represents a service delivery model that enables schools to satisfy dual federal mandates. The Individuals with Disabilities Education Act requires instruction of students with disabilities in the least restrictive environment, while the No Child Left Behind Act mandates that a "highly qualified teacher" deliver instruction in all content areas. Under co-teaching arrangements, a general education teacher provides content expertise, while a special education teacher supports a classroom that contains both students with disabilities and those without. This structure creates a non-restrictive learning environment without requiring dual licensure in both content areas and special education for either instructor. Importantly for identification, co-teaching assignment depends primarily on special education enrollment and scheduling constraints rather than teacher quality, providing plausibly exogenous variation in collaboration exposure.

Multiple strategies exist for dividing classroom responsibilities between co-teachers, though I remain agnostic on this choice. This stance reflects both the unobservable nature of specific strategies in my data and the reality that actual implementation likely emerges through continuous negotiation between paired teachers. The essential feature for my analysis is that co-teaching forces two teachers into close proximity, creating stronger peer influence opportunities than typical same-school teacher interactions typically do.

Existing co-teaching research emphasizes student outcomes within co-taught classrooms (Austin, 2001; Jones and Winters, 2024; Jones et al., 2025; Ballis and Heath, 2021; Bessette, 2008; Cook et al., 2017; De Backer et al., 2023; Andersen et al., 2020; Conderman and Hedin, 2014). Rather than contributing directly to this student-focused literature, I provide novel descriptive evidence on teacher selection into co-teaching, pairing mechanisms, and subsequent effects on teacher assignment and turnover patterns.

I find positive peer effects. Teachers experience a discrete improvement in student test score performance after returning to solo teaching following a co-teaching experience. Comparing the same teachers' value-added measures before co-teaching with their performance in the first year after returning to solo teaching, I find improvement averaging 0.04σ in student achievement gains. This improvement is larger (0.10σ) when the co-teaching partner had 16+ years of experience. Since I observe the same teachers both before and after collaboration, these effects cannot be attributed to selection of better teachers into co-teaching arrangements.

This result addresses a fundamental question in personnel economics: when do workers actually learn from each other, or do they simply benefit from immediate assistance during collaboration? Most peer effects research measures contemporaneous impacts while workers collaborate (Jackson and Bruegmann, 2009; Nix, 2020; De Grip and Sauermann, 2012; Battu et al., 2003; Azoulay et al., 2010; Mas and Moretti, 2009; Ichino and Maggi, 2000), making this distinction impossible. I provide clear evidence of genuine skill transfer as the collaboration experience persistently changes how teachers approach their individual work.

I identify collaboration as a mechanism for accelerating returns to experience. The literature establishes that worker productivity increases with tenure (Arrow, 1962; Chen, 2021; Haggag et al., 2017; Ost, 2014; Shaw and Lazear, 2008; Thompson, 2010, 2012; Guryan et al., 2009). However, methods for accelerating this process remain poorly understood. Using teachers to explicitly coach other teachers shows promise, but loses efficacy at scale (Kraft et al., 2018). I show that strategic pairing with experienced workers can compress years of learning into shorter periods. This has immediate implications for training programs, mentorship design, and team formation across industries.

My research design exploits plausibly exogenous variation in collaborative partnership assignment. Unlike most workplace collaboration, which involves voluntary sorting or managerial selection based on complementary skills, co-teaching assignment depends heavily on institutional factors: special education enrollment fluctuations, scheduling constraints, and administrative requirements for legal compliance. While not perfectly random, this assignment process is largely disconnected from teacher quality considerations that would bias peer effects estimates.

The institutional setting provides additional analytical advantages. I observe detailed productivity measures (student test scores) for individual workers both during and after collaboration. I can track the same workers across multiple years and different partnership configurations. I observe complete peer histories, allowing me to separate experience effects from partner-specific learning.

I document several patterns relevant to personnel policy. Most collaborations are brief. Fewer than 40% of partnerships continue into a second year. Organizations frequently pair inexperienced workers together, potentially missing opportunities for knowledge transfer. The largest productivity gains occur when experienced workers partner with less experienced colleagues, suggesting that mentorship structures may be more effective than peer collaboration among equals.

My findings extend beyond education to any setting where workers can learn from observing and interacting with colleagues. The results inform debates about team formation, training program design, and optimal allocation of experienced workers within organizations. They suggest that collaboration's value lies not just in immediate productivity gains, but in its capacity to accelerate individual skill development with lasting effects. This contrasts with findings in other contexts where collaboration benefits do not persist beyond the partnership (Chen, 2021).

2 Data

I use longitudinal administrative data from a State's Department of Education covering all students and teachers from 2012 to 2019. A unique classroom identifier links students to teachers. Student data include annual demographics, supplemental service classifications (special education, English language learners), and standardized test scores in math and ELA, which I standardize by grade and year. I define co-taught classrooms as classes with exactly two teachers: one flagged as "Team-Teacher" or "Co-teacher" (the special education teacher) and one general education teacher. For my analysis of co-teaching spillovers, I restrict the sample to teachers identified with a "First Year" flag to ensure accurate experience measures. This restriction ensures accurate experience measures, as the administrative "Years of Experience" field ignores career breaks and would misclassify returning teachers.

Table 1: Summary Statistics

	Co-Taught Sample			Event Study Sample		
of Teachers	10,100			804		
of Students	152,847			61,652		
of Classes	28,588			7,361		
	Obs	Mean	SD	Obs	Mean	SD
Test Z-Score	455,759	-0.387	0.965	171,843	-0.141	0.986
Special Ed	630,310	0.329		171,843	0.227	
ELL	630,310	0.097		171,843	0.089	
FRPL	630,310	0.606		171,843	0.506	

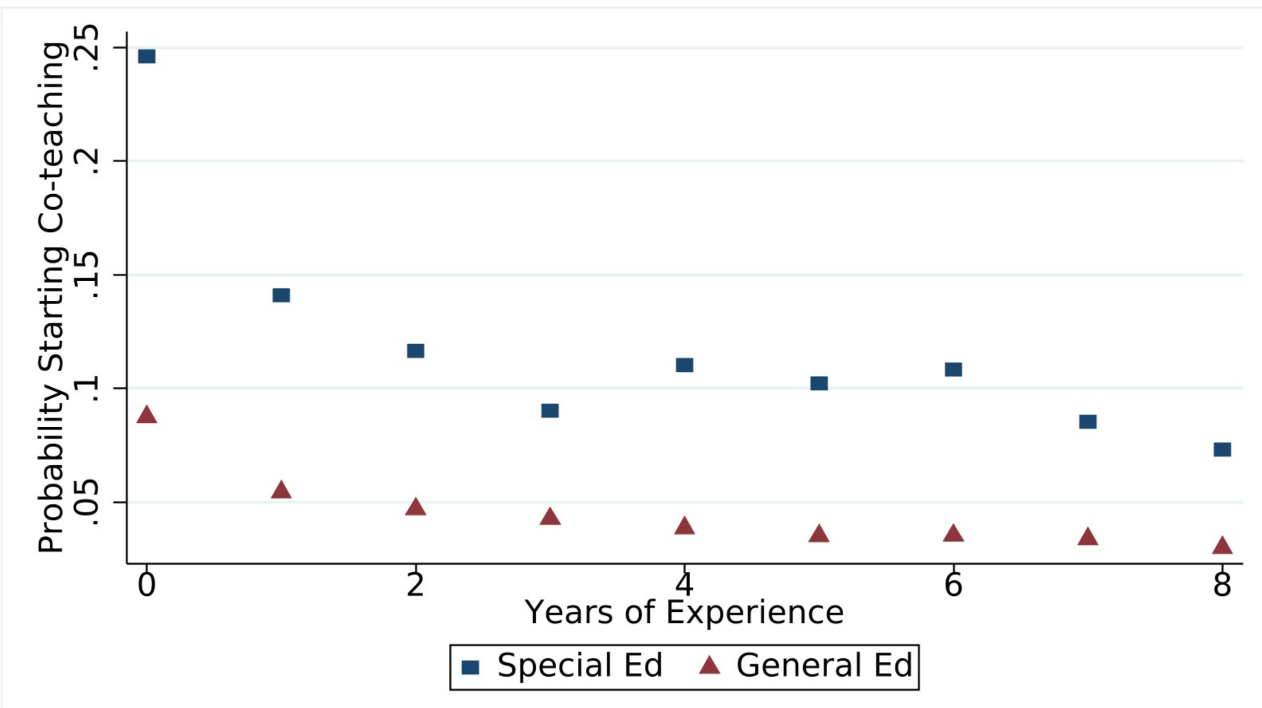
This table presents descriptive statistics for two samples: the full co-taught sample (all teachers and students in co-taught classrooms from 2012-2019) and the restricted event study sample (teachers with specific experience profiles used in the main analysis).

3 Descriptive Analysis

3.1 Teachers enter co-teaching early, but pairs break up often.

To identify the effects of co-teaching on teachers, I first establish selection patterns into co-teaching. In Figure 1, I calculate the probability that a new teacher begins co-teaching, given their years of experience, conditional on not having co-taught previously. This hazard rate indicates that special education teachers are more likely to be assigned to co-teaching pairs than general education teachers; however, both types are more likely to begin co-teaching early in their careers.

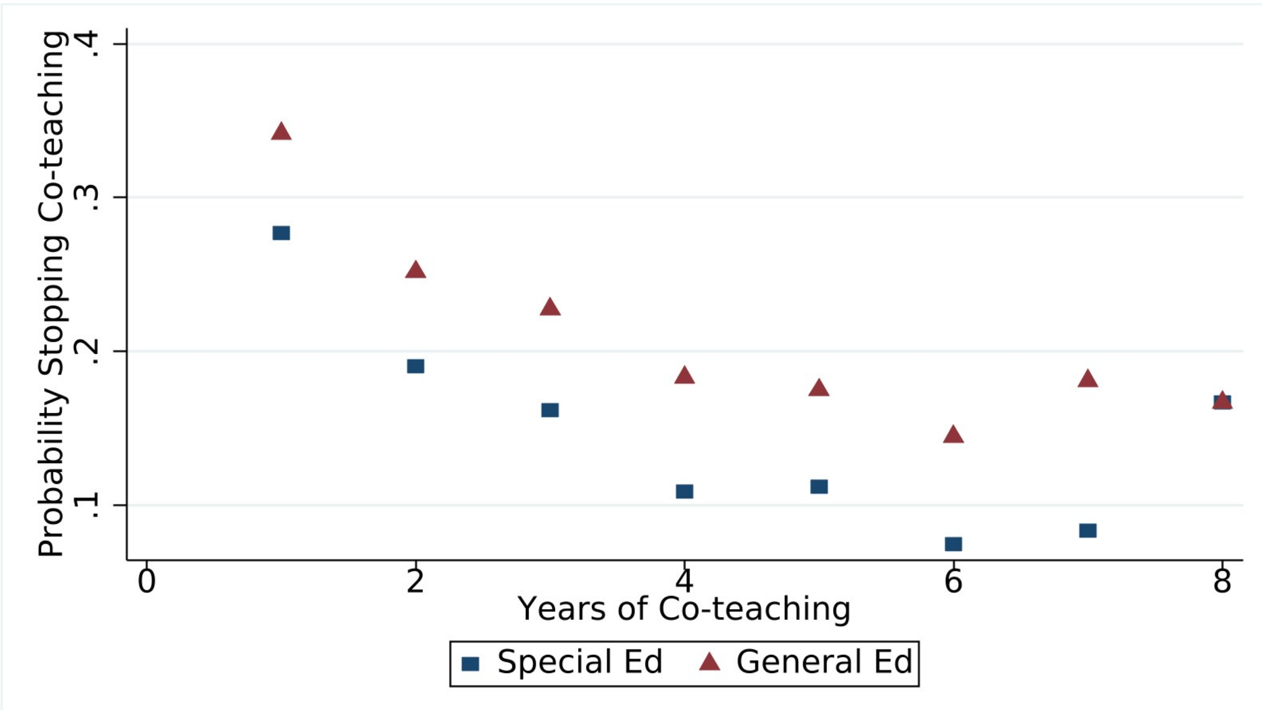
Figure 1: Hazard Rate for Entering Co-Teaching



This figure shows the probability that a teacher begins co-teaching for the first time, conditional on their years of experience and not having co-taught previously, separately for special education and general education teachers.

In Figure 2, I illustrate the probability of not co-teaching the following year, given that one co-teaches the prior year and teaches the following year. General education teachers are less attached to co-teaching, but the odds of transitioning out decrease with co-teaching experience for both types.

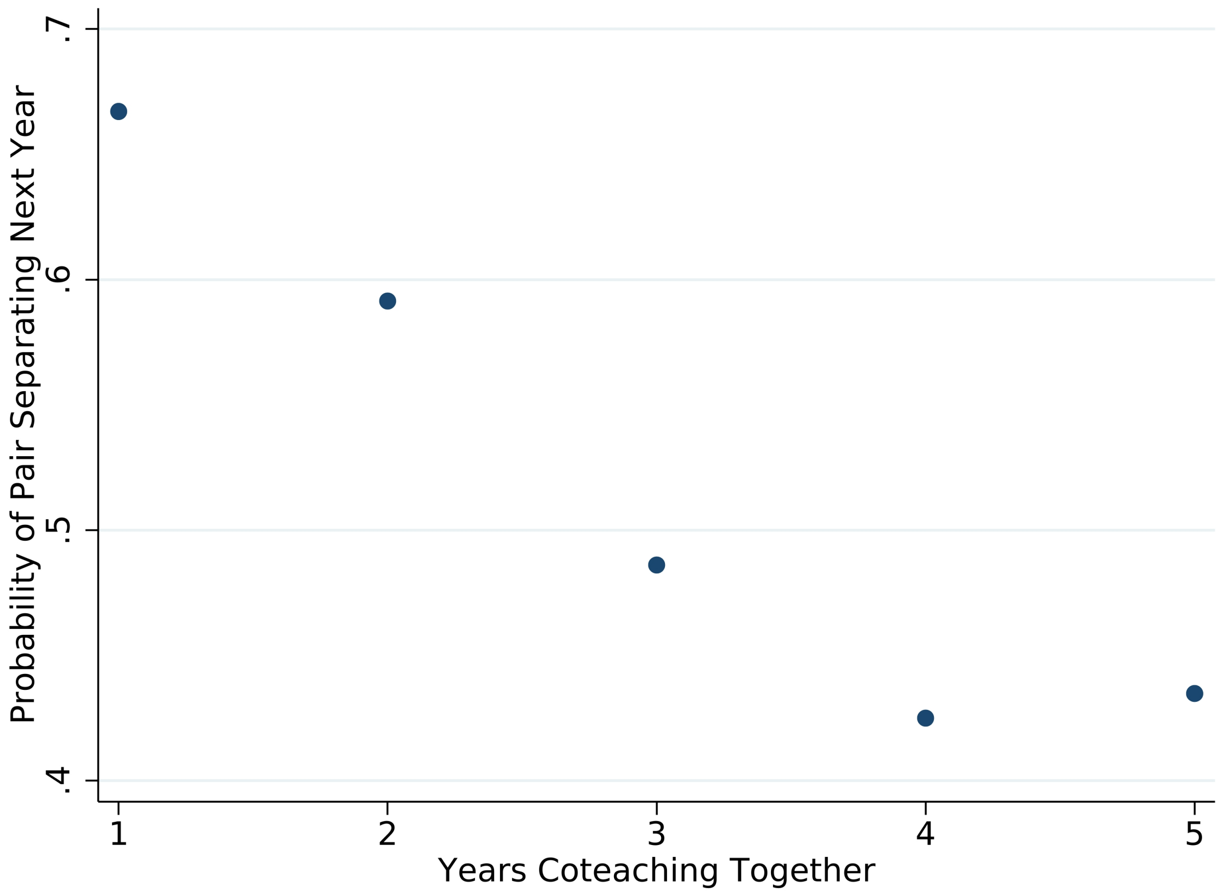
Figure 2: Hazard Rate for Continuing Co-Teaching



This figure displays the probability that a teacher stops co-teaching in the following year, conditional on co-teaching in the current year and continuing to teach, by teacher type and co-teaching experience.

Individual co-teaching pairs have a high likelihood of separation, with odds remaining above 40% even after five years together (Figure 3). Separation odds are particularly high after the first year, with over 60% of pairs not working together the following year. Some pairs reunite after time apart, but the pattern persists when I plot the hazard rate for pairs never working together again (Figure A.1). These high separation rates suggest that co-teaching assignment is largely driven by institutional needs rather than teacher preferences or administrator judgment about optimal matches.

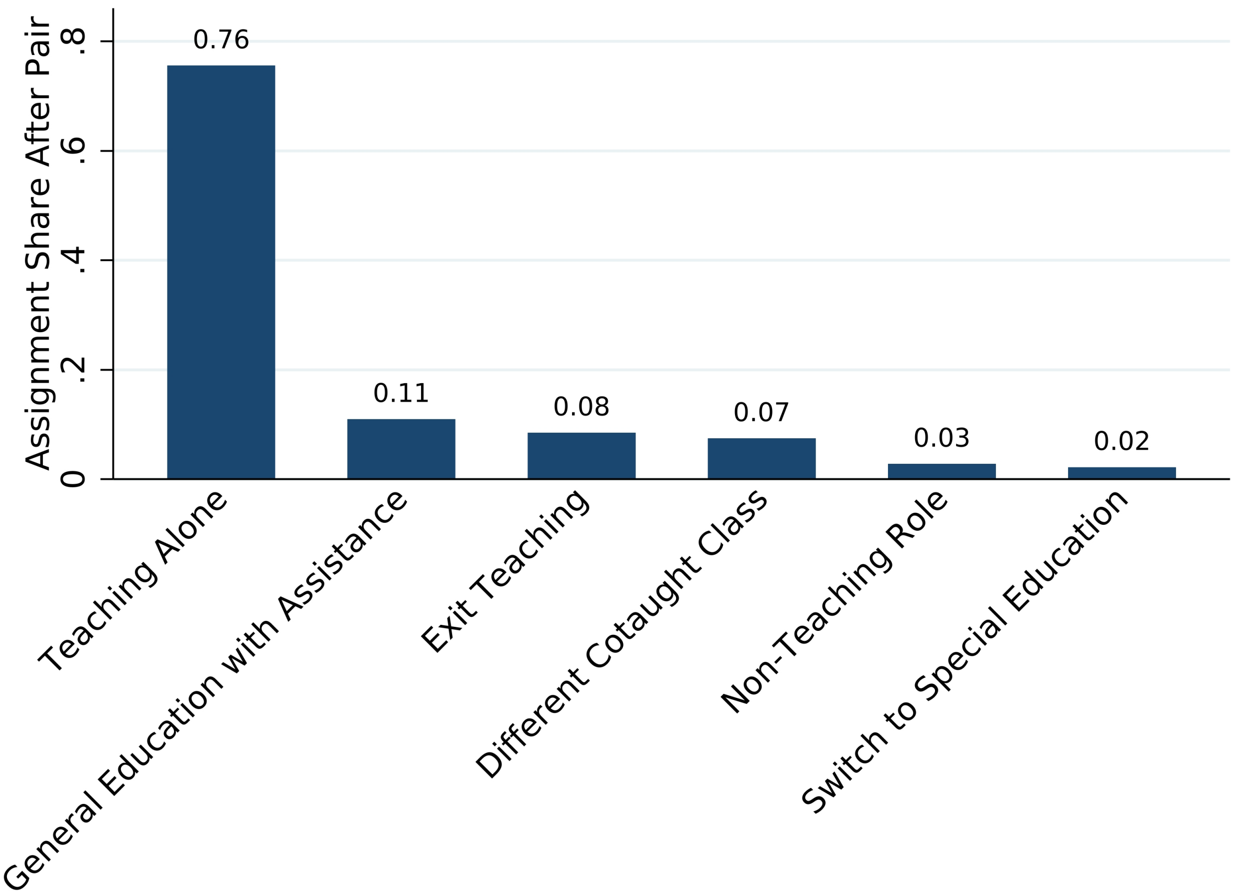
Figure 3: Hazard Rate for Co-teaching Pair Separation



This figure shows the probability that co-teaching pairs separate in the following year, conditional on the number of years they have worked together as a pair.

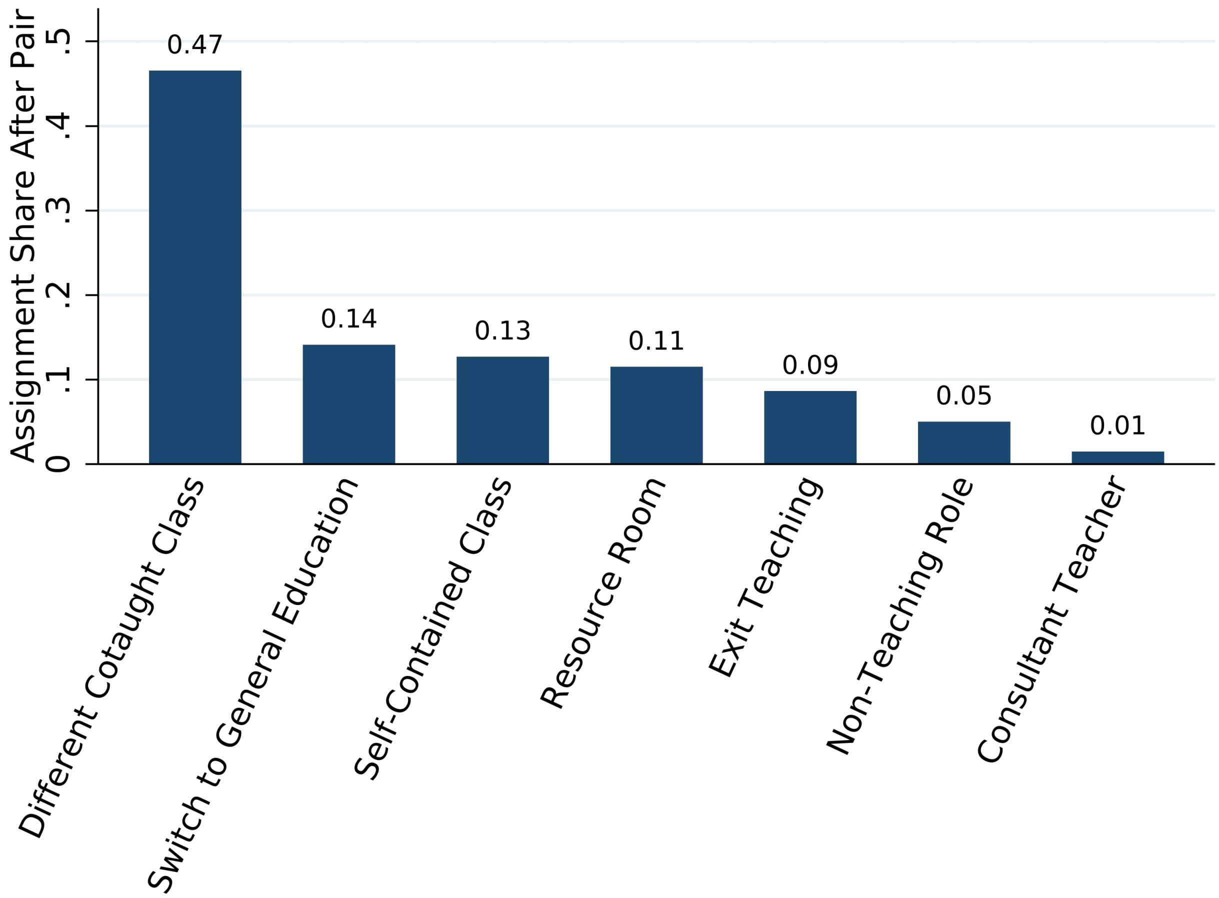
After pair break-ups, general education teachers typically return to traditional classrooms with 76% teaching alone (Figure 4). Special education teachers are likely to cycle through other co-teaching pairs with 47% teaching in different co-taught classes (Figure 5).

Figure 4: What General Education Teachers Do After Co-Teaching



This figure shows the distribution of classroom assignments for general education teachers in the year immediately following the end of a co-teaching partnership.

Figure 5: What Special Education Teachers Do After Co-Teaching



This figure displays the distribution of classroom assignments for special education teachers in the year immediately following the end of a co-teaching partnership.

A key identification concern is that co-teaching teachers may be systematically assigned different student populations. I test this using difference-in-differences comparing student assignment before and after co-teaching begins. General education teachers who co-teach initially receive 0.7 percentage points fewer students with disabilities, but 2 percentage points more after beginning co-teaching (Table 2). General education teachers who co-teach receive 2.2 percentage points more students with free or reduced-price lunch than those who do not co-teach (Table 4). This gap grows by an additional 2.4 percentage points after first co-teaching. I find no statistically significant difference in English Language Learner assignment (Table 3).

Table 2: Difference in Difference on Share Of Students With Disabilities

Dependent Variable:	Share Special Ed		
Teacher Type	All	General Education	Special Education
Ever Co-taught	0.008	-0.007**	-0.090***
	(0.005)	(0.002)	(0.011)
Ever Co-taught * Post	-0.006	0.020***	-0.312***
	(0.005)	(0.002)	(0.012)
(N)	210,263	181,736	28,411
Dependent Variable Mean	0.232	0.159	0.695

Standard errors in parentheses. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

This table uses difference-in-differences estimation to compare the share of students with disabilities assigned to teachers who ever co-teach versus those who never co-teach, before and after co-teaching begins.

Table 3: Difference in Difference on Share Of English Language Learners

Dependent Variable:	Share English Language Learners		
Teacher Type	All	General Education	Special Education
Ever Co-taught	0.002	0.001	0.010*
	(0.001)	(0.002)	(0.005)
Ever Co-taught *Post	-0.002	-0.002	-0.005
	(0.002)	(0.002)	(0.004)
(N)	210,263	181,736	28,411
Dependent Variable Mean	0.059	0.059	0.061

Standard errors in parentheses. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

This table uses difference-in-differences estimation to compare the share of English Language Learners assigned to teachers who ever co-teach versus those who never co-teach, before and after co-teaching begins.

Table 4: Difference in Difference on Share Of Students With Free or Reduced Price Lunch

Dependent Variable:	Share Free and Reduced Price Lunch		
Teacher Type	All	General Education	Special Education
Ever Co-taught	0.025***	0.022***	0.002
	(0.004)	(0.004)	(0.009)
Ever Co-taught * Post	0.024***	0.024***	-0.018*
	(0.004)	(0.004)	(0.009)
(N)	210,263	181,736	28,411
Dependent Variable Mean	0.507	0.489	0.622

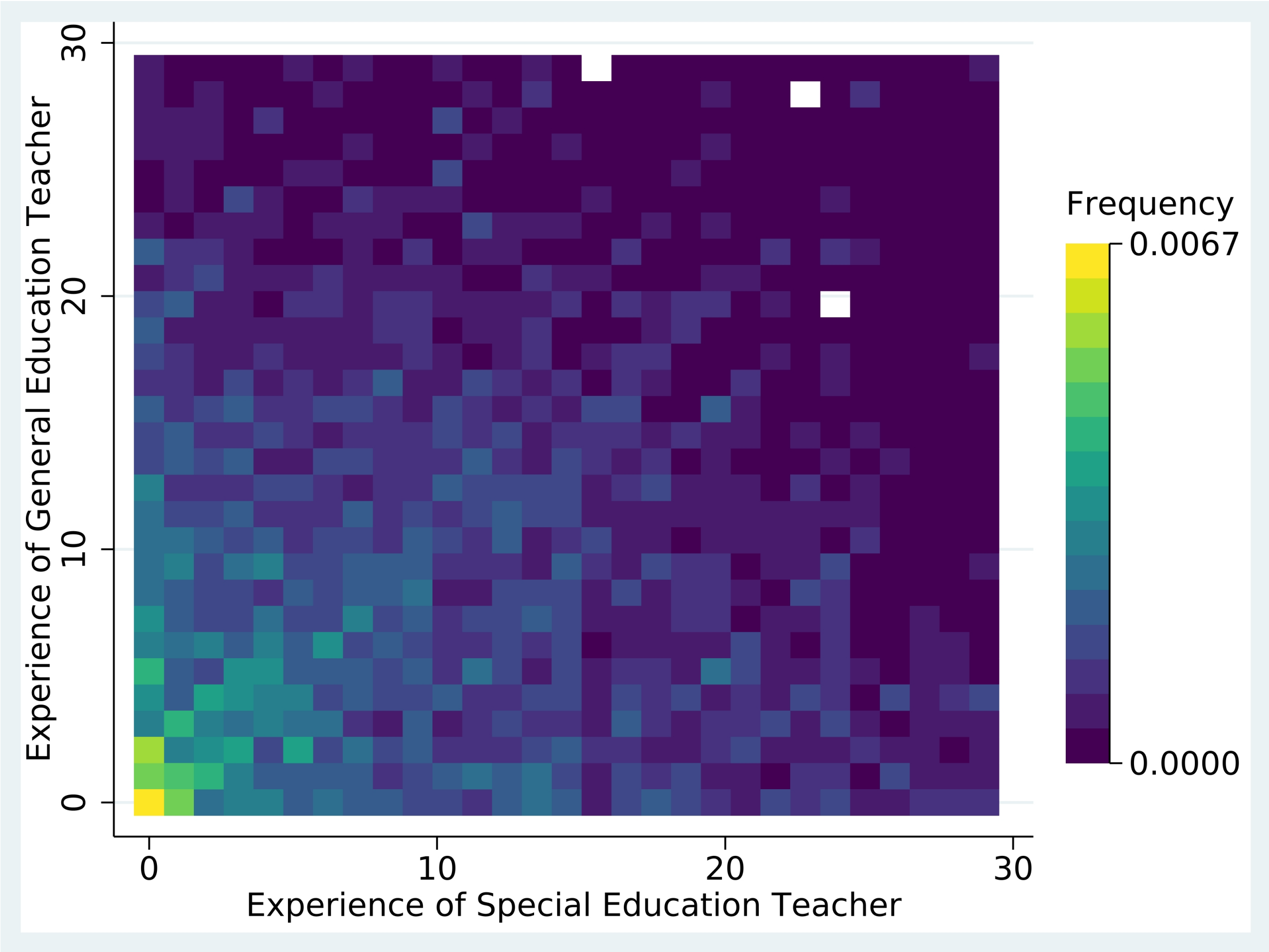
Standard errors in parentheses. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

This table uses difference-in-differences estimation to compare the share of economically disadvantaged students assigned to teachers who ever co-teach versus those who never co-teach, before and after co-teaching begins.

To investigate administrator pairing patterns, I create two heatmaps. The first measures pairing frequency between teachers of two ability levels (Figure 6). The second shows the difference between observed pair frequency and the frequency under random assignment of special education and general education co-teachers (Figure 7).

3.1.1 How do they differ by experience?

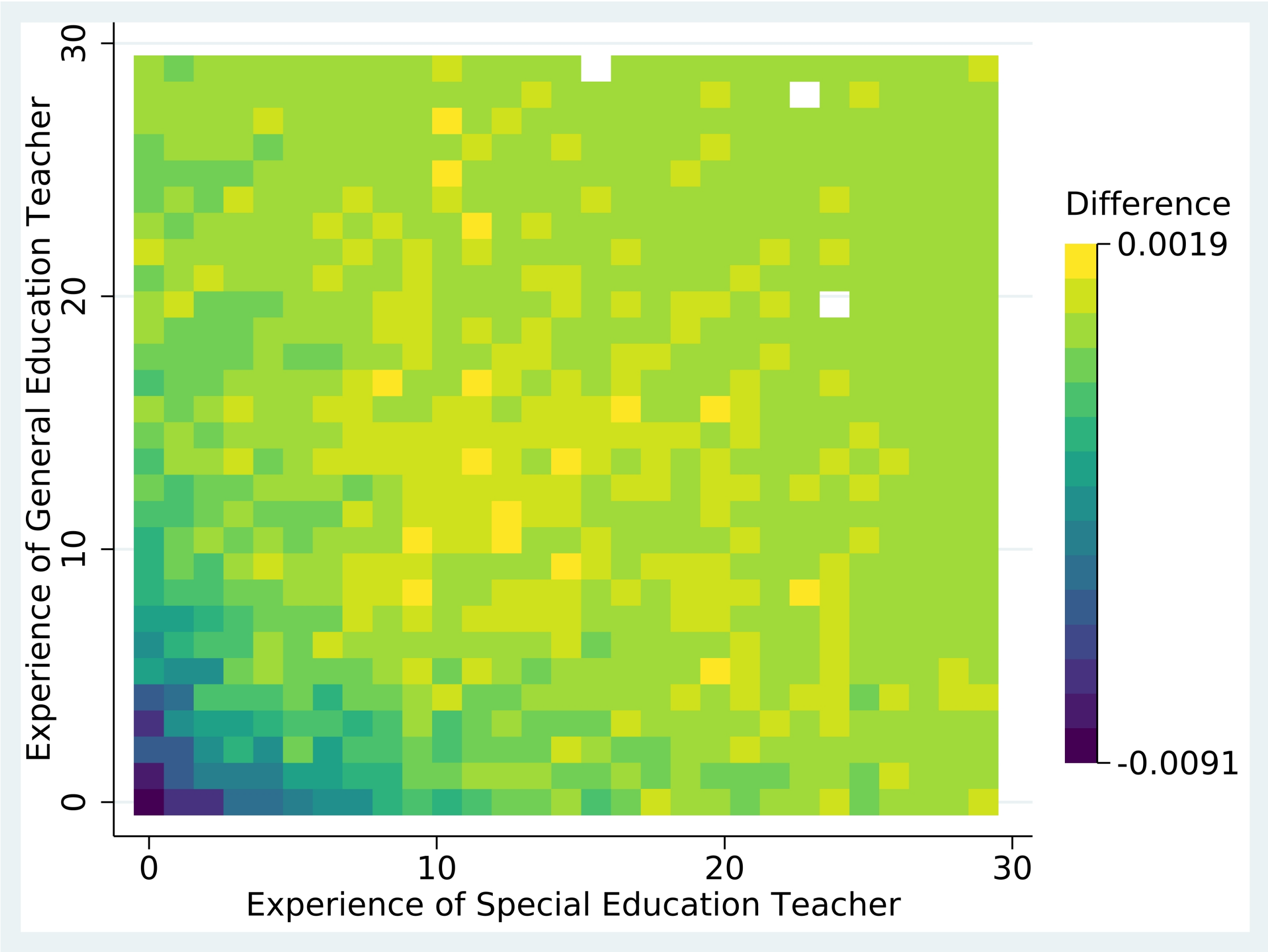
Figure 6: Pairing Frequency by Experience



This heatmap shows the observed frequency of co-teaching pairs based on the experience levels (in years) of both the special education and general education teachers in the partnership.

While two inexperienced teachers form the most common pairing type, this likely reflects the higher co-teaching entry rates of inexperienced teachers. Compared to random assignment frequency, administrators strategically avoid pairing two inexperienced teachers. Novice teacher pairs are 0.9 percentage points less likely than under random assignment. Experienced pairs appear slightly more common than random assignment would predict. These differences remain small as strategic pairing occurs infrequently.

Figure 7: Difference From Random Pairing of Teachers



This heatmap displays the difference between observed co-teaching pairing frequencies by experience level and the frequencies that would be expected under random assignment of available teachers to partnerships.

4 Teacher Development

4.1 Inside of Co-teaching

4.1.1 Are Persistent Pairs Better?

I find no evidence that match quality drives partnership duration. Instead, better individual teachers are more likely to remain in lasting partnerships.

A key question about successful pair matching is what drives success. One explanation is that successful pairs work better together and are less likely to separate. I test this by regressing co-taught student test scores on the maximum duration of a pair. As shown in Figure 5, Special education teachers are frequently used in multiple pairings so that I can separate the pair-specific effect from the teacher-specific effect. Longer-lasting pairs associate with higher test scores, but this relationship disappears, possibly even reverses, when I control for the teachers within the pair (Table 5).

Table 5: Regression of Duration on Student Test Scores

Dependent Variable (DV):	Student Test Scores		
Subject	Combined	Math	English
	Without Teacher Fixed Effects		
Years Spent as Pair (Total)	0.00809*	0.0125**	0.00649
	(0.00365)	(0.00421)	(0.00470)
N	142,805	61,646	81,143
	With Teacher Fixed Effects		
Years Spent as Pair (Total)	-0.00455	-0.0313*	0.000386
	(0.00688)	(0.0121)	(0.0105)
N	142,525	61,408	80,903

Standard errors in parentheses
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

This table regresses student test scores on the total number of years a co-teaching pair worked together, with and without teacher fixed effects, to test whether longer-lasting pairs produce better student outcomes.

4.1.2 Do Pairs Improve?

While matched pairs may not be inherently better fits, they may learn to work better together over time. It is challenging to identify returns to experience when year-specific events are present. Co-teaching assignment patterns likely vary over time; however, when controlling for pair fixed effects, year fixed effects are collinear with experience level fixed effects, unless pairs take years off. Relying on pairs taking years off requires assuming that those pairs that do take time off experience no decline in skill. To avoid this assumption, I investigate this using a two-step approach as described in Papay and Kraft (2015).[1] This two-step approach avoids collinearity between year and experience effects that would arise with standard fixed effects, while not requiring the strong assumption that interrupted partnerships maintain their skill levels.

First, I estimate year-specific fixed effects conditional on experience.

$$\pi_{spy} = \alpha[Year]_y + \gamma f(Exper_{py}) + \chi g(PairExperience_{ty}) + \phi h(\pi_{sp,y-1}) + \mu_{spy} \tag{1}$$

I then extract these year effects to estimate the impact of time together as a pair.

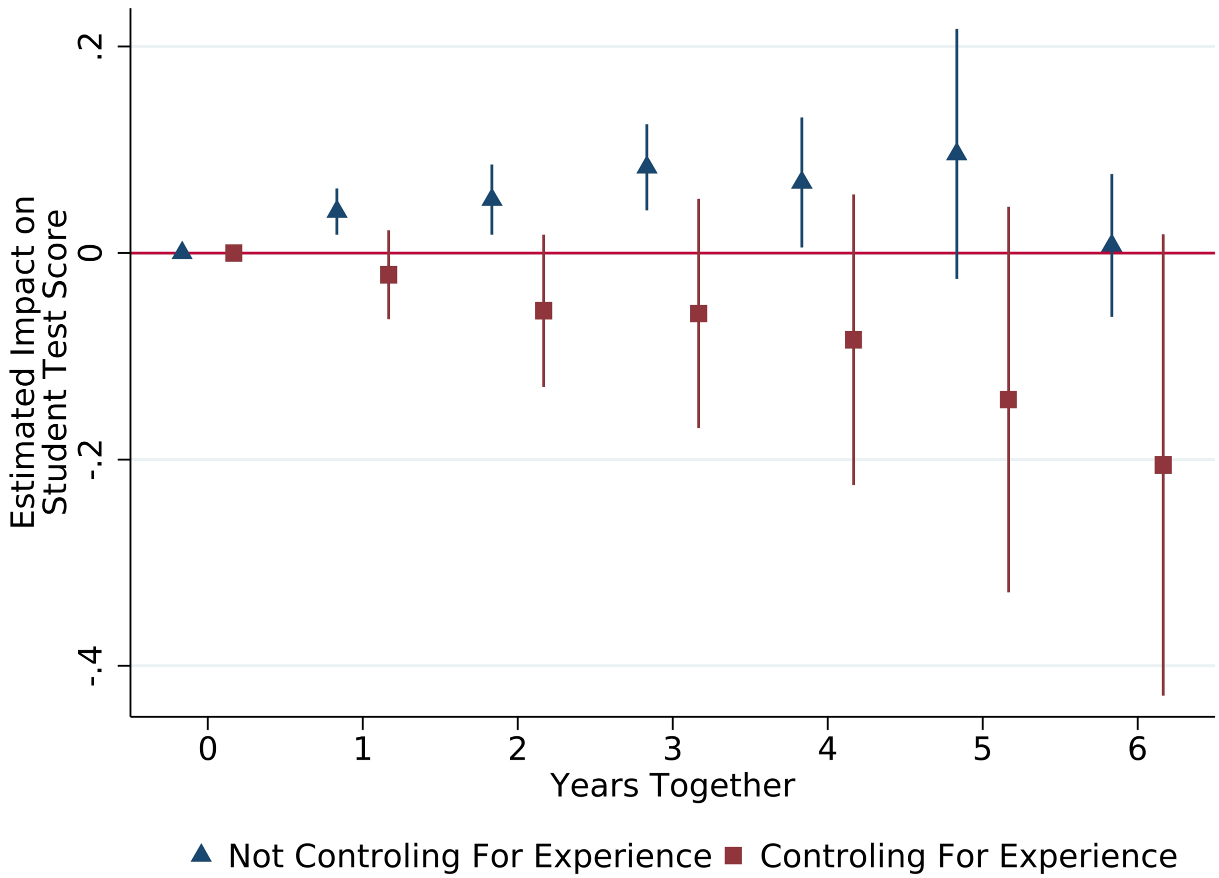
$$\pi_{spy} - \hat{\alpha}[Year]_y = \beta f(Exper_{py}) + \omega g(PairExperience_{py}) + \delta_p + h(\pi_{s,y-1}) + \epsilon_{spy} \tag{2}$$

π_{spy} is the student test score for subject s. $[Year]_y$ is the set of year dummy variables. $f(Exper_{py})$ is the functional form for returns to experience (fixed effects on the overall experience of each teacher in the pair separately). $g(PairExperience_{py})$ is the years a pair has co-taught together (variable of interest). δ_p is a pair fixed effect.

This method sidesteps potential collinearity between year fixed effects and pair experience together, assuming year effects are unrelated to pair quality conditional on experience (i.e. $COV(f(YEAR), \delta_p | g(PairExperience_{py})) = 0$).

[1] Ulitmately, both methods produce very similar results (Figure A.2).

Figure 8: Return To Experience in a Pair



This figure plots student test scores against years of co-teaching partnership duration using a two-step estimation method to separate pair-specific learning from general teaching experience effects.

Without controlling for general teacher experience, I observe a gradual increase in student test scores. This effect disappears when I introduce experience controls. This suggests that time together could be mistaken as a driver of improvement, but improvement would have occurred whether the experience was in the co-taught pair or outside it.

4.2 Outside of Co-teaching

4.2.1 Method

To identify persistent effects of co-teaching, I compare teachers' associated test scores before co-teaching to their performance after returning to solo teaching. The key identification challenge is constructing appropriate counterfactuals for teachers who co-teach.

I measure the impact of co-teaching on teacher-associated test scores using a stacked event study from Cengiz et al. (2019). This method accounts for staggered treatment issues while being selective about counterfactuals. I separate co-teaching teachers into "events" and compare them to teachers with matching profiles who do not co-teach. I separate co-teachers into different events by subject taught, year started teaching, year started co-teaching, and year left teaching. Within each event I use teachers who begin and leave teaching in the same year as co-teachers. I then "stack" events by combining them into one dataset. I ensure within-event comparisons by using event-specific time and teacher fixed effects. I restrict to events where teachers are observable teaching solo classes in one subject continuously. I also restrict co-teachers to the subject in which they co-taught. I only include "events" where co-teachers had between one and three years of experience before co-teaching. I restrict my analysis by prior experience to identify baseline teacher quality differences and minimize selection bias by excluding co-teachers with unusual profiles. I estimate the effect using:

$$\pi_{dtpsy} = \sum_{i=-3}^{2} \omega_i YTC_{tyi} + \alpha X_{psy} + \delta_{dt} + \gamma_{dy} + \epsilon_{dtpsy} \tag{3}$$

π_{dtpsy} is the student p's test score in event d for teacher t in subject s in year y. YTC_{tyi} is a dummy variable for being i years after first co-teaching. X_{psy} are student-year-subject level controls including demographics and a cubic polynomial for the student's test scores the two previous years. δ_{dt} are event-specific teacher fixed effects. γ_{dy} are event-specific year fixed effects.

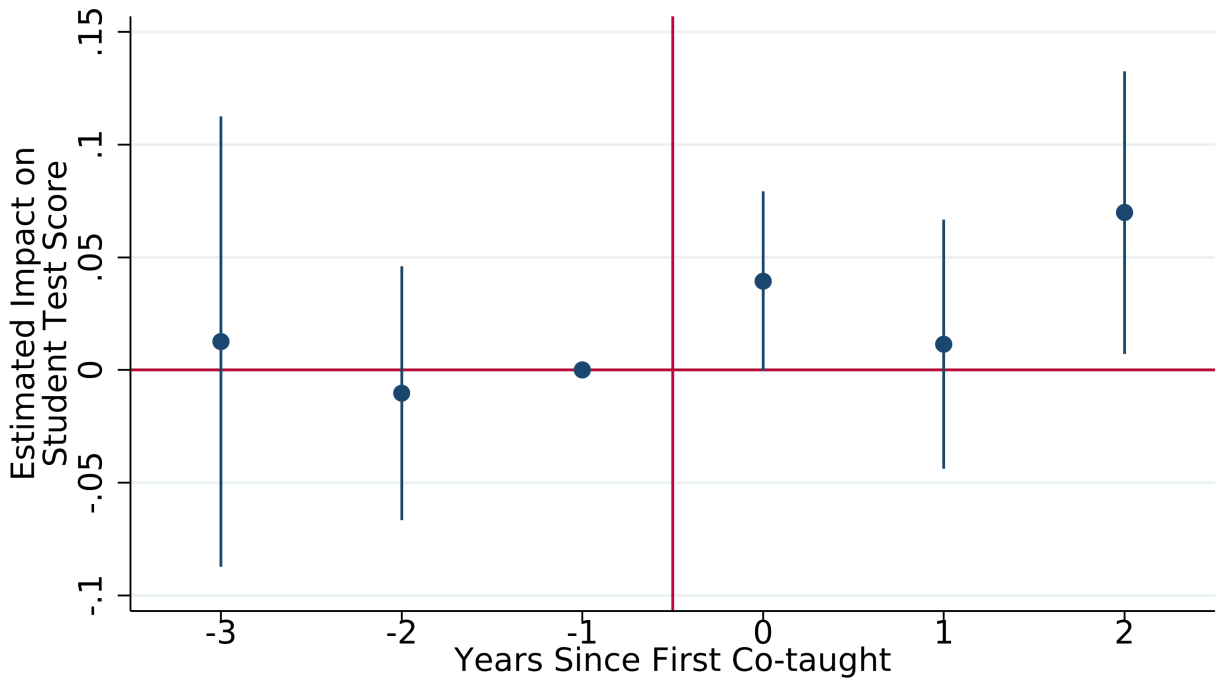
Using a stacked event study should solve the primary issue with staggered treatment, but it does not necessarily solve the issue of non-convex weighting. In a traditional two-way fixed-effect event study, there are two primary issues (Goodman-Bacon, 2021). First, the just-treated subjects are being compared to already-treated subjects, which can lead to bias when the treatment effect is time-varying. The stacked event study method solves this issue by only comparing treated subjects to untreated subjects. The second issue with traditional two-way fixed effects is the issue of weighting. Both the traditional and stacked event study implicitly weight the treated and untreated observations, which can be an issue when the weighting is non-convex (i.e., some observations have negative weights). Methods that solve this issue typically do not use concurrent controls for observations after treatment, which is necessary in my case, as student assignment varies (Wing et al., 2024). I initially ignored this potential issue and checked post-hoc to determine how this implicit weighting impacts my estimated effect (in Section 4.2.4).

4.2.2 Average Improvement

On average, I find modest increases in associated student test scores after a teacher first co-teaches (Figure 9). Prior to co-teaching, teachers who did not co-teach serve as plausible counterfactuals with relatively parallel trends. After co-teaching, co-teachers show associated student test score jumps in the first year. In the second year, the difference is no longer statistically significant, but becomes statistically significant again in the third year.

The parallel pre-trends support the identifying assumption, while the immediate jump in year 1 suggests skill transfer rather than gradual learning. The pattern in years 2-3 may reflect either persistence with noise or dynamic treatment effects.

Figure 9: Difference in Student Test Scores By Time Since First Co-taught



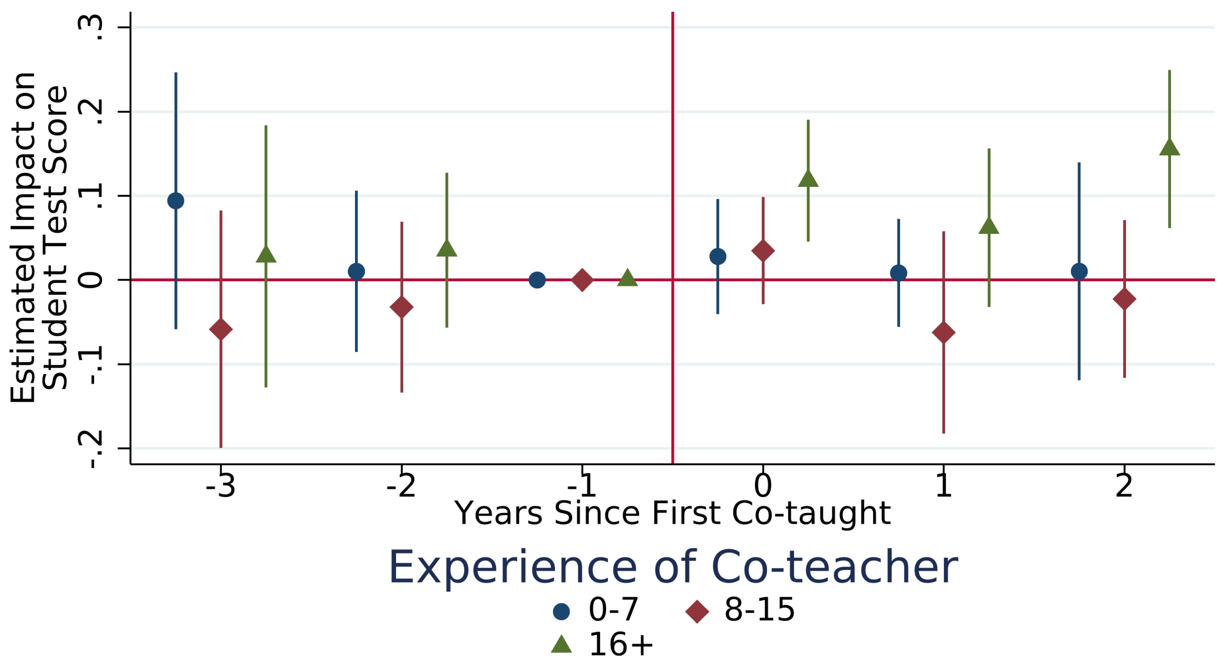
This event study compares student test score performance of teachers before and after their first co-teaching experience using a stacked difference-in-differences methodology with event-specific controls.

4.2.3 Experience of Co-teacher

The simplest mechanism for co-teaching's developmental benefit would be one teacher helping teach the other. If this is the mechanism, then the other teacher's identity should matter. To investigate this, I run separate regressions by the other teacher's experience level. Figure 10 shows the primary developmental benefit comes from very experienced co-teachers. Within each experience bucket, non-co-teachers serve as plausible counterfactuals prior to co-teaching, but estimated test score impact deviates for co-teachers who had partners with sixteen or more years of experience.

This pattern is consistent with complementary human capital models where experienced workers have more tacit knowledge to transfer, but also with diminishing returns as very experienced teachers may be less effective mentors.

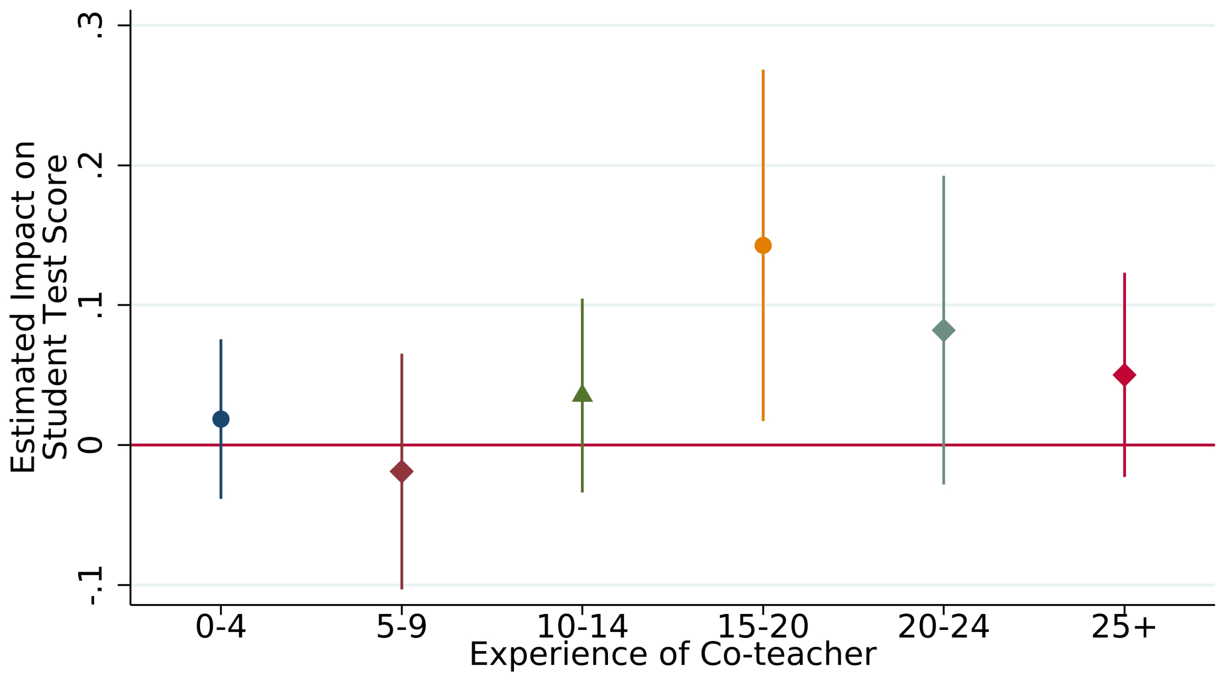
Figure 10: Difference in Student Test Scores By Time Since First Co-taught and Co-Teacher Experience



This figure disaggregates the main event study results by the experience level of the co-teaching partner, comparing effects across different partner experience categories.

To further differentiate co-teacher experience benefits, I run stacked difference-in-differences using smaller experience groupings. Generally, co-teacher benefits grow with experience until the 15-20 year bucket where returns begin declining (Figure 11).

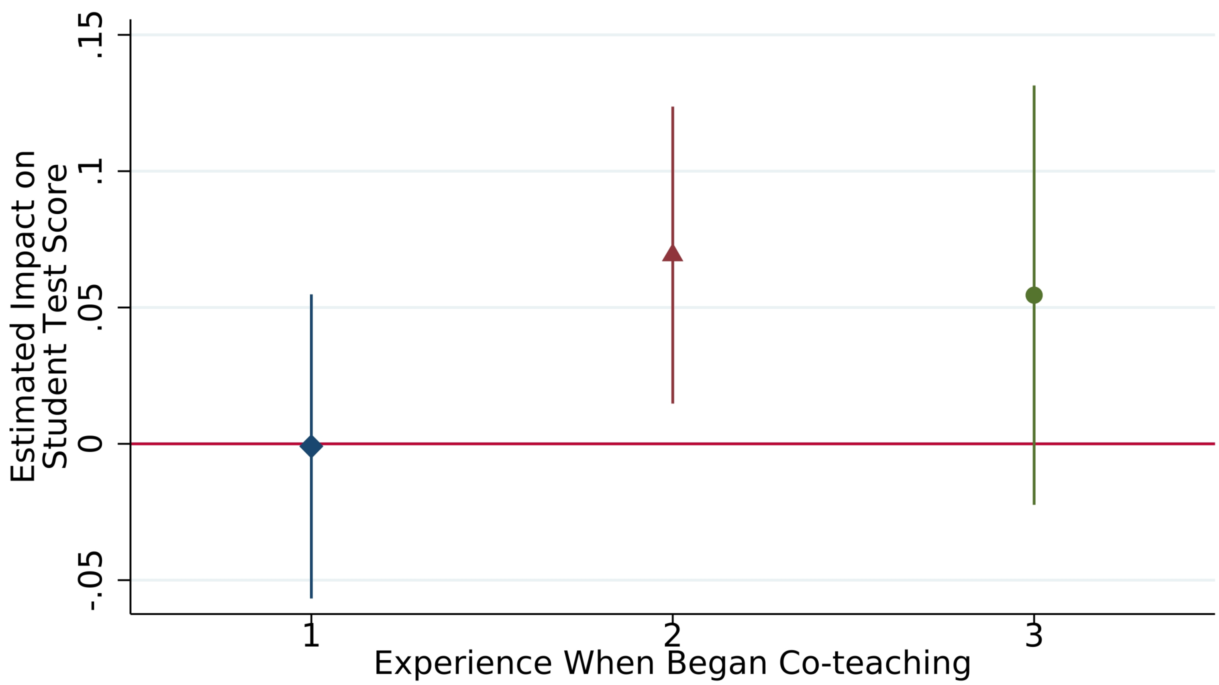
Figure 11: Difference in Student Test Scores By Time Since First Co-taught and Co-Teacher Experience



This figure shows the same analysis as Figure 10 but with more granular experience groupings for the co-teaching partner, using smaller experience bins to identify optimal experience levels.

I also investigate heterogeneity by timing of co-taught classroom assignment. On average, developmental benefits appear strongest when teachers have more than one year of experience prior to co-teaching (Figure 12). I find no difference in impact by co-teacher degree level (Figure A.3).

Figure 12: Difference in Student Test Scores By Time Since First Co-taught and Experience When First Co-taught



This figure examines heterogeneity in co-teaching effects based on when in a teacher's career they first experience co-teaching, comparing teachers with different amounts of prior solo teaching experience.

4.2.4 Robustness and Lack Thereof

These results depend on some of the choices I made in my analysis. While I believe they are necessary for an accurate estimation of the returns to co-teaching, I would like to be transparent about how my findings are and are not robust.

The positive returns to co-teaching only appear when controlling for students' prior test scores, but two years of test scores are not necessary. When I do not control for students' prior test scores there appears to be a dip in students' test scores, although it is not persistent or statistically significant (Figure A.4). It is unlikely this is an accurate representation of teacher development. Tables 2 -4 show that student assignment changes after a teacher begins co-teaching, and prior work assessing teacher quality demonstrates the importance of controlling for student assignment (Chetty et al., 2014). When using only one year of lagged test scores for the cubic polynomial rather than two years, the results are similar (Figure A.5).

Cohorts with four and five years of experience before co-teaching follow a different pattern than those who start after one to three years. For these cohorts, however, those who do not co-teach may not be an accurate counter-factual. Figure A.7 shows evidence of a negative pre-trend and Figure 1 shows that general education teachers are relatively less likely to begin co-teaching after so many years. I omit these cohorts to refine the analysis; however, the inclusion of cohorts that take more time to begin co-teaching does slightly mitigate the estimated benefits (Figure A.8).

Varying the method by which I weight events does impact the estimated magnitude of the returns to co-teaching. Still, it is unclear whether this bias is affecting my estimates positively or negatively (Table A.1). Simply using the implicit weighting produced by the stacked difference-in-difference estimates a difference of 0.0307σ. Weighting every event evenly reduces this estimate to 0.0171σ. Weighting each event by the number of observations within the event increases the estimated difference to 0.0545σ. The number of observations within each event varies from 20 to $7,675$, so weighting events evenly appears to be an unusual choice, but doing so would reduce the estimated effects.

5 Conclusion

Co-teaching creates lasting improvements in teacher productivity. Teachers who collaborate with experienced colleagues (16+ years) and then return to solo instruction perform 0.10σ better on student achievement measures. The effect persists, demonstrating genuine skill transfer rather than temporary assistance.

This answers a basic question in personnel economics: Do workers learn from collaboration or just benefit while collaborating? Previous studies could not distinguish between these explanations because they only measured effects during partnerships. By following the same teachers before, during, and after collaboration, I show that learning occurs.

The mechanism is knowledge transfer from experienced to inexperienced teachers. Partners with fewer than 16 years of experience produce no lasting benefits. This suggests that tacit knowledge accumulates slowly but transfers quickly through interaction.

Three implications follow for organizations. First, brief collaborations are sufficient—the high turnover rate of 60% after one year does not waste learning opportunities. Second, pairing novices together foregoes knowledge transfer. Third, collaboration works best after workers develop basic skills independently.

The findings extend beyond education. Any workplace where experienced workers can transfer tacit knowledge through collaboration may benefit from strategic pairing programs. The institutional features that drive co-teaching assignment—special education enrollment and scheduling constraints—exist in other settings where worker pairing depends on operational needs rather than quality matching.

Future research should identify what experienced workers transfer and how organizations can design collaboration to maximize learning. The persistence of effects after partnership ends shows that workplace collaboration can permanently upgrade individual capabilities, not just provide temporary productivity gains.

References

Andersen, S. C., Beuchert, L., Nielsen, H. S., and Thomsen, M. K. (2020). The effect of teacher's aides in the classroom: Evidence from a randomized trial. *Journal of the European Economic Association*, 18(1):469–505.

Arrow, K. J. (1962). The economic implications of learning by doing. *The Review of Economic Studies*, 29(3):155–173.

Austin, V. L. (2001). Teachers' beliefs about co-teaching. *Remedial and Special Education*, 22(4):245–255.

Azoulay, P., Graff Zivin, J. S., and Wang, J. (2010). Superstar extinction. *The Quarterly Journal of Economics*, 125(2):549–589.

Ballis, B. and Heath, K. (2021). The long-run impacts of special education. *American Economic Journal: Economic Policy*, 13(4):72–111.

Battu, H., Belfield, C. R., and Sloane, P. J. (2003). Human capital spillovers within the workplace: Evidence for great britain. *Oxford Bulletin of Economics and Statistics*, 65(5):575–594.

Bessette, H. J. (2008). Using students' drawings to elicit general and special educators' perceptions of co-teaching. *Teaching and Teacher Education*, 24(5):1376–1396.

Cengiz, D., Dube, A., Lindner, A., and Zipperer, B. (2019). The effect of minimum wages on low-wage jobs. *The Quarterly Journal of Economics*, 134(3):1405–1454.

Chen, Y. (2021). Team-specific human capital and team performance: evidence from doctors. *American Economic Review*, 111(12):3923–3962.

Chetty, R., Friedman, J. N., and Rockoff, J. E. (2014). Measuring the impacts of teachers i: Evaluating bias in teacher value-added estimates. *American economic review*, 104(9):2593–2632.

Conderman, G. and Hedin, L. R. (2014). Co-teaching with strategy instruction. *Intervention in School and Clinic*, 49(3):156–163.

Cook, S. C., McDuffie-Landrum, K. A., Oshita, L., and Cook, B. G. (2017). Co-teaching for students with disabilities: A critical and updated analysis of the empirical literature. In *Handbook of Special Education*, pages 233–248. Routledge.

De Backer, L., Schelfhout, W., Simons, M., and Vandervieren, E. (2023). Student teachers' peer team teaching experiences from a quantitative perspective: Perceptions, profiles and transition probabilities. *Teaching and Teacher Education*, 135:104361.

De Grip, A. and Sauermann, J. (2012). The effects of training on own and co-worker productivity: Evidence from a field experiment. *The Economic Journal*, 122(560):376–399.

Goodman-Bacon, A. (2021). Difference-in-differences with variation in treatment timing. *Journal of Econometrics*, 225(2):254–277.

Guryan, J., Kroft, K., and Notowidigdo, M. J. (2009). Peer effects in the workplace: Evidence from random groupings in professional golf tournaments. *American Economic Journal: Applied Economics*, 1(4):34–68.

Haggag, K., McManus, B., and Paci, G. (2017). Learning by driving: Productivity improvements by new york city taxi drivers. *American Economic Journal: Applied Economics*, 9(1):70–95.

Ichino, A. and Maggi, G. (2000). Work environment and individual background: Explaining regional shirking differentials in a large italian firm. *The Quarterly Journal of Economics*, 115(3):1057–1090.

Jackson, C. K. and Bruegmann, E. (2009). Teaching students and teaching each other: The importance of peer learning for teachers. *American Economic Journal: Applied Economics*, 1(4):85–108.

Jones, N., Kaler, L., Markham, J., Senese, J., and Winters, M. A. (2025). Service delivery models: Impacts for students with and without disabilities. *Educational Researcher*, 54(3):141–152.

Jones, N. and Winters, M. A. (2024). Are two teachers better than one?: The effect of coteaching on students with and without disabilities. *Journal of Human Resources*, 59(4):1180–1206.

Kraft, M. A., Blazar, D., and Hogan, D. (2018). The effect of teacher coaching on instruction and achievement: A meta-analysis of the causal evidence. *Review of Educational Research*, 88(4):547–588.

Mas, A. and Moretti, E. (2009). Peers at work. *American Economic Review*, 99(1):112–145.

Nix, E. (2020). Learning spillovers in the firm. Technical report, Working paper.

Ost, B. (2014). How do teachers improve? the relative importance of specific and general human capital. *American Economic Journal: Applied Economics*, 6(2):127–151.

Papay, J. P. and Kraft, M. A. (2015). Productivity returns to experience in the teacher labor market: Methodological challenges and new evidence on long-term career improvement. *Journal of Public Economics*, 130:105–119.

Shaw, K. and Lazear, E. P. (2008). Tenure and output. *Labour Economics*, 15(4):704–723.

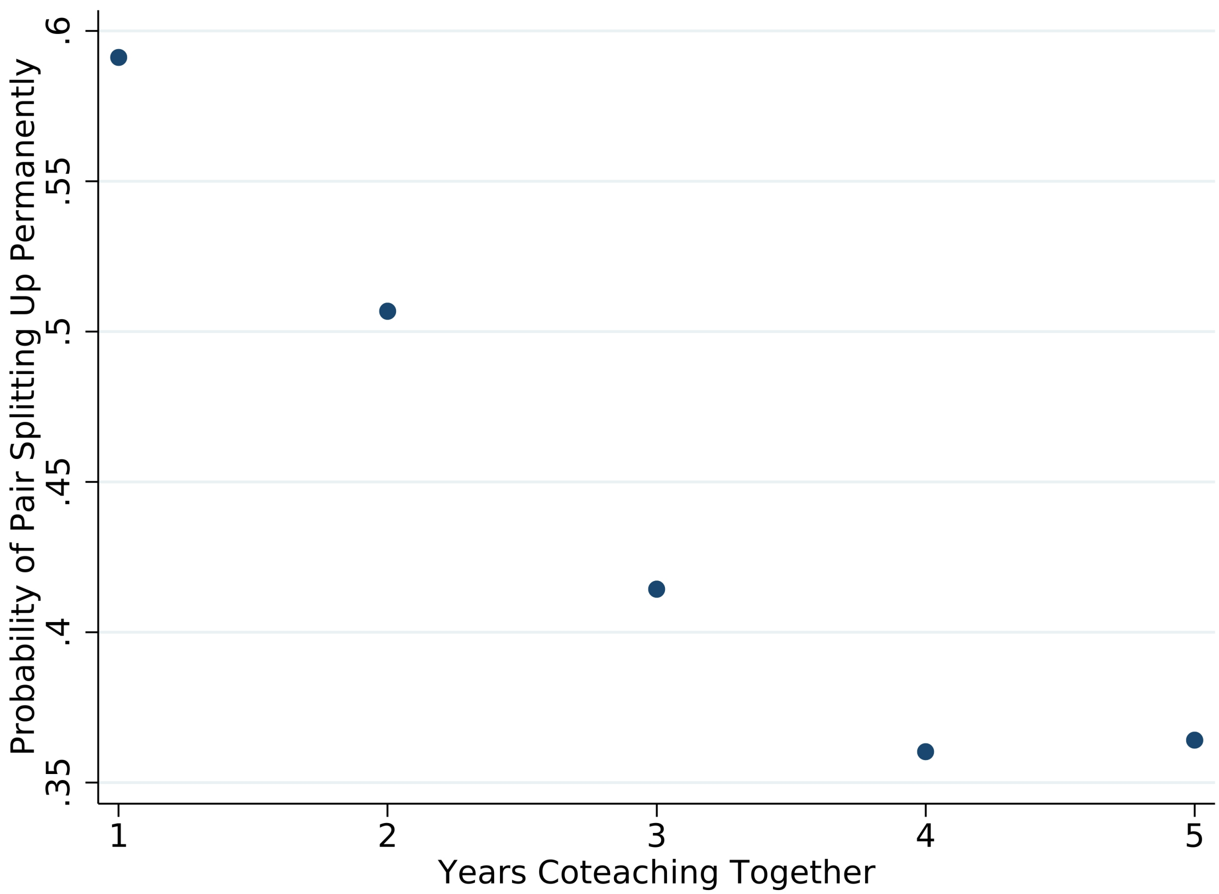
Thompson, P. (2010). Learning by doing. *Handbook of the Economics of Innovation*, 1:429–476.

Thompson, P. (2012). The relationship between unit cost and cumulative quantity and the evidence for organizational learning-by-doing. *Journal of Economic Perspectives*, 26(3):203–224.

Wing, C., Freedman, S. M., and Hollingsworth, A. (2024). Stacked difference-in-differences. Technical report, National Bureau of Economic Research.

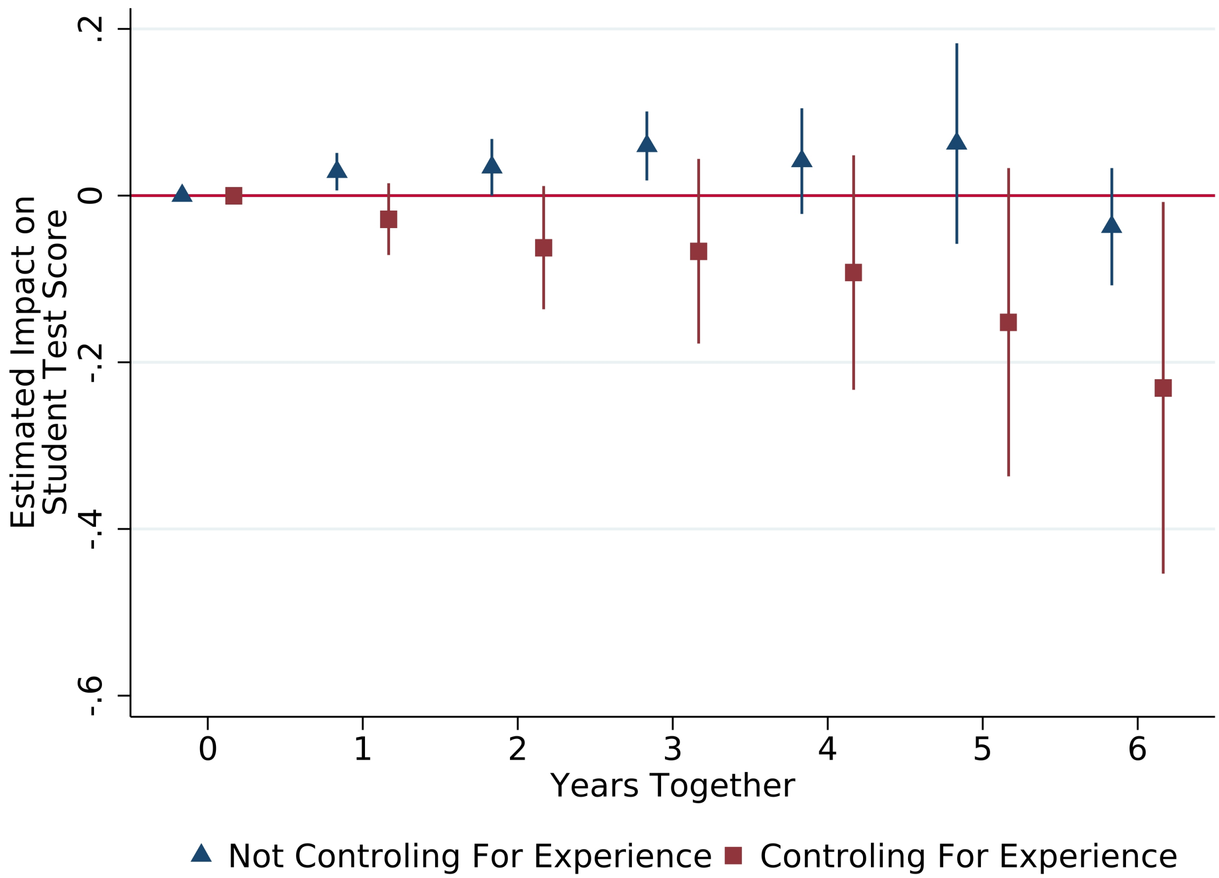
A Appendix

Figure A.1: Hazard Rate for Permanent Co-teaching Pair Separation



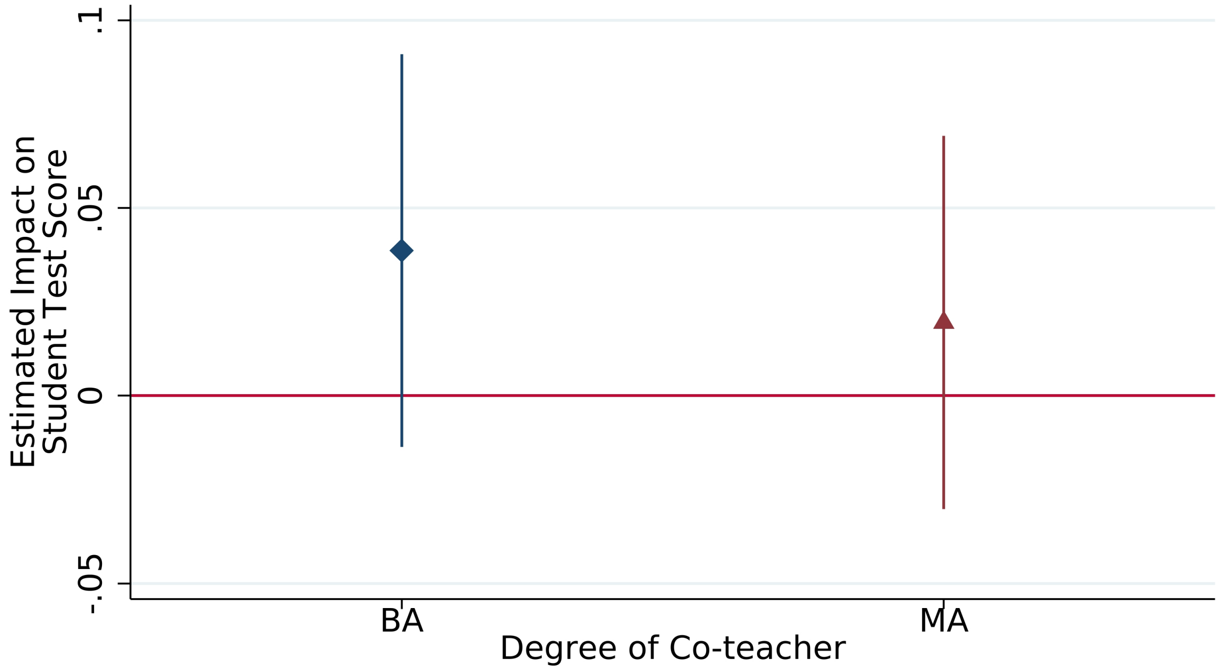
This figure shows the probability that co-teaching pairs never work together again (permanent separation) rather than temporary separation followed by potential reunion.

Figure A.2: Return To Experience in a Pair (Estimated in One Step)



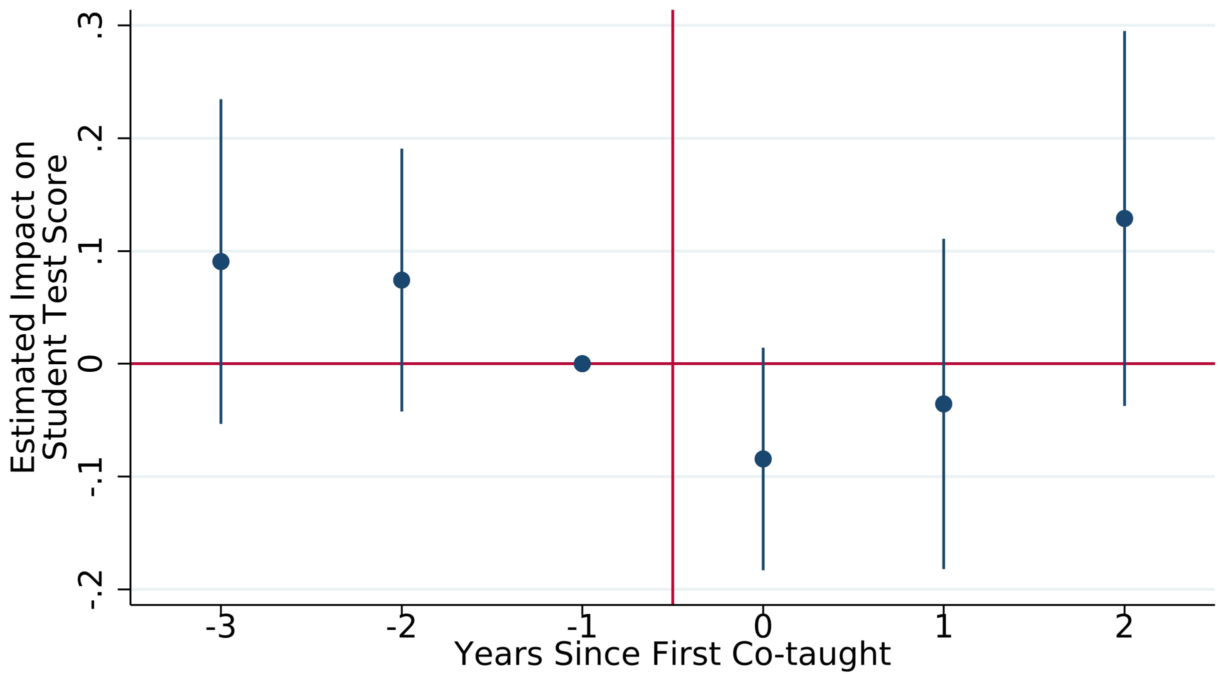
This figure presents the same analysis as Figure 8 but using a single-step estimation method rather than the two-step approach, as a robustness check on the methodology.

Figure A.3: Difference in Student Test Scores By Time Since First Co-taught and Co-teacher Degree



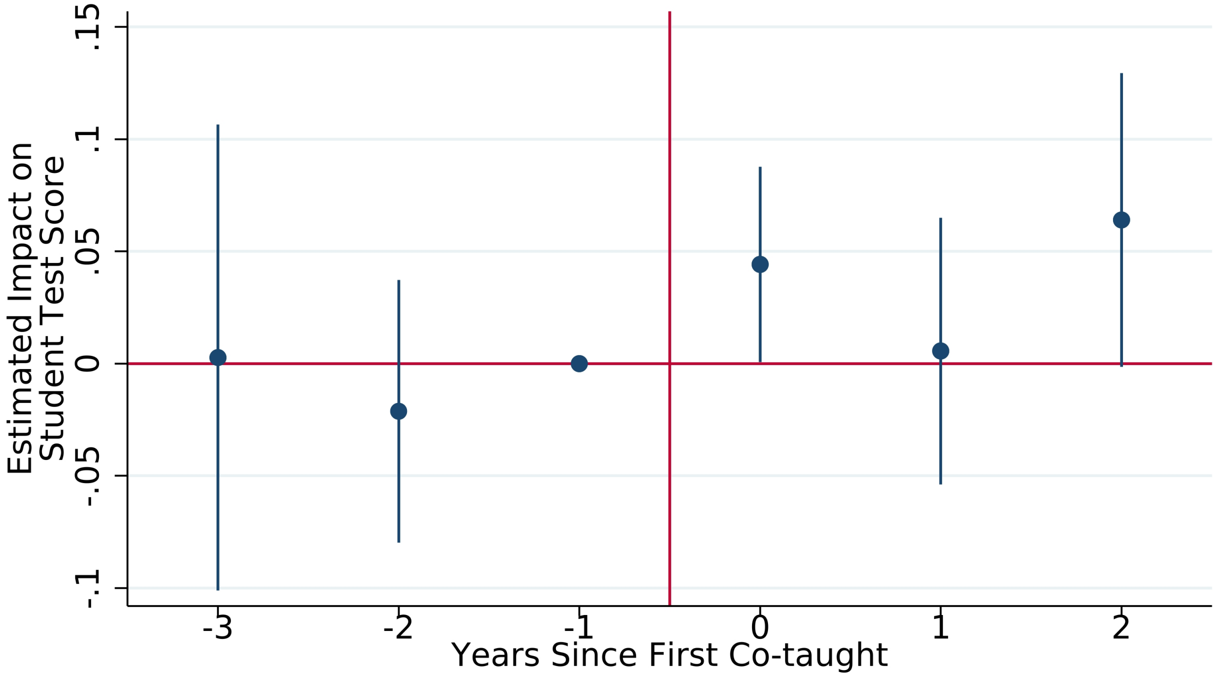
This figure examines whether the educational attainment (degree level) of the co-teaching partner affects the magnitude of developmental benefits from co-teaching.

Figure A.4: Difference in Student Test Scores By Time Since First Co-taught (No Lagged Test Scores)



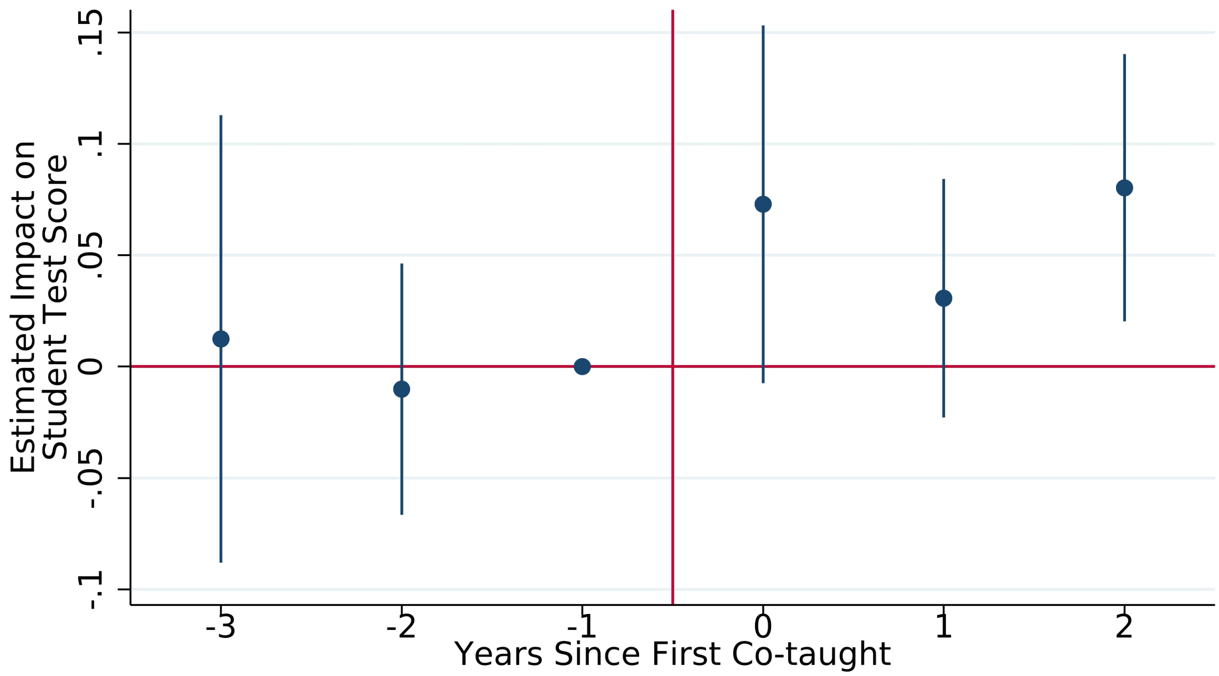
This robustness check presents the main event study results without including student prior test score controls in the regression specification.

Figure A.5: Difference in Student Test Scores By Time Since First Co-taught (One Year of Lagged Test Scores)



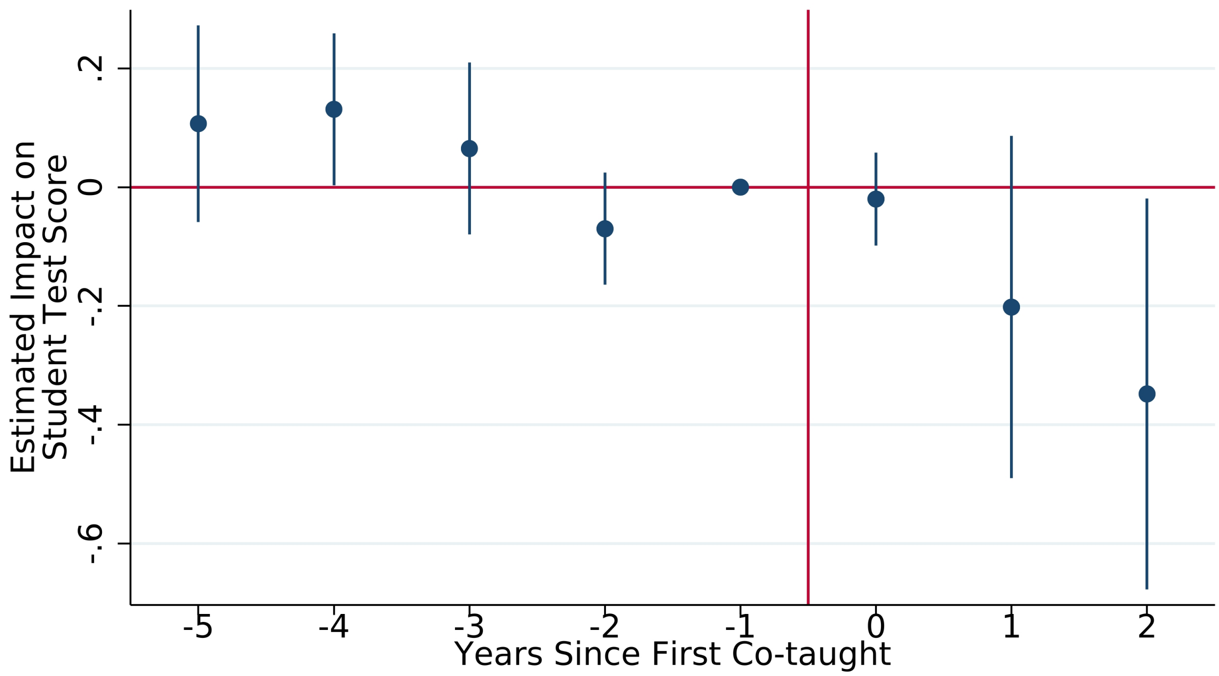
This robustness check uses only one year of prior student test scores in the control polynomial rather than the two years used in the main specification.

Figure A.6: Difference in Student Test Scores By Time Since First Co-taught (Controlling For Current Co-teaching)



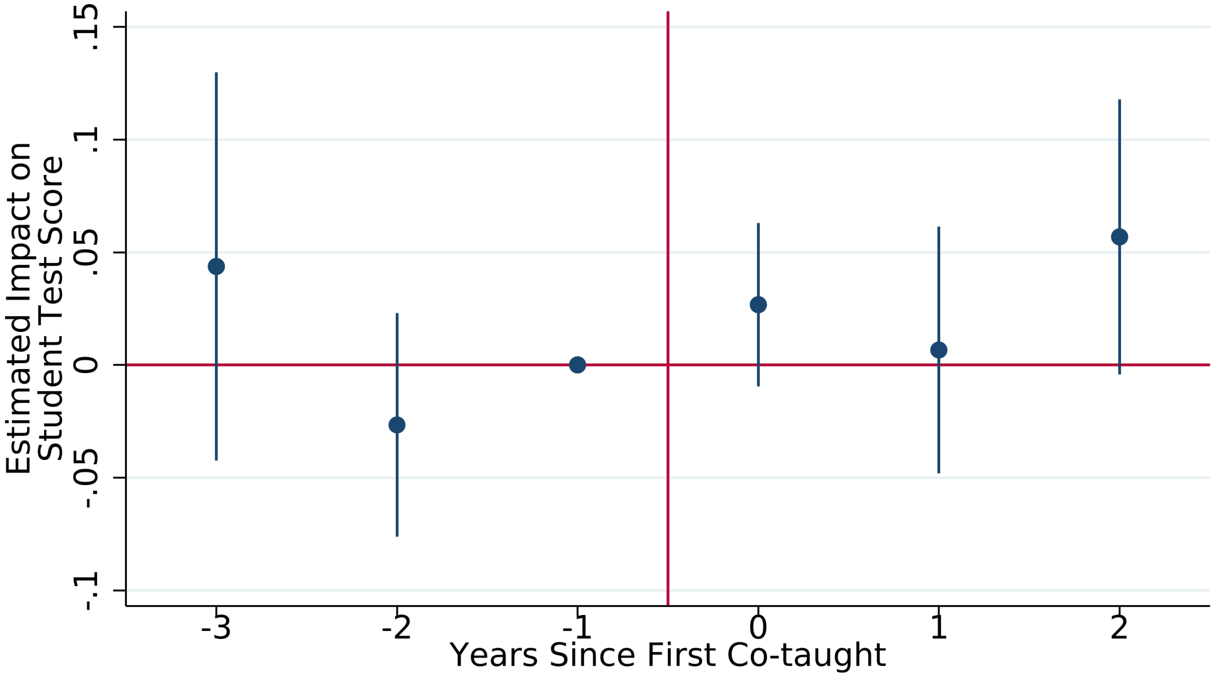
This specification adds indicator variables for whether teachers are currently co-teaching to separate persistent effects from contemporaneous collaboration benefits.

Figure A.7: Difference in Student Test Scores By Time Since First Co-taught (Only Using More Than Three Years of Prior Experience)



This figure restricts the sample to teachers who had more than three years of solo teaching experience before first co-teaching, testing the sensitivity of results to sample composition.

Figure A.8: Difference in Student Test Scores By Time Since First Co-taught (One or More Years of Prior Experience)



This figure expands the sample to include all teachers with at least one year of prior experience before co-teaching, rather than restricting to 1-3 years as in the main specification.

Table A.1: Difference in Difference Estimates Weighting Events Differently

Dependent Variable (DV):	Student Test Scores		
Method	Implicit Weights	Even Weights	Observation Weights
Estimated Difference	0.0307	0.0171	0.0545

This table compares the main difference-in-differences estimate under three different weighting schemes: implicit weights from the stacked methodology, equal weights across events, and weights proportional to the number of observations per event.