Supplement to "Can Teaching Be Taught? Improving Teachers' Pedagogical Skills at Scale in Rural Peru"

Juan F. Castro	Paul Glewwe	Alexandra Heredia-Mayo	Stephanie Majerowicz	Ricardo Montero
Universidad del Pacifico	University of Minnesota	Universidad del Pacifico	Universidad de los Andes	University of Minnesota

October, 2024

Appendix A: Additional Tables

Table A1. Descriptive Statistics

	Public Primary Schools		Randomized Expansion Sample		
	All	Monolingual Multigrade	All	Test Score Sample	Pedagogical Sample
	(1)	(2)	(3)	(4)	(5)
Math score (2015)	550.92	526.70	512.17	513.29	497.49
	(93.37)	(97.67)	(93.45)	(92.16)	(92.24)
Language score (2015)	552.39	529.19	519.02	519.47	511.11
	(69.62)	(69.13)	(66.78)	(65.70)	(66.40)
Number of students	88.07	28.78	28.66	45.96	29.13
	(161.70)	(24.06)	(23.48)	(25.21)	(23.83)
Number of teachers	4.74	1.91	1.90	2.57	1.89
	(6.79)	(1.18)	(1.14)	(1.23)	(1.15)
Number of sections (classes)	6.82	5.33	5.36	5.83	5.34
	(4.81)	(1.26)	(1.19)	(0.81)	(1.21)
Teacher-Student Ratio	0.09	0.10	0.10	0.06	0.09
	(0.09)	(0.09)	(0.07)	(0.03)	(0.07)
Rurality	1.90	2.25	2.35	2.28	2.43
	(1.05)	(0.79)	(0.78)	(0.83)	(0.72)
% students indigenous mother tongue	23.70	3.22	5.02	4.27	2.36
	(41.60)	(16.95)	(20.86)	(18.86)	(14.55)
Poverty rates	55.16	55.67	56.79	59.59	57.76
	(22.96)	(20.83)	(22.83)	(22.33)	(21.24)
Ceiling material	5.90	5.66	5.64	5.77	5.59
	(1.63)	(1.41)	(1.42)	(1.48)	(1.29)
Wall material	6.21	6.10	6.03	6.09	6.03
	(1.20)	(1.27)	(1.37)	(1.36)	(1.36)
Floor material	2.81	2.86	2.83	2.85	2.89
	(0.78)	(0.71)	(0.76)	(0.71)	(0.74)
% teachers with degree	0.97	0.98	0.98	0.96	0.98
	(0.11)	(0.10)	(0.10)	(0.12)	(0.09)
Internet access	0.20	0.08	0.08	0.10	0.08
	(0.40)	(0.27)	(0.28)	(0.30)	(0.27)
Receives textbooks	0.77	0.76	0.73	0.74	0.69
	(0.42)	(0.42)	(0.45)	(0.44)	(0.46)
Receives workbooks	0.70	0.69	0.66	0.69	0.62
	(0.46)	(0.46)	(0.47)	(0.46)	(0.49)
School-day length	8.16	8.22	8.16	8.15	8.16
	(0.86)	(0.98)	(0.83)	(0.88)	(0.90)
Electricity	0.66	0.54	0.53	0.54	0.53
	(0.47)	(0.50)	(0.50)	(0.50)	(0.50)
Water	0.59	0.50	0.48	0.53	0.49
	(0.49)	(0.50)	(0.50)	(0.50)	(0.50)
Sanitation	0.25	0.13	0.12	0.14	0.13
	(0.44)	(0.33)	(0.33)	(0.35)	(0.34)
Computers per Student	0.63	0.39	0.43	0.45	0.25
	(4.24)	(2.50)	(2.58)	(2.87)	(0.96)
Schools (with baseline data)	29,419	14,467	6,204	2,563	340
Schools (full sample)	30,530	14,467	6,218	2,567	340

Note: This table shows the descriptive statistics for the experimental sample compared to all Peruvian public primary schools and multigrade schools. Column 1 shows the average characteristics for all public primary schools in Peru, while Column 2 restricts the sample to monolingual multigrade primary schools, the target population of the coaching program. Columns 3 and 4 show descriptive statistics for the experimental sample, with Column 3 including all schools in the sample and Column 4 restricting the sample to the subset of schools with test scores in 2016. Finally, Column 5 shows the subsample for which we measure the pedagogical skills of teachers, which is missing data from 24 very remote schools that could not be surveyed. Rurality is a categorical variable that takes values 0 for Urban schools, and 1, 2 and 3 for increasingly rural schools. Ceiling, Wall and Floor Materials are categorical variables that take values up to 7, with higher values implying better materials. Standard deviations are shown in parentheses.

Table A2. Descriptive Statistics for Outcomes - Baseline

	Public Primary Schools			Experimental Sample	
	All	Not in Sample	In Sample	Control	Treated
Math					
Rasch	550.92	559.39	512.17	510.29	513.46
Level 1 (beginning)	0.40	0.37	0.54	0.55	0.53
Level 2 (in process)	0.37	0.39	0.32	0.31	0.33
Level 3 (satisfactory)	0.23	0.25	0.14	0.14	0.14
Reading					
Rasch	552.39	559.69	519.02	516.17	520.98
Level 1 (beginning)	0.14	0.12	0.23	0.24	0.23
Level 2 (in process)	0.54	0.53	0.58	0.58	0.58
Level 3 (satisfactory)	0.32	0.35	0.19	0.18	0.20
N	14,595	11,977	2,618	1,067	1,551

Note: This table presents the descriptive statistics of the ECE score in baseline (2015) for all public primary schools in Peru, as well as for the experimental sample. Only schools with more than 5 students in the tested grade take the test. The ECE scores are reported both as levels of subject mastery (beginning, in process, and satisfactory) and as a Rasch score with a nationally standardized mean of 500 and standard deviation of 100.

Table A.3. Robustness Checks on Treatment Effects on Students' Standardized Test Scores

	Preferred Specification (1)	Without Controls (2)	Region Fixed Effects (3)	School Fixed Effects (4)
Panel A. Mathematics Test Scores 2016				
Treatment	0.106	0.076	0.090	0.105
	(0.034)	(0.036)	(0.035)	(0.035)
Observations	22198	22198	22198	131668
Schools	2547	2547	2547	2547
R^2	0.142	0.133	0.089	0.282
Panel B. Mathematics Test Scores 2018				
Treatment	0.114	0.080	0.091	0.114
	(0.033)	(0.034)	(0.034)	(0.029)
Observations	18261	18261	18261	149938
Schools	2053	2053	2053	2547
R^2	0.182	0.176	0.123	0.273
Panel C. Reading Comprehension Test Scores 2016				
Treatment	0.075	0.041	0.062	0.078
	(0.032)	(0.033)	(0.033)	(0.032)
Observations	22199	22199	22199	131678
Schools	2547	2547	2547	2547
R^2	0.162	0.152	0.116	0.308
Panel D. Reading Comprehension Test Scores 2018				
Treatment	0.100	0.068	0.082	0.096
	(0.031)	(0.032)	(0.032)	(0.027)
Observations	18275	18275	18275	149962
Schools	2053	2053	2053	2547
R^2	0.168	0.162	0.117	0.296
Fixed Effects	School District	School District	Region	School
Controls	Yes	No	Yes	No

Note: This table shows the average treatment effect of the coaching program on standardized student test scores for a number of different specifications for robustness. Panels A and C show the effect after one year of treatment in 2016, while Panels B and D show the effects after three years of treatment in 2018. Column 1 shows the preferred specification that we showed in the main text which include school district fixed effects and two unbalanced controls (number of students and teachers). Column 2 removes the controls, and Column 3 includes region fixed effects instead of school fixed effects. Finally, Column 4 takes advantage of the panel data from 2010-2018 in order to include school fixed effects, without any additional controls. All results are over standardized exam scores and can be interpreted as standard deviations. Regressions are run at the student level, with robust standard errors clustered by school presented in parentheses.

Table A4. Selective Attrition Test:
Regression of Treatment Status on Pre-treatment Characteristics

	(1) All teachers	(2) Observed teachers at the end of 2017	(3) Non-observed teachers at the end of 2017
Age	0.001	0.003	-0.000
	(0.003)	(0.003)	(0.005)
Male	-0.001	0.022	-0.107
	(0.038)	(0.051)	(0.090)
Education			
Higher education	-0.059	0.069	-0.302
	(0.184)	(0.095)	(0.353)
Postgraduate	-0.102	0.027	---
	(0.194)	(0.127)	
Teacher career			
Contract	0.105	0.059	0.158
	(0.066)	(0.082)	(0.130)
2nd scale	-0.017	0.024	-0.066
	(0.059)	(0.075)	(0.117)
3rd scale	0.110	0.088	0.180
	(0.072)	(0.092)	(0.141)
4th scale	-0.015	0.048	-0.213
	(0.091)	(0.123)	(0.258)
5th scale	-0.037	0.040	---
	(0.159)	(0.197)	
Joint Significance Test - pvalue	0.506	0.941	0.136
N	646	444	202
R-squared	0.212	0.244	0.440

Note: All regressions include UGEL fixed effects. Standard errors clustered at the school level in parentheses.

Table A5: Move Decisions of Sample 1 Teachers after One Year

	Move decision (observed)	Likers	Movers	Remainers	Dislikers	Row Sum (observed)	Equation
Assigned to APM school	move to APM school	$p^L(1-\sigma)$	$p^M\mu$	0	0	0.1207	(A1)
	move to non-APM school	0	$p^M(1-\mu)$	0	p^D	0.2470	(A2)
	stayed in same school	$p^L\sigma$	0	p^R	0	0.6323	(A3)
Assigned to non-APM school	move to APM school	p^L	$p^M\mu$	0	0	0.0965	(A4)
	move to non-APM school	0	$p^M(1-\mu)$	0	$p^D(1-\nu)$	0.2793	(A5)
	stayed in same school	0	0	p^R	$p^D\nu$	0.6242	(A6)

The fraction of likers who stay in the same school is σ, and the fraction of dislikers who stay in the same school is ν. See Appendix C for how these relationships are derived.

Table A6 Pedagogical Skills Heterogeneous Treatment Effects (Sample 1)

	(1)	(2)	(3)	(4)	(5)
Treatment	0.300	0.196	0.301	0.260	0.223
	(0.097)	(0.229)	(0.111)	(0.152)	(0.140)
Experience	0.000	-0.003	0.000	-0.000	-0.000
	(0.009)	(0.011)	(0.009)	(0.009)	(0.009)
Contract Teacher	0.139	0.141	0.142	0.141	0.127
	(0.155)	(0.156)	(0.216)	(0.156)	(0.155)
Magisterial Level	0.109	0.110	0.109	0.096	0.108
	(0.044)	(0.044)	(0.044)	(0.058)	(0.044)
Sex (Men=1)	-0.300	-0.301	-0.300	-0.299	-0.382
	(0.095)	(0.095)	(0.095)	(0.095)	(0.140)
Age	-0.028	-0.027	-0.027	-0.027	-0.028
	(0.009)	(0.009)	(0.009)	(0.009)	(0.009)
Treatment #Experience		0.005			
		(0.011)			
Treatment #Contract			-0.007		
			(0.235)		
Treatment #M. Level				0.025	
				(0.078)	
Treatment #Sex					0.169
					(0.180)
R^2	0.37	0.37	0.37	0.37	0.37
N	455	455	455	455	455

Note: All regressions include UGEL fixed effects. Standard errors clustered at the school level are reported in parenthesis.

Table A7
Selective Attrition Test: Regression of Observed Indicator on Treatment Status, Pre-treatment Characteristics, and Interactions of Both Variables

	Observed in Year 2 (Yes=1)	
Treatment	-0.707	(0.408)
Treatment x Age	0.007	(0.005)
Treatment x Male	-0.040	(0.082)
Treatment x Higher Education	0.403	(0.338)
Treatment x Postgraduate	0.271	(0.350)
Treatment x Contract	-0.029	(0.115)
Treatment x 2nd scale	0.075	(0.100)
Treatment x 3rd scale	-0.054	(0.121)
Treatment x 4th scale	-0.007	(0.189)
Treatment x 5th scale	0.318	(0.247)
Age	-0.003	(0.003)
Male	0.008	(0.058)
Higher education	0.050	(0.295)
Postgraduate	0.344	(0.303)
Contract Teacher	-0.154	(0.086)
2nd scale	-0.068	(0.070)
3rd scale	-0.023	(0.082)
4th scale	-0.077	(0.154)
5th scale	0.016	(0.121)
Joint Significance Test: p-value (treatment and interactions between treatment and pre-treatment characteristics)	0.372	
N	646	
R-squared	0.26	

Note: The regression includes UGEL fixed effects. Standard errors clustered at the school level in parentheses.

Table A8: Teacher Skills Estimates using Regional Fixed Effects

	Sample 1		Sample 2
	(1)	(2)	(3)
Treatment	0.175	0.196	0.222
	(0.092)	(0.090)	(0.111)
Experience		-0.000	
		(0.008)	
Contract Teacher		0.088	
		(0.136)	
Teacher Career Level		0.083	
		(0.040)	
Sex (men=1)		-0.277	
		(0.086)	
Age		-0.024	
		(0.008)	
R^2	0.15	0.22	0.18
N	455	455	347

All regressions include regional fixed effects. Standard errors clustered at the school level are in parentheses.

Table A9: Regression of 2016 Students' ECE Scores on Teacher Pedagogical Practices Index (2016)

	ECE Scores			
	Math		Reading	
	(1)	(2)	(3)	(4)
Teacher Skill Index	0.042	0.060	0.043	0.072
	(0.015)	(0.042)	(0.015)	(0.042)
R^2	0.19	0.19	0.24	0.24
N	1,487	298	1,487	298
Grades	All	2nd	All	2nd

Note: These regressions measure the correlation between the teacher skill index and students' test scores in the schools those teachers are. The measurement of teachers' practices uses the same rubrics and methodology as in our sample, corresponds to a random sample of primary schools not related to APM or its randomized expansion. Columns (1) and (3) include all observed teachers in the school, while columns (2) and (4) only uses information on teachers in second grade (which is the grade evaluated in the ECE). These observational results support the idea that the skills measured are correlated with student success in standardized tests. The point estimates are higher when sample is limited to teachers in the corresponding grade, but precision is lost due to a much smaller sample. All regressions include regional fixed effects. Huber-White heteroscedasticity robust standard errors in parentheses.

Table A10: Differences in Teacher Test Scores across Movement Status

	Teachers who move into randomized sample					Teachers who move out of randomized sample				
	Into Control Schools		Into Treatment Schools		Pairwise t-test	Out of Control Schools		Out of Treatment Schools		Pairwise t-test
	N	Mean/ (SD)	N	Mean/ (SD)	Coef/ (P-value)	N	Mean/ (SD)	N	Mean/ (SD)	Coef/ (P-value)
Standardized teacher exam score in 2014 or 2015		(1)		(2)	(1) - (2)		(3)		(4)	(3) - (4)
Year 2016	1099	-0.130	1523	-0.112	-0.018	1630	0.014	2351	0.046	-0.032
		(0.936)		(0.929)	(0.635)		(0.948)		(0.992)	(0.311)
Year 2017	1332	-0.313	1823	-0.298	-0.015	1413	-0.339	2009	-0.314	-0.025
		(0.897)		(0.896)	(0.636)		(0.899)		(0.865)	(0.412)
Year 2018	1019	-0.323	1395	-0.355	0.032	1570	-0.191	2213	-0.155	-0.036
		(0.942)		(0.920)	(0.396)		(0.934)		(0.948)	(0.249)

Note: This table shows the difference in the test scores obtained by teachers in exams by whether they move into or out of the 6,218 randomized sample schools. Teacher exam scores come from tests designed to measure teachers' pedagogical skills and content knowledge taken in order to get into the civil service track or to get a promotion within it. These test were all taken in 2014 and 2015 prior to treatment and scares are standardized over the entire sample of teachers with test scores. Columns (1) and (2) show the average score of teachers who moved into schools in the randomized sample, regardless of where they were coming from, while columns (3) and (4) show the average score for those who moved out of the randomize sample school, regardless of where they moved to. The columns show the mean and standard deviation for the test scores of those teachers, and the following column shows the coefficient and the p-value on the pairwise t-test of the difference. For columns (1) and (2), moving in year 2016 means that teachers moved from 2015 to 2016 (and therefore were in an APM school in 2016) while in columns (3) and (4) show teachers who were in APM schools in that year and moved at the end of the year. Therefore, in both cases, teachers were in the APM schools in the year shown.

Appendix B: Definitions and Derivations for All the Treatment Effects for APM Program

I. The Basic Set-Up

The test score of student i at the end of year t (s_i^t) is determined by his or her test score at the end of the previous year (s_i^{t-1}) and the skills of the teacher that this student had in the current year (y_j^t), where j denotes the particular teacher that student i had in year t:

$$s_i^t = \sigma s_i^{t-1} + \pi y_j^t \qquad \text{(B.1)}$$

where σ is the impact of the previous year's skills and π is the impact of teacher skill (which, for simplicity, is assumed to be unidimensional).

Schools are randomly assigned in year 1 to be either APM schools (R = 1) or non-APM schools (R = 0). This school assignment does not change over time, and R always refers to assignment of schools.

The skill of teacher j at the end of year t, y_j^t, is assumed to be a linear function of his or her skills in the previous year (y_j^{t-1}), the skill gained from one more year of experience (λ_i), and whether he or she is treated in the current year (T_j^t). The treatment impact, δ^k, can vary by the type of teacher (remainer (R), liker (L), dislike (D) and mover (M)). Depreciation of teaching skills that are unrelated to the coaching program can be included in λ_i Equation (B.2) provides the general expression of y_j^t for year t:

$$y_j^t = y_j^{t-1} + \lambda_j + \delta^k T_j^t, \quad \text{for } k = R, L, D, M \qquad \text{(B.2)}$$

Applying (B.2) to year 1 yields the following expression for the skills of teacher j in that year:

$$y_j^1 = y_j^0 + \lambda_j + \delta^k T_j^1, \quad \text{for } k = R, L, D, M \qquad \text{(B.3)}$$

$$= \theta_j^1 + \delta^k T_j^1$$

where θ_j^1 is convenient notation for $y_j^0 + \lambda_j$.[1]

For year 2, we need to allow for interaction effects of treatments in different years; for example, the impact of a second year of exposure to the program on teachers' skills could be smaller than the impact for the first year of exposure. The equation for y_j^2 is:

$$y_j^2 = y_j^1 + \lambda_j + \delta^k T_j^2 + \gamma_{1,2}^k T_j^1 T_j^2, \quad \text{for } k = R, L, D, M \qquad \text{(B.4)}$$

$$= \theta_j^1 + \delta^k T_j^1 + \lambda_j + \delta^k T_j^2 + \gamma_{1,2}^k T_j^1 T_j^2$$

$$= \theta_j^2 + \delta^k (T_j^1 + T_j^2) + \gamma_{1,2}^k T_j^1 T_j^2$$

where θ_j^2 is convenient notation for $\theta_j^1 + \lambda_i = y_j^0 + 2\lambda_j$. If the impact of a second year of exposure to the program is smaller than that of the first year of exposure, then $\gamma_{1,2}^k$ would be < 0. Note also that $\gamma_{1,2}^k$ can include depreciation of teacher skills produced by the program.

For year 3, we need to allow for further interaction effects. The equation for y_j^3 is:

$$y_j^3 = y_j^2 + \lambda_j + \delta^k T_j^3 + \gamma_{1,2}^k (T_j^1 T_j^3 + T_j^2 T_j^3) + \gamma_{1,2,3}^k T_j^1 T_j^2 T_j^3, \quad \text{for } k = R, L, D, M \qquad \text{(B.5)}$$

$$= \theta_j^2 + \delta^k (T_j^1 + T_j^2) + \gamma_{1,2}^k T_j^1 T_j^2 + \lambda_j + \delta^k T_j^3 + \gamma_{1,2}^k (T_j^1 T_j^3 + T_j^2 T_j^3) + \gamma_{1,2,3}^k T_j^1 T_j^2 T_j^3$$

$$= \theta_j^3 + \delta^k (T_j^1 + T_j^2 + T_j^3) + \gamma_{1,2}^k (T_j^1 T_j^2 + T_j^1 T_j^3 + T_j^2 T_j^3) + \gamma_{1,2,3}^k T_j^1 T_j^2 T_j^3$$

where θ_j^3 is convenient notation for $\theta_j^2 + \lambda_i = y_j^0 + 3\lambda_j$. Note that the interaction effect for any combination of two years of training is assumed to be the same, regardless of whether the two years are year 1 and year 2, or year 1 and year 3, or year 2 and year 3. Allowing for different interaction effects for each possible pair of years would do little beyond complicating the notation.

[1] Note that "R" is used in two different ways, to indicate randomization of a school and to denote "remainer" teachers. When "R" is in normal size text, it indicates the former, and when it is a superscript it indicates the latter.

II. Definitions of Treatment Effects for Teachers' Skills (denoted by y)

We first define three standard treatment effects for teacher skills, then we explain how to define two of these three definitions when the focus is not on set of teachers but on the teachers in a set of schools, in a context where teachers often move between schools.

1. **Average Treatment Effect (ATE).**

We begin with the average treatment effect (ATE). This is defined as what happens when all teachers are treated, and the counterfactual is that no teachers are treated (the program does not exist). In the notation below, superscripts on y refer to number of years of the program, subscripts on y refer to the potential outcome (1 = treated, 0 = not treated), and the p^k terms (k = R, L, D and M) refer to the proportion of teachers in the population who are of type k. ATE for teacher skills in year t is defined as:

$$ATE_{tchr}(t) \equiv E[y_1^t - y_0^t] = E[y_1^t] - E[y^t | \text{ No program exists}] \quad (B.6)$$

Applying this definition to years 1, 2 and 3 yields the following more specific definitions:

$$ATE_{tchr}(1) \equiv E[y_1^1 - y_0^1] = \bar{\delta}, \text{ where } \bar{\delta} = \delta^R p^R + \delta^L p^L + \delta^D p^D + \delta^M p^M \quad (B.7)$$

$$ATE_{tchr}(2) \equiv E[y_1^2 - y_0^2] = 2\bar{\delta} + \bar{\gamma}_{1,2}, \text{ where } \bar{\gamma}_{1,2} = \gamma_{1,2}^R p^R + \gamma_{1,2}^L p^L + \gamma_{1,2}^D p^D + \gamma_{1,2}^M p^M \quad (B.8)$$

$$ATE_{tchr}(3) \equiv E[y_1^3 - y_0^3] = 3\bar{\delta} + 3\bar{\gamma}_{1,2} + \bar{\gamma}_{1,2,3}, \text{ where } \bar{\gamma}_{1,2,3} = \gamma_{1,2,3}^R p^R + \gamma_{1,2,3}^L p^L + \gamma_{1,2,3}^D p^D + \gamma_{1,2,3}^M p^M \quad (B.9)$$

2. **Intention to Treat Effect (ITT)**

Next, consider the intent to treat (ITT) effect. This is defined as the impact on teacher skills for the teachers who were offered the treatment in year 1, that is, the teachers who were in the treated schools in year 1. The counterfactual is being assigned to a control school in year 1. The general definition for ITT in year t is:

$$ITT_{tchr}(t) \equiv E[y^t | R_{tchr,\, year\, 1} = 1] - E[y^t | R_{tchr,\, year\, 1} = 0] \quad (B.10)$$

$R_{tchr,\, year\, 1}$ refers to the teacher's school in year 1, which can differ from his or her school in year t. Applying this definition to years 1, 2 and 3 yields the following, more specific, definitions:

$$ITT_{tchr}(1) \equiv E[y^1 | R_{tchr,\, year\, 1} = 1] - E[y^1 | R_{tchr,\, year\, 1} = 0] = \bar{\delta} \quad (B.11)$$

$$ITT_{tchr}(2) \equiv E[y^2 | R_{tchr,\, year\, 1} = 1] - E[y^2 | R_{tchr,\, year\, 1} = 0] \quad (B.12)$$

$$= \bar{\theta}^2 + [p^R(2\delta^R+\gamma^R_{1,2}) + p^L(2\delta^L+\gamma^L_{1,2}) + p^D\delta^D + p^M(\tau(2\delta^M+\gamma^M_{1,2})+(1-\tau)\delta^M)] - [\bar{\theta}^2 + p^L\delta^L + p^M\tau\delta^M]$$

$$= p^R(2\delta^R+\gamma^R_{1,2}) + p^L(\delta^L+\gamma^L_{1,2}) + p^D\delta^D + p^M(\delta^M + \tau\gamma^M_{1,2})$$

$$= \bar{\delta} + p^R(\delta^R+\gamma^R_{1,2}) + p^L\gamma^L_{1,2} + p^M\tau\gamma^M_{1,2}$$

$$ITT_{tchr}(3) \equiv E[y^3 | R_{tchr,\, year\, 1} = 1] - E[y^3 | R_{tchr,\, year\, 1} = 0] \quad (B.13)$$

$$= \bar{\theta}^3 + p^R(3\delta^R+3\gamma^R_{1,2}+\gamma^R_{1,2,3}) + p^L(3\delta^L+3\gamma^L_{1,2}+\gamma^L_{1,2,3}) + p^D\delta^D$$

$$+ p^M[\tau^2(3\delta^M+3\gamma^M_{1,2}+\gamma^M_{1,2,3}) + 2\tau(1-\tau)(2\delta^M+\gamma^M_{1,2}) + (1-\tau)^2\delta^M)]$$

$$- [\bar{\theta}^3 + p^L(2\delta^L+\gamma^L_{1,2}) + p^M[\tau^2(2\delta^M+\gamma^M_{1,2}) + 2\tau(1-\tau)\delta^M]$$

$$= p^R(3\delta^R+3\gamma^R_{1,2}+\gamma^R_{1,2,3}) + p^L(\delta^L+2\gamma^L_{1,2}+\gamma^L_{1,2,3}) + p^D\delta^D + p^M(\delta^M+2\tau\gamma^M_{1,2}+\tau^2\gamma^M_{1,2,3})$$

$$= \bar{\delta} + p^R(2\delta^R+3\gamma^R_{1,2}+\gamma^R_{1,2,3}) + p^L(2\gamma^L_{1,2}+\gamma^L_{1,2,3}) + p^M(2\tau\gamma^M_{1,2}+\tau^2\gamma^M_{1,2,3})$$

where $\bar{\theta}^2 = p^R\bar{\theta}^{2,R} + p^L\bar{\theta}^{2,L} + p^D\bar{\theta}^{2,D} + p^M\bar{\theta}^{2,M}$ and $\bar{\theta}^3 = p^R\bar{\theta}^{3,R} + p^L\bar{\theta}^{3,L} + p^D\bar{\theta}^{3,D} + p^M\bar{\theta}^{3,M}$. The intuition behind the $p^M(\tau(2\delta^M+\gamma^M_{1,2})+(1-\tau)\delta^M)$ term in the second line of (B.12) is that, of the movers in APM schools in year 2, a proportion τ were also in an APM school in year 1, so the combined effect of two years of treatment for them is $2\delta^M+\gamma^M_{1,2}$, and a proportion 1-τ were in non-APM schools in year 1, so the effect of treatment for one year for them is δ^M.

3. Average Causal Response (ACR)

The third treatment effect for teacher skills is similar to a local average treatment effect (LATE), but it differs from LATE because treatment can vary by 1, 2 or 3 years. Angrist and Imbens (1995) extended LATE to this case, and they called this treatment effect the average causal response (ACR). The general definition after t years is:

$$ACR_{tchr}(t) \equiv \sum_{s=1}^{t} E[y_s^t - y_{s-1}^t | T_1^t \geq s > T_0^t] \frac{Prob[T_1^t \geq s > T_0^t]}{\sum_{r=1}^{t} Prob[T_1^t \geq r > T_0^t]} \quad (B.14)$$

where T_0^t is the (potential) number of years of training up through year t for a teacher who was assigned to a non-APM school in year 1, and T_1^t is the (potential) number of years of training up through year t for a teacher assigned to an APM school in year 1.[2] The subscripts on y indicate the value of y given a (potential) number of *years* of treatment (which varies from 0 to 3), not the value of y given "treatment or no treatment" (a binary variable), as was the case for the definition for $ATE_{tchr}(t)$.

Applying this general definition to years 1, 2 and 3 yields:

$$ACR_{tchr}(1) \equiv E[y_1^1 - y_0^1 | T_1^1 \geq 1 > T_0^1] \frac{Prob[T_1^1 \geq 1 > T_0^1]}{Prob[T_1^1 \geq 1 > T_0^1]} \quad (B.15)$$

$$= E[y_1^1 - y_0^1 | T_1^1 \geq 1 > T_0^1]$$

$$= E[y^1 | R = 1] - E[y^1 | R = 0] = \bar{\delta}$$

[2] For the general case, possible values for both T_0^t and T_1^t are integers from 0 to t. However, for the APM program all teachers followed their random assignment in year 1, so possible values for T_0^t are 0 to t-1, and for T_1^t are 1 to t.

$$\mathrm{ACR_{tchr}}(2) \equiv \mathrm{E}[y_1^2 - y_0^2 |\ T_1^2 \geq 1 > T_0^2]\ \frac{\mathrm{Prob}[T_1^2 \geq 1 > T_0^2]}{\mathrm{Prob}[T_1^2 \geq 1 > T_0^2] + \mathrm{Prob}[T_1^2 = 2 > T_0^2]} \qquad \text{(B.16)}$$

$$+\ \mathrm{E}[y_2^2 - y_1^2 |\ T_1^2 = 2 > T_0^2]\ \frac{\mathrm{Prob}[T_1^2 = 2 > T_0^2]}{\mathrm{Prob}[T_1^2 \geq 1 > T_0^2] + \mathrm{Prob}[T_1^2 = 2 > T_0^2]}$$

$$= \frac{\delta^R p^R + \delta^D p^D + \delta^M (1-\tau) p^M}{p^R + p^D + (1-\tau) p^M} \times \frac{p^R + p^D + (1-\tau) p^M}{1 + p^R}$$

$$+ \frac{(\delta^R + \gamma_{1,2}^R) p^R + (\delta^L + \gamma_{1,2}^L) p^L + (\delta^M + \gamma_{1,2}^M) \tau p^M}{p^R + p^L + \tau p^M} \times \frac{p^R + p^L + \tau p^M}{1 + p^R}$$

$$= [\overline{\delta} + p^R(\delta^R + \gamma_{1,2}^R) + p^L \gamma_{1,2}^L + p^M \tau \gamma_{1,2}^M]/[1 + p^R] = \mathrm{ITT_{tchr}}(2)/[1 + p^R]$$

$$\mathrm{ACR_{tchr}}(3) \equiv \mathrm{E}[y_1^3 - y_0^3 |\ T_1^3 \geq 1 > T_0^3]\ \frac{\mathrm{Prob}[T_1^3 \geq 1 > T_0^3]}{\mathrm{Prob}[T_1^3 \geq 1 > T_0^3] + \mathrm{Prob}[T_1^3 = 2 > T_0^3] + \mathrm{Prob}[T_1^3 = 3 > T_0^3]} \qquad \text{(B.17)}$$

$$+\ \mathrm{E}[y_2^3 - y_1^3 |\ T_1^3 \geq 2 > T_0^3]\ \frac{\mathrm{Prob}[T_1^3 \geq 2 > T_0^3]}{\mathrm{Prob}[T_1^3 \geq 1 > T_0^3] + \mathrm{Prob}[T_1^3 = 2 > T_0^3] + \mathrm{Prob}[T_1^3 = 3 > T_0^3]}$$

$$+\ \mathrm{E}[y_3^3 - y_2^3 |\ T_1^3 \geq 3 > T_0^3]\ \frac{\mathrm{Prob}[T_1^3 \geq 3 > T_0^3]}{\mathrm{Prob}[T_1^3 \geq 1 > T_0^3] + \mathrm{Prob}[T_1^3 = 2 > T_0^3] + \mathrm{Prob}[T_1^3 = 3 > T_0^3]}$$

$$= \frac{\delta^R p^R + \delta^D p^D + \delta^M (1-\tau)^2 p^M}{p^R + p^D + (1-\tau)^2 p^M} \times \frac{p^R + p^D + (1-\tau)^2 p^M}{1 + 2p^R}$$

$$+ \frac{(\delta^R + \gamma_{1,2}^R) p^R + (\delta^M + \gamma_{1,2}^M) 2\tau(1-\tau) p^M}{p^R + 2\tau(1-\tau) p^M} \times \frac{p^R + 2\tau(1-\tau) p^M}{1 + 2p^R}$$

$$+ \frac{(\delta^R + 2\gamma_{1,2}^R + \gamma_{1,2,3}^R) p^R + (\delta^L + 2\gamma_{1,2}^L + \gamma_{1,2,3}^L) p^L + \tau^2 (\delta^M + 2\gamma_{1,2}^M + \gamma_{1,2,3}^M) p^M}{p^R + p^L + \tau^2 p^M} \times \frac{p^R + p^L + \tau^2 p^M}{1 + 2p^R}$$

$$= [\overline{\delta} + p^R(2\delta^R + 3\gamma_{1,2}^R + \gamma_{1,2,3}^R) + p^L(2\gamma_{1,2}^L + \gamma_{1,2,3}^L) + p^M(2\tau\gamma_{1,2}^M + \tau^2\gamma_{1,2,3}^M)]/[1 + 2p^R]$$

$$= \mathrm{ITT_{tchr}}(3)/[1 + 2p^R]$$

To see the intuition behind $ACR_{tchr}(t)$, consider $ACR_{tchr}(2)$ given in equation (B.16). The term $E[y_1^2 - y_0^2 | T_1^2 \geq 1 > T_0^2]$ is the impact on teacher skills of receiving one year of treatment, relative to having zero years of treatment, as indicated by the subscripts on the y terms, for teachers who would have had one or two years of treatment in year 2 if assigned to an APM school in year 1 ($T_1^2 \geq 1$), but would not have been treated in year 2 if assigned to a non-APM school in year 1 ($T_0^2 < 1$). Of the four teacher types, this includes all remainers and dislikers, and movers who randomly switched to a non-APM school in year 2 (for whom $T_0^2 = 0$ and $T_1^2 = 1$). The term $E[y_2^2 - y_1^2 | T_1^2 = 2 > T_0^2]$ is the impact on teacher skills of receiving a *second* year of the treatment, *relative to having one year of treatment*, as indicated by the subscripts on the y terms, for teachers who would have had two years of treatment in year 2 if assigned to an APM school in year 1 but only zero or one year of in year 2 if assigned to a non-APM school in year 1. This includes all remainers, all likers, and movers who randomly switched to APM schools in year 2 (for whom $T_0^2 = 1$ and $T_1^2 = 2$). Turning to the sum of the two probability terms in the denominator, $Prob[T_1^2 \geq 1 > T_0^2]$ is the probability that a teacher is a remainer, a disliker, or a mover who randomly switches to a non-APM school in year 2, and $Prob[T_1^2 = 2 > T_0^2]$ is the probability that a teacher is a remainer, a liker, or a mover who randomly switches to an APM school in year 2. Their sum is greater than 1; remainers are “counted twice” since they are included in both probabilities. Likers, dislikers, and movers are “counted” only once.

In effect, $ACR_{tchr}(2)$ is an average of: a) the (average) impact on teacher skills of going from no treatment to one year of treatment for remainers, dislikers, and those movers who randomly move to a non-APM school in year 2; and b) the (average) impact on those skills of going from one to two years of treatment for remainers, likers, and the movers who randomly move to APM schools in year 2. Thus, $ACR_{tchr}(2)$ is the average of the impact on teacher skills for each additional year of treatment brought about by random assignment to an APM school in year 1, with remainers getting “double weight” since that assignment raises their years of treatment by two years, while for all others random assignment increases years of treatment by only one year. Importantly, note that, for any t, $ACR_{tchr}(t)$ is a *per year* (*not* a cumulative) impact that averages only over years of treatment induced by random assignment to an APM school in year 1. To obtain a cumulative impact over t years, multiply $ACR_{tchr}(t)$ by the years of coaching induced by random assignment to an APM school, which is the denominator in equation (B.14). Thus, the cumulative impact will be equal to $ITT_{tchr}(t)$.

4. Average Treatment Effect (on teacher skills) for teachers in treated schools (ATE_{sch})

The three treatment effects discussed so far, in effect, follow teachers who move to other schools. But many teacher training or coaching programs focus on particular schools, so it is useful to define treatment effects for the teachers in the schools that are implementing the APM program. As explained above, the number of teaching positions in a given school rarely changes. If the number of teaching positions in all schools is fixed, the proportion of teachers in treated schools in years 2 and 3 who are movers is $(\mu/\tau)p^M$, where μ is the proportion of all movers who move to an APM school in year 2 or year 3 (which is determined by the application process that also determines the proportions of teachers who are remainers, likers, dislikers and movers), and the proportion of teachers in control schools in years 2 and 3 who are movers is $[(1-\mu)/(1-\tau)]p^M$.[3]

There are two possibilities for treatment effects that focus on schools. The first is an average treatment effect (ATE) on teacher skills for those schools, where the counterfactual is no program at all, which we denote as ATE_{sch}. This is defined as follows for year t:

$$ATE_{sch}(t) \equiv E[y^t| R = 1] - E[y^t| \text{Program does not exist}] \qquad (B.18)$$

Applying this general definition for years 1, 2 and 3 yields:

$$ATE_{sch}(1) \equiv E[y^1| R = 1] - E[y^1| \text{Program does not exist}]$$

$$= \bar{\theta}^1 + \bar{\delta} - \bar{\theta}^1 = \bar{\delta}$$

$$ATE_{sch}(2) \equiv E[y^2| R = 1] - E[y^2| \text{Program does not exist}] \quad (B.19)$$

$$= (\bar{\theta}^{2,R} + 2\delta^R + \gamma^R_{1,2})p^R + (\bar{\theta}^{2,L} + 2\delta^L + \gamma^L_{1,2})p^L + (\bar{\theta}^{2,L} + \delta^L)p^L((1-\tau)/\tau)$$

[3] This definition of μ implies that, among all teachers in APM and non-APM schools, the proportion who are movers in APM schools in year 2 or 3 is μp^M. Focusing on APM schools only, this proportion must be divided by τ, yielding $(\mu/\tau)p^M$. A similar derivation shows that the proportion of movers in non-APM schools is $[(1-\mu)/(1-\tau)]p^M$.

$$+ (\bar{\theta}^{2,M} + 2\delta^M + \gamma^M_{1,2})\tau(\mu/\tau)p^M + (\bar{\theta}^{2,M} + \delta^M)(1-\tau)(\mu/\tau)p^M - [\bar{\theta}^{2,R}p^R + \bar{\theta}^{2,L}p^L + \bar{\theta}^{2,D}p^D + \bar{\theta}^{2,M}p^M]$$

$$= (2\delta^R + \gamma^R_{1,2})p^R + (\delta^L(1+\tau) + \gamma^L_{1,2}\tau)p^L/\tau + [(1+\tau)\delta^M + \tau\gamma^M_{1,2}](\mu/\tau)p^M$$

$$+ \bar{\theta}^{2,L}p^L((1-\tau)/\tau) - \bar{\theta}^{2,D}p^D + \bar{\theta}^{2,M}p^M((\mu/\tau) - 1)$$

$$ATE_{sch}(3) \equiv E[y^3 | R = 1] - E[y^3 | \text{Program does not exist}] \quad \text{(B.20)}$$

$$= (\bar{\theta}^{3,R} + 3\delta^R + 3\gamma^R_{1,2} + \gamma^R_{1,2,3})p^R + (\bar{\theta}^{3,L} + 3\delta^L + 3\gamma^L_{1,2} + \gamma^L_{1,2,3})p^L + (\bar{\theta}^{3,L} + 2\delta^L + \gamma^L_{1,2})p^L((1-\tau)/\tau)$$

$$+ (\bar{\theta}^{3,M} + 3\delta^M + 3\gamma^M_{1,2} + \gamma^M_{1,2,3})\tau^2(\mu/\tau)p^M + (\bar{\theta}^{3,M} + 2\delta^M + \gamma^M_{1,2})2\tau(1-\tau)(\mu/\tau)p^M + (\bar{\theta}^{3,M} + \delta^M)(1-\tau)^2(\mu/\tau)p^M$$

$$- [\bar{\theta}^{3,R}p^R + \bar{\theta}^{3,L}p^L + \bar{\theta}^{3,D}p^D + \bar{\theta}^{3,M}p^M]$$

$$= (3\delta^R + 3\gamma^R_{1,2} + \gamma^R_{1,2,3})p^R + (\delta^L(2+\tau) + (1+2\tau)\gamma^L_{1,2} + \tau\gamma^L_{1,2,3})p^L/\tau + [(1+2\tau)\delta^M + \tau(2+\tau)\gamma^M_{1,2} + \gamma^M_{1,2,3}\tau^2](\mu/\tau)p^M$$

$$+ \bar{\theta}^{3,L}p^L((1-\tau)/\tau) - \bar{\theta}^{3,D}p^D + \bar{\theta}^{3,M}p^M((\mu/\tau) - 1)$$

The first line of the final expressions for $ATE_{sch}(2)$ and $ATE_{sch}(3)$ are the treatment effect, and the last line is the composition effect.

5. Intent to Treat Effect (on teacher skills) for teachers in treated schools (ITT_{sch})

The second treatment effect for teacher skills that focuses on schools is an ITT effect; it is similar to ATE_{sch} except that the counterfactual is the skills of teachers in non-APM schools:

$$ITT_{sch}(t) \equiv E[y^t | R = 1] - E[y^t | R = 0] \quad \text{(B.21)}$$

Applying this general definition for years 1, 2 and 3 yields:

$$\mathrm{ITT}_{sch}(1) \equiv E[y^1 | R = 1] - E[y^1 | R = 0] \qquad \text{(B.22)}$$

$$= \bar{\theta}^1 + \bar{\delta} - \bar{\theta}^1 = \bar{\delta}$$

$$\mathrm{ITT}_{sch}(2) \equiv E[y^2 | R = 1] - E[y^2 | R = 0] \qquad \text{(B.23)}$$

$$= (\bar{\theta}^{2,R} + 2\delta^R + \gamma^R_{1,2})p^R + (\bar{\theta}^{2,L} + 2\delta^L + \gamma^L_{1,2})p^L + (\bar{\theta}^{2,L} + \delta^L)p^L((1\text{-}\tau)/\tau)$$

$$+ (\bar{\theta}^{2,M} + 2\delta^M + \gamma^M_{1,2})\tau(\mu/\tau)p^M + (\bar{\theta}^{2,M} + \delta^M)(1\text{-}\tau)(\mu/\tau)p^M$$

$$- [\bar{\theta}^{2,R}p^R + (\delta^D\tau + \bar{\theta}^{2,D})p^D/(1\text{-}\tau) + (\delta^M\tau + \bar{\theta}^{2,M}p^M)((1\text{-}\mu)/(1\text{-}\tau))]$$

$$= (2\delta^R + \gamma^R_{1,2})p^R + (\delta^L(1+\tau) + \tau\gamma^L_{1,2})(p^L/\tau) - \delta^D\tau(p^D/(1\text{-}\tau)) + ((\delta^M(1+\tau) + \gamma^M_{1,2}\tau)(\mu/\tau) - \delta^M\tau(1\text{-}\mu)/(1\text{-}\tau))p^M$$

$$+ \bar{\theta}^{2,L}(p^L/\tau) - \bar{\theta}^{2,D}(p^D/(1\text{-}\tau)) + \bar{\theta}^{2,M}p^M((\mu/\tau) - (1\text{-}\mu)/(1\text{-}\tau))$$

$$\mathrm{ITT}_{sch}(3) \equiv E[y^3 | R = 1] - E[y^3 | R = 0] \qquad \text{(B.24)}$$

$$= (\bar{\theta}^{3,R} + 3\delta^R + 3\gamma^R_{1,2} + \gamma^R_{1,2,3})p^R + (\bar{\theta}^{3,L} + 3\delta^L + 3\gamma^L_{1,2} + \gamma^L_{1,2,3})p^L + (\bar{\theta}^{3,L} + 2\delta^L + \gamma^L_{1,2})p^L((1\text{-}\tau)/\tau)$$

$$+ (\bar{\theta}^{3,M} + 3\delta^M + 3\gamma^M_{1,2} + \gamma^M_{1,2,3})\tau^2(\mu/\tau)p^M + (\bar{\theta}^{3,M} + 2\delta^M + \gamma^M_{1,2})2\tau(1\text{-}\tau)(\mu/\tau)p^M + (\bar{\theta}^{3,M} + \delta^M)(1\text{-}\tau)^2(\mu/\tau)p^M$$

$$- [\bar{\theta}^{3,R}p^R + (\delta^D\tau + \bar{\theta}^{3,D})p^D/(1\text{-}\tau) + ((2\delta^M + \gamma^M_{1,2})\tau^2 + 2\tau(1\text{-}\tau)\delta^M + \bar{\theta}^{3,M})p^M((1\text{-}\mu)/(1\text{-}\tau))]$$

$$= (3\delta^R + 3\gamma^R_{1,2} + \gamma^R_{1,2,3})p^R + (\delta^L(2+\tau) + \gamma^L_{1,2}(1+2\tau) + \gamma^L_{1,2,3}\tau)(p^L/\tau) - \delta^D\tau(p^D/(1\text{-}\tau))$$

$$+ [(\mu/\tau)(\delta^M(1+2\tau) + \gamma^M_{1,2}\tau(2+\tau) + \gamma^M_{1,2,3}\tau^2) - ((1\text{-}\mu)/(1\text{-}\tau)(\delta^M 2\tau + \gamma^M_{1,2}\tau^2)]p^M$$

$$+ \bar{\theta}^{3,L}(p^L/\tau) - \bar{\theta}^{3,D}(p^D/(1\text{-}\tau)) + \bar{\theta}^{3,M}p^M((\mu/\tau) - (1\text{-}\mu)/(1\text{-}\tau))$$

The first line of $\mathrm{ITT}_{sch}(2)$ is (first two lines of $\mathrm{ITT}_{sch}(3)$ are) the (net) treatment effect, and the last line is the composition effect.

III. Definitions of Treatment Effects for Students' Skills (denoted by s)

Focusing on students' skills is simplified by the fact that students are assumed not to change schools, and that the schools they are in always follow their (the schools') random assignment.

We define three treatment effects for student skills. The first two, ATE_{stud} and ITT_{stud}, correspond to the two treatment effects defined for their schools (ATE_{sch} and ITT_{sch}). These treatment effects for years 2 and 3 are complex because there are several possible "histories" for students' teachers in those years. For example, in year 2 a student's teacher in a treated school could be a liker who was in an APM school in years 1 and 2, or a liker who was in a non-APM school in year 1 but in an APM school in year 2. Another example is a student in a treated school in year 3; if he or she was taught by treated teacher in year 1 (who by definition had had one year of APM at that time), then by a teacher in year 2 who had APM in year 2 but not year 1, and by a teacher in year 3 who had APM in years 2 and 3 but not in year 1, he or she has been exposed to four years of teacher treatment, and his or her cumulative gain in learning from exposure to those teachers will be averaged over the four years. The general definition of ATE_{stud} for year t is:

$$ATE_{stud}(t) \equiv E[s^t | R = 1] - E[s^t | \text{Program does not exist}] \qquad (B.25)$$

Applying this to years 1, 2 and 3 yield the specific treatment effects for those years:

$$ATE_{stud}(1) \equiv E[s^1 | R = 1] - E[s^1 | \text{Program does not exist}] \qquad (B.26)$$

$$= E[\sigma s^0 + \pi y^1 | R = 1] - E[\sigma s^0 + \pi y^1 | \text{Program does not exist}]$$

$$= E[\pi y^1 | R = 1] - E[\pi y^1 | \text{Program does not exist}]$$

$$= \pi E[y^1 | R = 1] - \pi E[y^1 \mid \text{Program does not exist}]$$

$$= \pi(\bar{\theta}^1 + \bar{\delta}) - \pi\bar{\theta}^1$$

$$= \pi\bar{\delta}, \text{ where } \bar{\delta} = \delta^R p^R + \delta^L p^L + \delta^D p^D + \delta^M p^M$$

To obtain $ATE_{stud}(2)$, one can use the results for $ATE_{sch}(2)$ in equation (B.19):

$$ATE_{stud}(2) \equiv E[s^2 | R = 1] - E[s^2 | \text{Program does not exist}] \qquad (B.26)$$

$$= E[\sigma s^1 + \pi y^2 | R = 1] - E[\sigma s^1 + \pi y^2 | \text{Program does not exist}]$$

$$= \sigma(E[s^1 | R = 1] - E[s^1 | \text{Program does not exist}])$$

$$+ \pi(E[y_j^2 | R = 1] - E[y^2 | \text{Program does not exist}])$$

$$= \sigma\pi\bar{\delta} + \pi[(2\delta^R + \gamma^R_{1,2})p^R + (\delta^L(1+\tau) + \gamma^L_{1,2}\tau)p^L/\tau + ((1+\tau)\delta^M + \tau\gamma^M_{1,2})(\mu/\tau)p^M]$$

$$+ \pi[\bar{\theta}^{2,L}p^L((1-\tau)/\tau) - \bar{\theta}^{2,D}p^D + \bar{\theta}^{2,M}p^M((\mu/\tau) - 1)]$$

The first line is the treatment effect and the second line is the composition effect.

Year 3 is slightly more complicated since movers continue to move but likers and dislikers (and remainers) do not move between years 2 and 3. Using the results for $ATE_{sch}(3)$ from (B.20).

$$ATE_{stud}(3) \equiv E[s^3 | R = 1] - E[s^3 | \text{Program does not exist}] \qquad (B.27)$$

$$= E[\sigma s^2 + \pi y^3 | R = 1] - E[\sigma s^2 + \pi y^3 | \text{Program does not exist}]$$

$$= \sigma(E[s^2 | R = 1] - E[s^2 | \text{Program does not exist}])$$

$$+ \pi(E[y^3 | R = 1] - E[y^3 | \text{Program does not exist}])$$

$$= \sigma^2\pi\bar{\delta} + \sigma\pi[(2\delta^R + \gamma^R_{1,2})p^R + (\delta^L(1+\tau) + \gamma^L_{1,2}\tau)(p^L/\tau) + ((1+\tau)\delta^M + \tau\gamma^M_{1,2})(\mu/\tau)p^M]$$

$$+ \sigma\pi[\bar{\theta}^{2,L}p^L((1-\tau)/\tau) - \bar{\theta}^{2,D}p^D + \bar{\theta}^{2,M}p^M((\mu/\tau) - 1)]$$

$$+ \pi[(3\delta^R + 3\gamma^R_{1,2} + \gamma^R_{1,2,3})p^R + (\delta^L(2+\tau) + (1+2\tau)\gamma^L_{1,2} + \tau\gamma^L_{1,2,3})(p^L/\tau) + ((1+2\tau)\delta^M + \tau(2+\tau)\gamma^M_{1,2} + \gamma^M_{1,2,3}\tau^2)(\mu/\tau)p^M]$$

$$+ \pi[\bar{\theta}^{3,L}p^L((1-\tau)/\tau) - \bar{\theta}^{3,D}p^D + \bar{\theta}^{3,M}p^M((\mu/\tau) - 1)]$$

$$= \sigma^2\pi\bar{\delta} + \sigma\pi[(2\delta^R + \gamma^R_{1,2})p^R + (\delta^L(1+\tau) + \gamma^L_{1,2}\tau)(p^L/\tau) + ((1+\tau)\delta^M + \tau\gamma^M_{1,2})(\mu/\tau)p^M]$$

$$+ \pi[(3\delta^R + 3\gamma^R_{1,2} + \gamma^R_{1,2,3})p^R + (\delta^L(2+\tau) + (1+2\tau)\gamma^L_{1,2} + \tau\gamma^L_{1,2,3})(p^L/\tau) + ((1+2\tau)\delta^M + \tau(2+\tau)\gamma^M_{1,2} + \gamma^M_{1,2,3}\tau^2)(\mu/\tau)p^M]$$

$$+ \sigma\pi[\bar{\theta}^{2,L}p^L((1\text{-}\tau)/\tau) - \bar{\theta}^{2,D}p^D + \bar{\theta}^{2,M}p^M((\mu/\tau) - 1)] + \pi[\bar{\theta}^{3,L}p^L((1\text{-}\tau)/\tau) - \bar{\theta}^{3,D}p^D + \bar{\theta}^{3,M}p^M((\mu/\tau) - 1)]$$

The first two lines are the treatment effect and the last line is the composition effect.

Next, turn to ITT_{stud}. The general definition is:

$$ITT_{stud}(t) \equiv E[s^t| R = 1] - E[s^t| R = 0] \qquad (B.28)$$

Applying this to years 1, 2 and 3 yield the specific treatment effects for those years:

$$ITT_{stud}(1) \equiv E[s^1| R = 1] - E[s^1| R = 0] \qquad (B.29)$$

$$= E[\sigma s^0 + \pi y^1| R = 1] - E[\sigma s^0 + \pi y^1| R = 0]$$

$$= E[\pi y^1| R = 1] - E[\pi y^1| R = 0]$$

$$= \pi E[y^1| R = 1] - \pi E[y^1| R = 0]$$

$$= \pi(\bar{\theta}^1 + \bar{\delta}) - \pi\bar{\theta}^1 = \pi\bar{\delta}$$

$$ITT_{stud}(2) \equiv E[s^2| R = 1] - E[s^2| R = 0] \qquad (B.30)$$

$$= E[\sigma s^1 + \pi y^2| R_{)} = 1] - E[\sigma s^1 + \pi y^2| R = 0]$$

$$= \sigma(E[s^1| R = 1] - E[s^1| R = 0) + \pi(E[y^2| R = 1] - E[y^2| R = 0])$$

$$= \sigma\pi\bar{\delta} + \pi[(2\delta^R + \gamma^R_{1,2})p^R + (\delta^L(1+\tau) + \tau\gamma^L_{1,2})(p^L/\tau) - \delta^D\tau(p^D/(1\text{-}\tau)) + ((\delta^M(1+\tau) + \gamma^M_{1,2}\tau)(\mu/\tau) - \delta^M\tau(1\text{-}\mu)/(1\text{-}\tau))p^M]$$

$$+ \pi[\bar{\theta}^{2,L}(p^L/\tau) - \bar{\theta}^{2,D}(p^D/(1\text{-}\tau)) + \bar{\theta}^{2,M}p^M((\mu/\tau) - (1\text{-}\mu)/(1\text{-}\tau))]$$

The first line is the (net) treatment effect and the second line is the composition effect.

$$ITT_{stud}(3) \equiv E[s^3|\, R = 1] - E[s^3|\, R = 0] \qquad \text{(B.31)}$$

$$= E[\sigma s_i^2 + \pi y_j^3|\, R = 1] - E[\sigma s^2 + \pi y^3|\, R = 0]$$

$$= \sigma(E[s^2|\, R = 1] - E[s^2|\, R = 0]) + \pi(E[y^3|\, R = 1] - E[y^3|\, R = 0])$$

$$= \sigma^2\pi\bar{\delta} + \sigma\pi[(2\delta^R + \gamma^R_{1,2})p^R + (\delta^L(1+\tau) + \tau\gamma^L_{1,2})(p^L/\tau) - \delta^D\tau(p^D/(1\text{-}\tau)) + ((\delta^M(1+\tau) + \gamma^M_{1,2}\tau)(\mu/\tau) - \delta^M\tau(1\text{-}\mu)/(1\text{-}\tau))p^M]$$

$$+ \sigma\pi[\bar{\theta}^{2,L}(p^L/\tau) - \bar{\theta}^{2,D}(p^D/(1\text{-}\tau)) + \bar{\theta}^{2,M}p^M((\mu/\tau) - (1\text{-}\mu)/(1\text{-}\tau))]$$

$$+ \pi[(3\delta^R + 3\gamma^R_{1,2} + \gamma^R_{1,2,3})p^R + (\delta^L(2+\tau) + \gamma^L_{1,2}(1+2\tau) + \gamma^L_{1,2,3}\tau)(p^L/\tau) - \delta^D\tau(p^D/(1\text{-}\tau))]$$

$$+ \pi[(\mu/\tau)(\delta^M(1+2\tau) + \gamma^M_{1,2}\tau(2+\tau) + \gamma^M_{1,2,3}\tau^2) - ((1\text{-}\mu)/(1\text{-}\tau)(\delta^M 2\tau + \gamma^M_{1,2}\tau^2)]p^M$$

$$+ \pi[\bar{\theta}^{3,L}(p^L/\tau) - \bar{\theta}^{3,D}(p^D/(1\text{-}\tau)) + \bar{\theta}^{3,M}p^M((\mu/\tau) - (1\text{-}\mu)/(1\text{-}\tau))]$$

$$= \sigma^2\pi\bar{\delta} + \sigma\pi[(2\delta^R + \gamma^R_{1,2})p^R + (\delta^L(1+\tau) + \tau\gamma^L_{1,2})(p^L/\tau) - \delta^D\tau(p^D/(1\text{-}\tau)) + (\delta^M((1+\tau) + \gamma^M_{1,2}\tau)(\mu/\tau) - \delta^M\tau(1\text{-}\mu)/(1\text{-}\tau))p^M]$$

$$+ \pi[(3\delta^R + 3\gamma^R_{1,2} + \gamma^R_{1,2,3})p^R + (\delta^L(2+\tau) + \gamma^L_{1,2}(1+2\tau) + \gamma^L_{1,2,3}\tau)(p^L/\tau) - \delta^D\tau(p^D/(1\text{-}\tau))]$$

$$+ \pi[(\mu/\tau)(\delta^M(1+2\tau) + \gamma^M_{1,2}\tau(2+\tau) + \gamma^M_{1,2,3}\tau^2) - ((1\text{-}\mu)/(1\text{-}\tau)(\delta^M 2\tau + \gamma^M_{1,2}\tau^2)]p^M$$

$$+ \sigma\pi[\bar{\theta}^{2,L}(p^L/\tau) - \bar{\theta}^{2,D}(p^D/(1\text{-}\tau)) + \bar{\theta}^{2,M}p^M((\mu/\tau) - (1\text{-}\mu)/(1\text{-}\tau))]$$

$$+ \pi[\bar{\theta}^{3,L}(p^L/\tau) - \bar{\theta}^{3,D}(p^D/(1\text{-}\tau)) + \bar{\theta}^{3,M}p^M((\mu/\tau) - (1\text{-}\mu)/(1\text{-}\tau))]$$

The first three lines are the (net) treatment effect and the last two lines are the composition effect.

The third treatment effect for students is the (average) impact of an additional year of teacher training on student learning, averaged over all additional years of that training that a student experiences. In effect, this is a transfer of the ACR_{tchr} treatment effects on teacher skill onto student learning, which is complicated by the many different "histories" a student can have in terms of treated teachers in years 2 and 3. We call these treatment effects ACR_{stud} effects, though they differ from ACR_{tchr} (and so differ from the ACR effects of Angrist and Imbens, 1995) since students are not *directly* treated but instead are *indirectly* treated by exposure to treated teachers.

The general definition of $ACR_{students}$ in year t (1, 2 or 3) is:

$$ACR_{stud}(t) \equiv \frac{E[s^t|R=1] - E[s^t|R=0]}{E[h_{tchr}(t)|R=1] - E[h_{tchr}(t)|R=0]} \qquad \text{(B.32)}$$

where $h_{tchr}(t)$ is the cumulative "history" from year 1 to year t of a student's exposure to teachers with APM coaching. For example, a student in a treated school in year 2 had a coached teacher in year 1, but in year 2 the teacher could have one or two years of coaching (i.e. one if the teacher was in a non-APM school in year 1), so the student could have $h_{tchr}(2)$ of either 2 or 3. The expected value of $h_{tchr}(t)$ averages over the types of teachers in the school from year 1 to year t.

For year 1, $ACR_{stud}(1) = ATT_{stud}(t) = ITT_{stud}(t)$ since all teachers follow their random assignment in year 1, so:

$$ACR_{stud}(1) = \pi\bar{\delta} \qquad \text{(B.33)}$$

For year 2, the definition in (B.32) gives (using the derivations in (B.30)):

$$ACR_{stud}(2) \equiv \frac{E[s^2|R=1] - E[s^2|R=0]}{E[h_{tchr}(2)|R=1] - E[h_{tchr}(2)|R=0]} \qquad \text{(B.34)}$$

$$= \frac{\sigma\pi\bar{\delta} + \pi[(2\delta^R + \gamma^R_{1,2})p^R + (\delta^L(1+\tau) + \tau\gamma^L_{1,2})(p^L/\tau) - \delta^D\tau(p^D/(1-\tau)) + ((\delta^M(1+\tau) + \gamma^M_{1,2}\tau)(\mu/\tau) - \delta^M\tau(1-\mu)/(1-\tau))p^M]}{1 + 2p^R + (p^L/\tau)(\tau+1) + p^M(\mu/\tau)(1+\tau) - [\tau p^D/(1-\tau) + \tau p^M((1-\mu)/(1-\tau))]}$$

$$+ \frac{\pi[\bar{\theta}^{2,L}(p^L/\tau) - \bar{\theta}^{2,D}(p^D/(1-\tau)) + \bar{\theta}^{2,M}p^M((\mu/\tau) - (1-\mu)/(1-\tau))]}{1 + 2p^R + (p^L/\tau)(\tau+1) + p^M(\mu/\tau)(1+\tau) - [\tau p^D/(1-\tau) + \tau p^M((1-\mu)/(1-\tau))]}$$

For year 3, applying the definition in (B.32) yields (using the derivations in (B.31)):

$$\text{ACR}_{\text{stud}}(3) \equiv \frac{E[s^3|R=1]-E[s^3|R=0]}{E[h_{tchr}(3)|R=1]-E[h_{tchr}(3)|R=0]} \qquad \text{(B.35)}$$

$$=\pi\frac{\sigma^2\bar{\delta}+\left((3+2\sigma)\delta^R+(3+\sigma)\gamma_{1,2}^R+\gamma_{1,2,3}^R\right)p^R+\left(\delta^L(\sigma(1+\tau)+2+\tau)+\gamma_{1,2}^L(\sigma\tau+2\tau+1)+\tau\gamma_{1,2,3}^L\right)(p^L/\tau)+\left(\delta^M(\sigma(1+\tau)+2\tau+1)+\gamma_{1,2}^M\tau(\sigma+2+\tau)+\gamma_{1,2,3}^M\tau^2\right)p^M(\mu/\tau)}{[1+5p^R+(3+2\tau)p^L/\tau+(2+3\tau)p^M(\mu/\tau)]-[2\tau p^D/(1-\tau)+3\tau p^M((1-\mu)/(1-\tau))]}$$

$$-\pi\frac{\delta^D p^D(\tau/(1-\tau))(\sigma+1)+(\tau(\sigma+2)\delta^M+\tau^2\gamma_{1,2}^M)p^M((1-\mu)/(1-\tau))}{[1+5p^R+(3+2\tau)p^L/\tau+(2+3\tau)p^M(\mu/\tau)]-[2\tau p^D/(1-\tau)+3\tau p^M((1-\mu)/(1-\tau))]}$$

$$+\pi\frac{[(\sigma\bar{\theta}^{2,L}+\bar{\theta}^{3,L})(p^L/\tau)+(\sigma\bar{\theta}^{2,M}+\bar{\theta}^{3,M})p^M(\mu/\tau)]-[(\sigma\bar{\theta}^{2,D}+\bar{\theta}^{3,D})p^D/(1-\tau)+(\sigma\bar{\theta}^{2,M}+\bar{\theta}^{3,M})p^M((1-\mu)/(1-\tau))]}{[1+5p^R+(3+2\tau)p^L/\tau+(2+3\tau)p^M(\mu/\tau)]-[2\tau p^D/(1-\tau)+3\tau p^M((1-\mu)/(1-\tau))]}$$

$$=\frac{ITT_{stud}(3)}{[1+5p^R+(3+2\tau)p^L/\tau+(2+3\tau)p^M(\mu/\tau)-[2\tau p^D/(1-\tau)+3\tau p^M((1-\mu)/(1-\tau))]}$$

IV. How Does These Treatment Effects Simplify when There Are No Likers or Dislikers?

Basically, p^L and p^D both = 0, so $p^R + p^M = 1$.

1. **ATE$_{\text{tchr}}$**

$$\text{ATE}_{\text{tchr}}(1) \equiv E[y_1^1 - y_0^1] = \bar{\delta}, \text{ where } \bar{\delta} = \delta^R p^R + \delta^M p^M \qquad \text{(B.36)}$$

$$\text{ATE}_{\text{tchr}}(2) \equiv E[y_1^2 - y_0^2] = 2\bar{\delta} + \bar{\gamma}_{1,2}, \text{ where } \bar{\gamma}_{1,2} = \gamma_{1,2}^R p^R + \gamma_{1,2}^M p^M \qquad \text{(B.37)}$$

$$\text{ATE}_{\text{tchr}}(3) \equiv E[y_1^3 - y_0^3] = 3\bar{\delta} + 3\bar{\gamma}_{1,2} + \bar{\gamma}_{1,2,3}, \text{ where } \bar{\gamma}_{1,2,3} = \gamma_{1,2,3}^R p^R + \gamma_{1,2,3}^M p^M \qquad \text{(B.38)}$$

2. $\mathbf{ITT_{tchr}}$

$$ITT_{tchr}(1) \equiv E[y^1 | R_{tchr,\ year\ 1} = 1] - E[y^1 | R_{tchr,\ year\ 1} = 0] = \bar{\delta}, \text{ where } \bar{\delta} = \delta^R p^R + \delta^M p^M \quad (B.39)$$

$$ITT_{tchr}(2) \equiv E[y^2 | R_{tchr,\ year\ 1} = 1] - E[y^2 | R_{tchr,\ year\ 1} = 0] = \bar{\delta} + p^R(\delta^R + \gamma^R_{1,2}) + p^M \tau \gamma^M_{1,2} \quad (B.40)$$

$$ITT_{tchr}(3) \equiv E[y^3 | R_{tchr,\ year\ 1} = 1] - E[y^3 | R_{tchr,\ year\ 1} = 0] = \bar{\delta} + p^R(2\delta^R + 3\gamma^R_{1,2} + \gamma^R_{1,2,3}) + p^M(2\tau\gamma^M_{1,2} + \tau^2\gamma^M_{1,2,3}) \quad (B.41)$$

3. Average Causal Response (ACR)/Local Average Treatment Effect (LATE)

$$ACR_{tchr}(1) \equiv E[y^1_1 - y^1_0 | T^1_1 \geq 1 > T^1_0] \frac{\text{Prob}[T^1_1 \geq 1 > T^1_0]}{\text{Prob}[T^1_1 \geq 1 > T^1_0]} = \bar{\delta}, \text{ where } \bar{\delta} = \delta^R p^R + \delta^M p^M \quad (B.42)$$

$$ACR_{tchr}(2) \equiv E[y^2_1 - y^2_0 | T^2_1 \geq 1 > T^2_0] \frac{\text{Prob}[T^2_1 \geq 1 > T^2_0]}{\text{Prob}[T^2_1 \geq 1 > T^2_0] + \text{Prob}[T^2_1 = 2 > T^2_0]} \quad (B.43)$$

$$+ E[y^2_2 - y^2_1 | T^2_1 = 2 > T^2_0] \frac{\text{Prob}[T^2_1 = 2 > T^2_0]}{\text{Prob}[T^2_1 \geq 1 > T^2_0] + \text{Prob}[T^2_1 = 2 > T^2_0]}$$

$$= [\bar{\delta} + p^R(\delta^R + \gamma^R_{1,2}) + p^M \tau \gamma^M_{1,2}]/(1 + p^R) = ITT_{tchr}(2)/(1 + p^R)$$

$$ACR_{teachers,3\ years} \equiv E[y^3_1 - y^3_0 | T^3_1 \geq 1 > T^3_0] \frac{\text{Prob}[T^3_1 \geq 1 > T^3_0]}{\text{Prob}[T^3_1 \geq 1 > T^3_0] + \text{Prob}[T^3_1 = 2 > T^3_0] + \text{Prob}[T^3_1 = 3 > T^3_0]} \quad (B.44)$$

$$+ E[y^3_2 - y^3_1 | T^3_1 \geq 2 > T^3_0] \frac{\text{Prob}[T^3_1 \geq 2 > T^3_0]}{\text{Prob}[T^3_1 \geq 1 > T^3_0] + \text{Prob}[T^3_1 = 2 > T^3_0] + \text{Prob}[T^3_1 = 3 > T^3_0]}$$

$$+ E[y^3_3 - y^3_2 | T^3_1 \geq 3 > T^3_0] \frac{\text{Prob}[T^3_1 \geq 3 > T^3_0]}{\text{Prob}[T^3_1 \geq 1 > T^3_0] + \text{Prob}[T^3_1 = 2 > T^3_0] + \text{Prob}[T^3_1 = 3 > T^3_0]}$$

$$= [\bar{\delta} + p^R(2\delta^R + 3\gamma^R_{1,2} + \gamma^R_{1,2,3}) + p^M(2\tau\gamma^M_{1,2} + \tau^2\gamma^M_{1,2,3})]/(1 + 2p^R) = ITT_{teachers,3\ years}/(1 + 2p^R)$$

4. **ATE_{sch}**

$$\text{ATE}_{sch}(1) \equiv E[y^1 | R = 1] - E[y^1 | \text{Program does not exist}] \quad \text{(B.45)}$$

$$= \bar{\delta}, \text{ where } \bar{\delta} = \delta^R p^R + \delta^M p^M$$

$$\text{ATE}_{sch}(2) \equiv E[y^2 | R = 1] - E[y^2 | \text{Program does not exist}] \quad \text{(B.46)}$$

$$= (2\delta^R + \gamma^R_{1,2})p^R + ((1+\tau)\delta^M + \tau\gamma^M_{1,2})p^M$$

Note that there is no composition effect, which also implies that $\mu = \tau$.

$$\text{ATE}_{sch}(3) \equiv E[y^3 | R = 1] - E[y^3 | \text{Program does not exist}] \quad \text{(B.47)}$$

$$= (3\delta^R + 3\gamma^R_{1,2} + \gamma^R_{1,2,3})p^R + [(1+2\tau)\delta^M + \tau(2+\tau)\gamma^M_{1,2} + \gamma^M_{1,2,3}\tau^2]p^M$$

Note that there is no composition effect, which also implies that $\mu = \tau$.

5. **ITT_{sch}**

$$\text{ITT}_{sch}(1) \equiv E[y^1 | R = 1] - E[y^1 | R = 0] = \bar{\delta}, \text{ where } \bar{\delta} = \delta^R p^R + \delta^M p^M \quad \text{(B.48)}$$

$$\text{ITT}_{sch}(2) \equiv E[y^2 | R = 1] - E[y^2 | R = 0] \quad \text{(B.49)}$$

$$= (2\delta^R + \gamma^R_{1,2})p^R + (\delta^M + \gamma^M_{1,2}\tau)p^M$$

Again, there is no composition effect, which again implies that $\mu = \tau$.

$$\text{ITT}_{sch}(3) \equiv E[y^3 | R = 1] - E[y^3 | R = 0] \quad \text{(B.50)}$$

$$= (3\delta^R + 3\gamma^R_{1,2} + \gamma^R_{1,2,3})p^R + (\delta^M + 2\tau\gamma^M_{1,2} + \tau^2\gamma^M_{1,2,3})p^M$$

Again, there is no composition effect, which again implies that $\mu = \tau$.

6. **ATE_{stud}**

$$\text{ATE}_{\text{stud}}(1) \equiv E[s^1 | R = 1] - E[s^1 | \text{Program does not exist}] \quad \text{(B.51)}$$

$$= \pi\bar{\delta}, \text{ where } \bar{\delta} = \delta^R p^R + \delta^M p^M$$

$$\text{ATE}_{\text{stud}}(2) \equiv E[s^2 | R = 1] - E[s^2 | \text{Program does not exist}] \quad \text{(B.52)}$$

$$= \sigma\pi\bar{\delta} + \pi[(2\delta^R + \gamma^R_{1,2})p^R + ((1+\tau)\delta^M + \tau\gamma^M_{1,2})p^M]$$

Again, there is no composition effect, which again implies that $\mu = \tau$.

$$\text{ATE}_{\text{stud}}(3) \equiv E[s^3 | R = 1] - E[s^3 | \text{Program does not exist}] \quad \text{(B.53)}$$

$$= \sigma^2\pi\bar{\delta} + \sigma\pi[(2\delta^R + \gamma^R_{1,2})p^R + ((1+\tau)\delta^M + \tau\gamma^M_{1,2})p^M]$$

$$+ \pi[(3\delta^R + 3\gamma^R_{1,2} + \gamma^R_{1,2,3})p^R + ((1+2\tau)\delta^M + \tau(2+\tau)\gamma^M_{1,2} + \gamma^M_{1,2,3}\tau^2)p^M]$$

Again, there is no composition effect, which again implies that $\mu = \tau$.

7. **ITT_{stud}**

$$\text{ITT}_{\text{stud}}(1) \equiv E[s^1 | R = 1] - E[s^1 | R = 0] = \pi\bar{\delta}, \text{ where } \bar{\delta} = \delta^R p^R + \delta^M p^M \quad \text{(B.54)}$$

$$\text{ITT}_{\text{stud}}(2) \equiv E[s^2 | R = 1] - E[s^2 | R = 0] \quad \text{(B.55)}$$

$$= \sigma\pi\bar{\delta} + \pi[(2\delta^R + \gamma^R_{1,2})p^R + (\delta^M + \gamma^M_{1,2}\tau)p^M]$$

Again, no composition effect, which again implies that $\mu = \tau$.

$$\text{ITT}_{\text{stud}}(3) \equiv E[s^3 | R = 1] - E[s^3 | R = 0] \quad \text{(B.56)}$$

$$= \sigma^2\pi\bar{\delta} + \sigma\pi[(2\delta^R + \gamma^R_{1,2})p^R + (\delta^M + \gamma^M_{1,2}\tau)p^M]$$

$$+ \pi[(3\delta^R + 3\gamma^R_{1,2} + \gamma^R_{1,2,3})p^R + (\delta^M + 2\tau\gamma^M_{1,2} + \gamma^M_{1,2,3}\tau^2)p^M]$$

Again, there is no composition effect, which again implies that $\mu = \tau$.

8. ACR_{stud}

$$\text{ACR}_{\text{stud}}(1) \equiv \frac{E[s^1|R=1] - E[s^1|R=0]}{E[h_{tchr}(1)|R=1] - E[h_{tchr}(1)|R=0]} = \pi\bar{\delta}, \text{ where } \bar{\delta} = \delta^R p^R + \delta^M p^M \quad \text{(B.57)}$$

$$\text{ACR}_{\text{stud}}(2) \equiv \frac{E[s^2|R=1] - E[s^2|R=0]}{E[h_{tchr}(2)|R=1] - E[h_{tchr}(2)|R=0]} \quad \text{(B.58)}$$

$$= \frac{\sigma\pi\bar{\delta} + \pi[(2\delta^R + \gamma^R_{1,2})p^R + (\delta^M + \gamma^M_{1,2}\tau)p^M]}{1 + 2p^R + p^M}$$

$$= \frac{\text{ITT}_{\text{stud}}(2)}{1 + 2p^R + p^M}$$

$$\text{ACR}_{\text{stud}}(3) \equiv \frac{E[s^3|R=1] - E[s^3|R=0]}{E[h_{tchr}(3)|R=1] - E[h_{tchr}(3)|R=0]} \quad \text{(B.59)}$$

$$= \pi\frac{\sigma^2\bar{\delta} + \left((3+2\sigma)\delta^R + (3+\sigma)\gamma^R_{1,2} + \gamma^R_{1,2,3}\right)p^R + (\delta^M(\sigma+1) + \gamma^M_{1,2}\tau(\sigma+2) + \gamma^M_{1,2,3}\tau^2)p^M}{1 + 5p^R + 2p^M}$$

$$= \frac{\text{ITT}_{\text{stud}}(3)}{1 + 5p^R + 2p^M}$$

V. What Do OLS and IV Regressions Estimate?

Now turn to what we are able to estimate, starting with regressions based on the teacher skill data and then turning to student test scores.

1. Applied to Teacher Skill Variables

We have two samples of teachers, one that (imperfectly) follows the teachers who were in APM and non-APM schools in year 1 (Sample 1), and one that focuses on the teachers who are in the APM and non-APM schools in any given year (Sample 2). Start with the Sample 1 teachers.

OLS Applied to Sample 1 Teachers

The OLS estimate for Sample 1 applied to year t, can be denoted by $\hat{\beta}_1^y{}_{\text{OLS, year t}}$, where the "1" subscript indicates Sample 1 teachers. Regressing teacher skills on a constant and a variable for assignment to an APM school in year 1 yields (for any value of t) the OLS estimate $\hat{\beta}_1^y{}_{\text{OLS, year t}}$:

$$\hat{\beta}_1^y{}_{\text{OLS, year t}} = E[y^t | R_{\text{tchr, year 1}} = 1] - E[y^t | R_{\text{tchr, year 1}} = 0] \qquad \text{(B.60)}$$

Start with year 1 (we do not have the data, but we show for completeness). Using (B.11), and noting that teachers follow their random assignment in year 1, we have:

$$\hat{\beta}_1^y{}_{\text{OLS,t=1}} = E[y^1 | R_{\text{tchr, year 1}} = 1] - E[y^1 | R_{\text{tchr, year 1}} = 0] = \bar{\delta} \qquad \text{(B.61)}$$

So, if we have data on teacher skills at the end of year 1 (which, unfortunately, we do not have), we can estimate all the teacher skill treatment effects that we described above for year 1 ($ATE_{tchr}(1)$, $ITT_{tchr}(1)$, $ACR_{tchr}(1)$ $ATE_{sch}(1)$, and $ITT_{sch}(1)$), because these are all equal to $\bar{\delta}$ due to perfect compliance in year 1.

Next, turn to year 2. The OLS estimate for Sample 1 is:

$$\hat{\beta}_1^y{}_{\text{OLS,t=2}} = E[y^2 | R_{\text{tchr, year 1}} = 1] - E[y^2 | R_{\text{tchr, year 1}} = 0] \qquad \text{(B.62)}$$

Using (B.12) we have:

$$\hat{\beta}_1^y{}_{OLS,t=2} = E[y^2|\ R_{tchr,\ year\ 1} = 1] - E[y^2|\ R_{tchr,\ year\ 1} = 0] \qquad (B.63)$$

$$= \bar{\delta} + p^R(\delta^R + \gamma_{1,2}^R) + p^L\gamma_{1,2}^L + p^M\tau\gamma_{1,2}^M$$

This equals to $ITT_{tchr}(2)$, so we can estimate $ITT_{tchr}(2)$ by applying OLS to the Sample 1 teachers in year 2.

Next, turn to year 3. The OLS estimate for Sample 1 is:

$$\hat{\beta}_1^y{}_{OLS,t=3} = E[y^3|\ R_{tchr,\ year\ 1} = 1] - E[y^3|\ R_{tchr,\ year\ 1} = 0] \qquad (B.64)$$

Again, for Sample 1, we are following the same teachers over time, so their proportions do not change. Using (B.13), we have:

$$\hat{\beta}_1^y{}_{OLS,t=3} = E[y^3|\ R_{tchr,\ year\ 1} = 1] - E[y^3|\ R_{tchr,\ year\ 1} = 0] \qquad (B.65)$$

$$= \bar{\delta} + p^R(2\delta^R + 3\gamma_{1,2}^R + \gamma_{1,2,3}^R) + p^L(2\gamma_{1,2}^L + \gamma_{1,2,3}^L) + p^M(2\tau\gamma_{1,2}^M + \tau^2\gamma_{1,2,3}^M)$$

This equals to $ITT_{tchr}(3)$, so we can estimate $ITT_{tchr}(3)$ by applying OLS to the Sample 1 teachers in year 3.

OLS Applied to Sample 2 Teachers

The OLS estimate of the impact of the APM program on the skills of Sample 2 teachers in year t, which can be denoted as $\hat{\beta}_2^y{}_{OLS,\ year\ t}$, is equal to $E[y^t|\ R = 1] - E[y^t|\ R = 0]$, where R refers to the random assignment (in year 1) of the school in which the teacher is in year t. That is, it compares the teachers who are in treated and control schools in year t, regardless of their random assignment (regardless of the schools in which they were teaching) in year 1.

For year 1, this is the same as OLS applied to Sample 1 teachers, since $y^t = y^1$ and $R = R_{\text{tchr, year 1}}$, so there is no need to show this again.

Next, turn to year 2. The OLS estimate for Sample 2 is:

$$\hat{\beta}_2^y{}_{\text{OLS,t=2}} = E[y^2 | R = 1] - E[y^2 | R = 0] \qquad \text{(B.66)}$$

For Sample 2, the proportions of teachers who are in the APM and non-APM schools will change, so we need to account for that. All likers will move to APM schools and all dislikers will move to non-APM schools. So the proportion of remainer, liker and mover teachers in APM schools will be p^R (no change), p^L/τ and $p^M(\mu/\tau)$, where μ is the proportion of all movers who end up in APM schools. Similarly, the proportion of remainer, disliker and mover teachers in non-APM schools will be p^R (no change), $p^D/(1\text{-}\tau)$ and $p^M((1\text{-}\mu)/(1\text{-}\tau))$.

To calculate $\hat{\beta}_2^y{}_{\text{OLS,t=2}}$, start with $E[y^2| R = 1]$:

$$E[y^2| R = 1] = \bar{\theta}^{2,R}p^R + \bar{\theta}^{2,L}(p^L/\tau) + \bar{\theta}^{2,M}p^M(\mu/\tau) \qquad \text{(B.67)}$$

$$+ p^R(2\delta^R + \gamma^R_{1,2}) + (p^L/\tau)(\delta^L(1+\tau) + \tau\gamma^L_{1,2}) + (p^M(\mu/\tau))(\delta^M(1+\tau) + \tau\gamma^M_{1,2})$$

Next, work out $E[y^2| R = 0]$:

$$E[y^2| R = 0] = \bar{\theta}^{2,R}p^R + \bar{\theta}^{2,D}(p^D/(1\text{-}\tau) + \bar{\theta}^{2,M}p^M((1\text{-}\mu)/(1\text{-}\tau)) + (p^D/(1\text{-}\tau))\tau\delta^D + p^M((1\text{-}\mu)/(1\text{-}\tau))\tau\delta^M \quad \text{(B.68)}$$

Equations (B.67) and (B.68) can then be used to obtain $\hat{\beta}_2^y{}_{\text{OLS,t=2}}$:

$$\hat{\beta}_2^y{}_{\text{OLS,t=2}} = E[y^2| R = 1] - E[y^2| R = 0] \qquad \text{(B.69)}$$

$$= p^R(2\delta^R + \gamma^R_{1,2}) + (p^L/\tau)(\delta^L(1+\tau) + \tau\gamma^L_{1,2}) + (p^M(\mu/\tau))(\delta^M(1+\tau) + \tau\gamma^M_{1,2}) - [(p^D/(1\text{-}\tau))\tau\delta^D + p^M((1\text{-}\mu)/(1\text{-}\tau))\tau\delta^M]$$

$$+ \bar{\theta}^{2,L}(p^L/\tau) + \bar{\theta}^{2,M}p^M(\mu/\tau) - [\bar{\theta}^{2,D}(p^D/(1\text{-}\tau) + \bar{\theta}^{2,M}p^M((1\text{-}\mu)/(1\text{-}\tau))]$$

$$= p^R(2\delta^R + \gamma^R_{1,2}) + (p^L/\tau)(\delta^L(1+\tau) + \tau\gamma^L_{1,2}) - (p^D/(1\text{-}\tau))\tau\delta^D + p^M[\delta^M(\mu\text{-}\tau^2)/(\tau\text{-}\tau^2) + \mu\gamma^M_{1,2}]$$

$$+ \bar{\theta}^{2,L}(p^L/\tau) - \bar{\theta}^{2,D}(p^D/(1\text{-}\tau) + \bar{\theta}^{2,M}p^M[(\mu/\tau) - ((1\text{-}\mu)/(1\text{-}\tau))]$$

This is equal to $ITT_{sch}(2)$, which means that we can estimate $ITT_{sch}(2)$ by applying OLS to the Sample 2 teachers in year 2.

Finally, turn to year 3. The OLS estimate for Sample 2 teachers in year 3 is:

$$\hat{\beta}^y_{2\ OLS,t=3} = E[y^3 | R = 1] - E[y^3 | R = 0] \qquad (B.70)$$

This is similar to year 2, except we need to account for the fact that movers can move again between years 2 and 3. However, the proportions of the 4 types of teachers are the same as in year 3, we just have to adjust for 3 types of movers.

To calculate $\hat{\beta}^y_{2\ OLS,t=3}$, start with $E[y^3| R = 1]$:

$$E[y^3| R = 1] = \bar{\theta}^{3,R}p^R + \bar{\theta}^{3,L}(p^L/\tau) + \bar{\theta}^{3,M}p^M(\mu/\tau) \qquad (B.71)$$

$$+ p^R(3\delta^R + 3\gamma^R_{1,2} + \gamma^R_{1,2,3}) + (p^L/\tau)(\delta^L(2+\tau) + (2\tau+1)\gamma^L_{1,2} + \tau\gamma^L_{1,2,3}) + (p^M(\mu/\tau))(\delta^M(1+2\tau) + \tau(2+\tau)\gamma^M_{1,2} + \tau^2\gamma^M_{1,2,3})$$

Then calculate $E[y^3| R = 0]$:

$$E[y^3| R = 0] = \bar{\theta}^{2,R}p^R + \bar{\theta}^{2,D}(p^D/(1\text{-}\tau)) + \bar{\theta}^{2,M}p^M((1\text{-}\mu)/(1\text{-}\tau)) \qquad (B.72)$$

$$+ (p^D/(1\text{-}\tau))\tau\delta^D + p^M((1\text{-}\mu)/(1\text{-}\tau))(\delta^M 2\tau + \tau^2\gamma^M_{1,2})$$

Equations (B.71) and (B.72) can then be used to obtain $\hat{\beta}^y_{2\ OLS,t=3}$:

$$\hat{\beta}_{2}^{y}{}_{\text{OLS,t=3}} = E[y^3 | R = 1] - E[y^3 | R = 0] \quad \text{(B.73)}$$

$$= p^R(3\delta^R + 3\gamma^R_{1,2} + \gamma^R_{1,2,3}) + (p^L/\tau)(\delta^L(2+\tau) + (2\tau+1)\gamma^L_{1,2} + \tau\gamma^L_{1,2,3}) + (p^M(\mu/\tau))(\delta^M(1+2\tau) + \tau(2+\tau)\gamma^M_{1,2} + \tau^2\gamma^M_{1,2,3})$$

$$- [(p^D/(1-\tau))\tau\delta^D + p^M((1-\mu)/(1-\tau))(\delta^M 2\tau + \tau^2\gamma^M_{1,2})]$$

$$\bar{\theta}^{3,L}(p^L/\tau) + \bar{\theta}^{3,M}p^M(\mu/\tau) - [\bar{\theta}^{3,D}(p^D/(1-\tau) + \bar{\theta}^{3,M}p^M((1-\mu)/(1-\tau))]$$

The first two lines are the (net) treatment effect, and the last line is the composition effect.

This is equal to $ITT_{sch}(3)$, which means that we can estimate $ITT_{sch}(3)$ by applying OLS to the Sample 2 teachers in year 3.

IV Applied to Sample 1 Teachers

Consider a simple IV estimation. The equation of interest is the impact of years of participation in APM on teacher skills. We can write this equation as follows, where $T_j^{Tot,t}$ is the number of years that teacher j has been exposed to the program at time t:

$$y_j^t = \beta T_j^{Tot,t} + u_j \quad \text{(B.74)}$$

The first stage regression is random assignment to an APM school in year 1:

$$T_j^{Tot,t} = \alpha R_{tchr,\, year\ 1,\, j} + v_j \quad \text{(B.75)}$$

where $R_{tchr,\, year\ 1,\, j}$ denotes teacher j's random assignment in year 1. Simple IV regression of these equations estimates β as follows:

$$\hat{\beta}_{1}^{y}{}_{\text{IV, year t}} = \frac{Cov(y^t, R_{tchr,year\ 1})}{Cov(T^{Tot,t}, R_{tchr,year\ 1})} = \frac{E[y^t|R_{tchr,year\ 1}=1] - E[y^t|R_{tchr,year\ 1}=0]}{E[T^{Tot,t}|R_{tchr,year\ 1}=1] - E[T^{Tot,t}|R_{tchr,year\ 1}=0]} \quad \text{(B.76)}$$

where the second equality follows from the definition of covariance and the fact that R equals either 0 or 1. Note that $\hat{\beta}_1^y{}_{\text{IV, year t}}$ estimates a "per year" effect of the treatment; the cumulative effect is obtained by multiplying this by the years of coaching induced by a school's random assignment to APM (the denominator in (B.76)), which yields $ITT_{tchr}(t)$ (and which can be obtained by running an OLS regression of teacher schools in year t on a constant term and a dummy variable indicating the random assignment of the school where the teacher was in year 1).

For year 1, applying this is straightforward. Since all teachers follow their random assignment in year 1, the denominator of $\hat{\beta}_1^y{}_{\text{IV,t=1}}$ is 1, The numerator can be obtained from the derivations for ITT given in Section II, which implies that:

$$\hat{\beta}_1^y{}_{\text{IV,t=1}} = E[y^1 |\, R_{\text{tchr, year 1}} = 1] - E[y^1 |\, R_{\text{tchr, year 1}} = 0] = \bar{\delta} \qquad \text{(B.77)}$$

This is equal to all the treatment effects defined in Section II, including $ACR_{tchr}(1)$.

For year 2, we need to estimate:

$$\hat{\beta}_1^y{}_{\text{IV,t=2}} = \frac{E[y^2 | R_{\text{tchr ,year 1}}=1] - E[y^2 | R_{\text{tchr ,year 1}}=0]}{E[T^{\text{Tot},2} | R_{\text{tchr ,year 1}}=1] - E[T^{\text{Tot},2} | R_{\text{tchr ,year 1}}=0]} \qquad \text{(B.78)}$$

The numerator can be obtained from equation (B.12):

$$E[y^2 |\, R_{\text{tchr, year 1}} = 1] - E[y^2 |\, R_{\text{tchr, year 1}} = 0] = \bar{\delta} + p^R(\delta^R + \gamma_{1,2}^R) + p^L\gamma_{1,2}^L + p^M\tau\gamma_{1,2}^M \qquad \text{(B.79)}$$

The denominator is straightforward to calculate:

$$E[T^{\text{Tot},2} |\, R_{\text{tchr, year 1}} = 1] - E[T^{\text{Tot},2} |\, R_{\text{tchr, year 1}} = 0] \qquad \text{(B.80)}$$

$$= (2p^R + 2p^L + p^D + p^M(2\tau + (1\text{-}\tau)) - (0p^R + p^L + 0p^D + \tau p^M)$$

$$= 2p^R + p^L + p^D + p^M = 1 + p^R$$

Thus $\hat{\beta}_1^y{}_{IV,t=2} = (\bar{\delta} + p^R(\delta^R+\gamma^R_{1,2}) + p^L\gamma^L_{1,2} + p^M\tau\gamma^M_{1,2})/(1 + p^R)$, and so it estimates $ACR_{tchr}(2)$.

For year 3, we need to estimate:

$$\hat{\beta}_1^y{}_{IV,\ t=3} = \frac{E[y^3|R_{tchr\ ,year\ 1}=1] - E[y^3|R_{tchr\ ,year\ 1}=0]}{E[T^{Tot,3}|R_{tchr\ ,year\ 1}=1] - E[T^{Tot,3}|R_{tchr\ ,year\ 1}=0]} \qquad (B.81)$$

The numerator can be obtained from equation (B.13):

$$E[y^3|\ R_{tchr,\ year\ 1} = 1] - E[y^3|\ R_{tchr,\ year\ 1} = 0] \quad (B.82)$$

$$= \bar{\delta} + p^R(2\delta^R+3\gamma^R_{1,2}+\gamma^R_{1,2,3}) + p^L(2\gamma^L_{1,2}+\gamma^L_{1,2,3}) + p^M(2\tau\gamma^M_{1,2}+\tau^2\gamma^M_{1,2,3})$$

The denominator is again straightforward to calculate:

$$E[T^{Tot,3}|\ R_{tchr,\ year\ 1} = 1] - E[T^{Tot,3}|\ R_{tchr,\ year\ 1} = 0] \qquad (B.83)$$

$$= (3p^R + 3p^L + p^D + p^M(3\tau^2 + 4\tau(1\text{-}\tau) + (1\text{-}\tau)^2)) - (0p^R + 2p^L + 0p^D + p^M(2\tau^2 + 2\tau(1\text{-}\tau) + 0(1\text{-}\tau)^2))$$

$$= 3p^R + p^L + p^D + p^M = 1 + 2p^R$$

Thus, $\hat{\beta}_1^y{}_{IV,t=3} = (\bar{\delta} + p^R(2\delta^R+3\gamma^R_{1,2}+\gamma^R_{1,2,3}) + p^L(2\gamma^L_{1,2}+\gamma^L_{1,2,3}) + p^M(2\tau\gamma^M_{1,2}+\tau^2\gamma^M_{1,2,3}))/(1 + 2p^R)$, so $\hat{\beta}_1^y{}_{IV,t=3}$ estimates $ACR_{tchr}(3)$.

IV Applied to Sample 2 Teachers

If we have an "open" system, it is not possible to apply IV to Sample 2 teachers because some of them will come from outside of our randomization sample and thus the instrumental variable (random assignment) does not exist for some of those teachers. While one could argue that such teachers could be treated as not randomly assigned to the treatment groups, so that $R^1 = 0$, I do not think that this is correct, because teachers who want to move into our set of schools are a (self-)selected group of teachers who may, for example, be attracted by the APM program.

However, if our system is “closed”, so that we do have a valid IV for all Sample 2 teachers, it is possible to apply IV estimation to Sample 2. For year 1, as always all teachers follow their random assignment and this could again estimate $\bar{\delta}$, so $\beta_{2,IV,year\ 1} = \bar{\delta}$.

For year 2, we simply use the same approach as for Sample 1 teachers, except that $R_{tchr,\ year\ 1}$ is replaced by R; that is, the focus is on the teachers in the APM and non-APM schools in year t; not the teachers who were on those schools in year 1. For the teachers in Sample 2, IV estimates:

$$\beta_{2,IV,year\ 2} = \frac{E[y^2|R=1] - E[y^2|R=0]}{E[T^{Tot,2}|R=1] - E[T^{Tot,2}|R=0]} \qquad (B.84)$$

We simply need to work this out for the Sample 2 teachers. The defining characteristics of those teachers is where they were in year 2. Those in the treated schools in year 2 have $T^2 = 1$, and those in the control schools have $T^2 = 0$. Thus, the numerator of the above expression was derived in equation (B.69), which we show again:

$$E[y^2|\ R = 1] - E[y^2|\ R = 0] \qquad (B.69)$$

$$= p^R(2\delta^R + \gamma^R_{1,2}) + (p^L/\tau)(\delta^L(1+\tau) + \tau\gamma^L_{1,2}) - (p^D/(1-\tau))\tau\delta^D + p^M[\delta^M(\mu-\tau^2)/(\tau-\tau^2) + \mu\gamma^M_{1,2})$$

$$+ \bar{\theta}^{2,L}(p^L/\tau) + \bar{\theta}^{2,M}p^M(\mu/\tau) - [\bar{\theta}^{2,D}(p^D/(1-\tau) + \bar{\theta}^{2,M}p^M((1-\mu)/(1-\tau))]$$

Next, consider the denominator for this IV estimate:

$$E[T^{Tot,2}|\ R = 1] - E[T^{Tot,2}|\ R = 0] \qquad (B.85)$$

$$= p^R 2 + (p^L/\tau)[2\tau + (1\text{-}\tau)] + (p^M(\mu/\tau))[2\tau + (1\text{-}\tau)] - [(p^D/(1\text{-}\tau))\tau + p^M((1\text{-}\mu)/(1\text{-}\tau))\tau]$$

$$= 2p^R + (p^L/\tau)(1+\tau) + p^M(\mu\text{-}\tau^2)/(\tau\text{-}\tau^2) - (p^D/(1\text{-}\tau))\tau$$

Combining the numerator and denominator for the IV estimate for Sample 2, for a closed system, yields:

$$\beta_{2,IV,year\ 2} \qquad (B.86)$$

$$= \frac{p^R(2\delta^R+\gamma^R_{1,2})+(p^L/\tau)(\delta^L(1+\tau)+\tau\gamma^L_{1,2}) - (p^D/(1-\tau))\tau\delta^D + p^M[\delta^M(\mu-\tau^2)/(\tau-\tau^2) + \mu\gamma^M_{1,2}) + \bar{\theta}^{2,L}(p^L/\tau) + \bar{\theta}^{2,M}p^M(\mu-\tau)/(\tau(1-\tau)) - \bar{\theta}^{2,D}p^D/(1-\tau)}{2p^R + (p^L/\tau)(1+\tau) + p^M(\mu-\tau^2)/(\tau-\tau^2) + - (p^D/(1-\tau))\tau}$$

Unlike OLS estimation for year 2 (for a "closed" or "open" system), IV estimation for Sample 2 teachers in year 2 for a "closed" system is exactly equal to IV estimation for Sample 1 teachers in year 2. The intuition is that, in a closed system, Sample 1 and Sample 2 teachers are the same population of teachers and are equally distributed in APM and non-APM schools when random assignment, the defining feature of IV estimation, was done in year 1. In contrast, OLS estimates for Sample 2 (but not Sample 1) teachers are defined in terms of where teachers were in year 2.

For year 3, the IV estimate for Sample 2 teachers for a closed system is:

$$\beta_{2,IV,year\ 3} = \frac{E[y^3|R=1] - E[y^3|R=0]}{E[T^{Tot,3}|R=1] - E[T^{Tot,3}|R=0]} \qquad (B.87)$$

The numerator is from equation (B.73), which is:

$$E[y^3|\ R = 1] - E[y^3|\ R = 0] \qquad (B.73)$$

$$= p^R(3\delta^R + 3\gamma^R_{1,2} + \gamma^R_{1,2,3}) + (p^L/\tau)(\delta^L(2+\tau) + (2\tau+1)\gamma^L_{1,2} + \tau\gamma^L_{1,2,3}) + (p^M(\mu/\tau))(\delta^M(1+2\tau) + \tau(2+\tau)\gamma^M_{1,2} + \tau^2\gamma^M_{1,2,3})$$

$$- [(p^D/(1\text{-}\tau))\tau\delta^D + p^M((1\text{-}\mu)/(1\text{-}\tau))(\delta^M 2\tau + \tau^2\gamma^M_{1,2})]$$

$$\bar{\theta}^{3,L}(p^L/\tau) + \bar{\theta}^{3,M}p^M(\mu/\tau) - [\bar{\theta}^{3,D}(p^D/(1\text{-}\tau) + \bar{\theta}^{3,M}p^M((1\text{-}\mu)/(1\text{-}\tau))]$$

The denominator is:

$$E[T^{Tot,2}|\ R = 1] - E[T^{Tot,2}|\ R = 0] \qquad \text{(B.88)}$$

$$= p^R 3 + (p^L/\tau)(3\tau + 2(1\text{-}\tau)]) + (p^M(\mu/\tau))[3\tau^2 + 2\times 2\tau(1\text{-}\tau) + (1\text{-}\tau)^2] - [(p^D/(1\text{-}\tau))\tau + p^M((1\text{-}\mu)/(1\text{-}\tau))2\tau]$$

$$= 3p^R + (p^L/\tau)(2\text{+}\tau) + p^M(\mu\text{+}\mu\tau\text{-}2\tau^2)/(\tau\text{-}\tau^2) - (p^D/(1\text{-}\tau))\tau$$

Combining the numerator and the denominator gives, for Sample 2 for a closed system:

$$\beta_{2,\text{IV,year 3}} \qquad \text{(B.89)}$$

$$= \frac{p^R\left(3\delta^R+3\gamma^R_{1,2}+\gamma^R_{1,2,3}\right)+(p^L/\tau)\left(\delta^L(2+\tau)+(2\tau+1)\gamma^L_{1,2}+\tau\gamma^L_{1,2,3}\right) - (p^D/(1-\tau))\tau\delta^D + (p^M/(\tau-\tau^2))[\delta^M(\mu(1+\tau)-2\tau^2) + \gamma^M_{1,2}(\mu\left(2\tau-\tau^2\right)-\tau^3)]}{3p^R + (p^L/\tau)(2+\tau) + p^M(\mu+\mu\tau-2\tau^2)/(\tau-\tau^2) - (p^D/(1-\tau))\tau}$$

$$+ \frac{\bar{\theta}^{3,L}(p^L/\tau) + \bar{\theta}^{3,M}p^M(\mu-\tau)/(\tau(1-\tau)) - \bar{\theta}^{3,D}p^D/(1-\tau)}{3p^R + (p^L/\tau)(2+\tau) + p^M(\mu+\mu\tau-2\tau^2)/(\tau-\tau^2) - (p^D/(1-\tau))\tau}$$

2. Applied to Student Test Scores

Since students are assumed not to move between schools, there is only one sample of students, who are classified by the schools in which they are enrolled.

An OLS regression of students' test scores in year t on a constant term and the type of school (APM or non-APM) that the student is in that year will yield the following coefficient for the type of school, which we can denote by $\hat{\beta}^s_{OLS}$, year t:

$$\hat{\beta}^s_{OLS,\text{ year t}} = E[s^t|\ R = 1] - E[s^t|\ R = 0] \qquad \text{(B.90)}$$

Consider $\hat{\beta}^{s}_{OLS, \text{ year t}}$ separately for years 1, 2 and 3.

Start with year 1. Applying equation (B.1) yields $s_i^1 = \sigma s_i^{t0} + \pi y_j$ Since teachers are unable to move in year 1, no teachers in non-APM schools are trained and all teachers in APM schools are trained. Thus:

$$E[s^1| R = 1] = \sigma E[s^0| R = 1] + \pi E[y^1| R = 1] \qquad \text{(B.91)}$$

$$= \sigma E[s^0| R = 1] + \pi[(\bar{\theta}^{1,R}p^R + \bar{\theta}^{1,L}p^L + \bar{\theta}^{1,D}p^D + \bar{\theta}^{1,M}p^M) + (\delta^R p^R + \delta^L p^L + \delta^D p^D + \delta^M p^M)]$$

$$E[s^1| R = 0] = \sigma E[s^0| R = 0] + \pi(\bar{\theta}^{1,R}p^R + \bar{\theta}^{1,L}p^L + \bar{\theta}^{1,D}p^D + \bar{\theta}^{1,M}p^M) \qquad \text{(B.92)}$$

$$\hat{\beta}^{s}_{OLS,\, t=1} = E[s^1| R = 1] - E[s^1| R = 0] = \pi(\delta^R p^R + \delta^L p^L + \delta^D p^D + \delta^M p^M) \qquad \text{(B.93)}$$

$$= \pi\bar{\delta}$$

where $\bar{\delta}$ is the population-weighted average of the four δ terms. Note also that $E[s^0| R = 1] = E[s^0| R = 0]$ since R was randomly assigned.

Thus, for year 1 OLS produces an unbiased estimate of both $ATE_{stud}(1)$, which also equals $ITT_{stud}(1)$ and $ACR_{stud}(1)$, which is intuitively plausible since neither teachers nor students move in year 1.

Next, turn to year 2. The OLS estimate of β_2, which we denote by $\hat{\beta}^{s}_{OLS,\, t=2}$, is derived as follows, using equations (B.90), (B.93) and (B.23).

$$\hat{\beta}^{s}_{OLS,\, t=2} = E[s^2| R = 1] - E[s^2| R = 0] \qquad \text{(B.94)}$$

$$= E[\sigma s^1 + \pi y^2| R = 1] - E[\sigma s^1 + \pi y^2| R = 0]$$

$$= \sigma\{E[s^1| R = 1] - E[s^1| R = 0]\} + \pi\{E[y^2| R = 1] - E[y^2| R = 0]\}$$

$$= \sigma\pi\bar{\delta}$$

$$+ \pi[(2\delta^R + \gamma^R_{1,2})p^R + (\delta^L(1+\tau) + \tau\gamma^L_{1,2})(p^L/\tau) - \delta^D\tau(p^D/(1\text{-}\tau)) + ((\delta^M(1+\tau) + \gamma^M_{1,2}\tau)(\mu/\tau) - \delta^M\tau(1\text{-}\mu)/(1\text{-}\tau))p^M$$

$$+ \bar{\theta}^{2,L}(p^L/\tau) - \bar{\theta}^{2,D}(p^D/(1\text{-}\tau)) + \bar{\theta}^{2,M}p^M((\mu/\tau) - (1\text{-}\mu)/(1\text{-}\tau))]$$

The first and second lines are the direct effect on students while the third line is the composition effect due to teachers switching schools. This equals $ITT_{stud}(2)$.

Finally, turn to year 3. The OLS estimate of β_3, which we denote by $\hat{\beta}^s_{OLS,\ t=3}$, is derived as follows, using equations (B.90), (B.94) and (B.23).

$$\hat{\beta}^s_{OLS,\ t=3} = E[s^3 | R = 1] - E[s^3 | R = 0] \qquad (B.95)$$

$$= E[\sigma s^2 + \pi y^3 | R = 1] - E[\sigma s^2 + \pi y^3 | R = 0]$$

$$= \sigma\{E[s^2 | R = 1] - E[s^2 | R = 0]\} + \pi\{E[y^3 | R = 1] - E[y^3 | R = 0]\}$$

$$= \sigma\{\sigma\pi\bar{\delta}$$

$$+ \pi[(2\delta^R + \gamma^R_{1,2})p^R + (\delta^L(1+\tau) + \tau\gamma^L_{1,2})(p^L/\tau) - \delta^D\tau(p^D/(1\text{-}\tau)) + ((\delta^M(1+\tau) + \gamma^M_{1,2}\tau)(\mu/\tau) - \delta^M\tau(1\text{-}\mu)/(1\text{-}\tau))p^M$$

$$+ \bar{\theta}^{2,L}(p^L/\tau) - \bar{\theta}^{2,D}(p^D/(1\text{-}\tau)) + \bar{\theta}^{2,M}p^M((\mu/\tau) - (1\text{-}\mu)/(1\text{-}\tau))]\}$$

$$+ \pi\{(3\delta^R + 3\gamma^R_{1,2} + \gamma^R_{1,2,3})p^R + (\delta^L(2+\tau) + \gamma^L_{1,2}(1+2\tau) + \gamma^L_{1,2,3}\tau)(p^L/\tau) - \delta^D\tau(p^D/(1\text{-}\tau))$$

$$+ [(\mu/\tau)(\delta^M(1+2\tau) + \gamma^M_{1,2}\tau(2+\tau) + \gamma^M_{1,2,3}\tau^2) - ((1\text{-}\mu)/(1\text{-}\tau)(\delta^M 2\tau + \gamma^M_{1,2}\tau^2)]p^M$$

$$+ \bar{\theta}^{3,L}(p^L/\tau) - \bar{\theta}^{3,D}(p^D/(1\text{-}\tau)) + \bar{\theta}^{3,M}p^M((\mu/\tau) - (1\text{-}\mu)/(1\text{-}\tau))\}$$

The first, second, fourth and fifth lines are the direct effect on students, while the third and sixth lines are the composition effect due to teachers switching schools. This equals $ITT_{stud}(3)$.

Finally, consider IV estimates of student test scores. The variable being instrumented is the student's exposure to teachers with APM coaching. More specifically, as explained above, it is the "history" from year 1 to year t of students' exposure to treated teachers, which is denoted by $h_{tchr}(t)$. Random assignment (R) is the instrument. For simple IV estimation with a constant term and no other variables in the first stage and second stage equations, the IV estimate fro year t, denoted by $\hat{\beta}^{s}_{IV,t}$, is:

$$\hat{\beta}^{s}_{IV,t} \equiv \frac{Cov(s^t,R)}{Cov(h_{tchr}(t),R)} \qquad (B.96)$$

$$= \frac{E[s^t|R=1]-E[s^t|R=0]}{E[h_{tchr}(t)|R=1]-E[h_{tchr}(t)|R=0]}$$

where the second line uses the fact that R is a binary variable.

For year 1, we have:

$$\hat{\beta}^{s}_{IV,1} \equiv \frac{E[s^1|R=1]-E[s^1|R=0]}{E[h_{tchr}(1)|R=1]-E[h_{tchr}(1)|R=0]} \qquad (B.97)$$

All teachers follow their random assignment, so the denominator equals 1. The numerator is given in equation (B.29), which implies, as one would expect from full compliance in year 1:

$$\hat{\beta}^{s}_{IV,1} = \pi\bar{\delta} \qquad (B.98)$$

For year 2, the definition in (B.107) gives:

$$\hat{\beta}^{s}_{IV,2} = \frac{E[s^2|R=1]-E[s^2|R=0]}{E[h_{tchr}(2)|R=1]-E[h_{tchr}(2)|R=0]} \qquad (B.99)$$

The derivations above in equation (B.34) show that:

$$\hat{\beta}^{s}_{IV,2} = \frac{\sigma\pi\bar{\delta}+\pi[(2\delta^{R}+\gamma^{R}_{1,2})p^{R}+(\delta^{L}(1+\tau)+\tau\gamma^{L}_{1,2})(p^{L}/\tau) - \delta^{D}\tau(p^{D}/(1-\tau)) + ((\delta^{M}(1+\tau)+\gamma^{M}_{1,2}\tau)(\mu/\tau) - \delta^{M}\tau(1-\mu)/(1-\tau))p^{M}]}{1 + 2p^{R} + (p^{L}/\tau)(\tau+1)+ p^{\mu}(\mu/\tau)(1+\tau)-[\tau p^{D}/(1-\tau) + \tau p^{M}((1-\mu)/(1-\tau))]} \quad \text{(B.100)}$$

$$+ \frac{\pi[\bar{\theta}^{2,L}(p^{L}/\tau) - \bar{\theta}^{2,D}(p^{D}/(1-\tau)) + \bar{\theta}^{2,M}p^{M}((\mu/\tau)-(1-\mu)/(1-\tau))]}{1 + 2p^{R} + (p^{L}/\tau)(\tau+1)+ p^{M}(\mu/\tau)(1+\tau)-[\tau p^{D}/(1-\tau) + \tau p^{M}((1-\mu)/(1-\tau))]}$$

This is equal to ACR$_{stud}$(2).

For year 3, applying the definition in (B.96) yields:

$$\hat{\beta}^{s}_{IV,3} = \frac{E[s^{3}|R=1]- E[s^{3}|R=0]}{E[h_{tchr}(3)|R=1]- E[h_{tchr}(3)|R=0]} \quad \text{(B.101)}$$

The derivations in equation (B.35) show that

$$\hat{\beta}^{s}_{IV,3} = \quad \text{(B.102)}$$

$$\pi\frac{\sigma^{2}\bar{\delta} + \left((3+2\sigma)\delta^{R}+(3+\sigma)\gamma^{R}_{1,2}+\gamma^{R}_{1,2,3}\right)p^{R}+ \left(\delta^{L}(\sigma(1+\tau)+2+\tau)+\gamma^{L}_{1,2}(\sigma\tau+2\tau+1)+\tau\gamma^{L}_{1,2,3}\right)(p^{L}/\tau) + (\delta^{M}(\sigma(1+\tau)+2\tau+1)+\gamma^{M}_{1,2}\tau(\sigma+2+\tau)+\gamma^{M}_{1,2,3}\tau^{2})p^{M}(\mu/\tau)}{[1+5p^{R} +(3+2\tau)p^{L}/\tau +(2+3\tau)p^{M}(\mu/\tau)] - [2\tau p^{D}/(1-\tau) + 3\tau p^{M}((1-\mu)/(1-\tau))]}$$

$$-\pi\frac{\delta^{D}p^{D}(\tau/(1-\tau))(\sigma+1) + (\tau(\sigma+2)\delta^{M}+ \tau^{2}\gamma^{M}_{1,2})p^{M}((1-\mu)/(1-\tau))}{[1+5p^{R} +(3+2\tau)p^{L}/\tau +(2+3\tau)p^{M}(\mu/\tau)] - [2\tau p^{D}/(1-\tau) + 3\tau p^{M}((1-\mu)/(1-\tau))]}$$

$$+\pi\frac{[(\sigma\bar{\theta}^{2,L}+\bar{\theta}^{3,L})(p^{L}/\tau)+(\sigma\bar{\theta}^{2,M}+\bar{\theta}^{3,M})p^{M}(\mu/\tau)] - [(\sigma\bar{\theta}^{2,D}+\bar{\theta}^{3,D})p^{D}/(1-\tau) +(\sigma\bar{\theta}^{2,M}+\bar{\theta}^{3,M})p^{M}((1-\mu)/(1-\tau))]}{[1+5p^{R} +(3+2\tau)p^{L}/\tau +(2+3\tau)p^{M}(\mu/\tau)] - [2\tau p^{D}/(1-\tau) + 3\tau p^{M}((1-\mu)/(1-\tau))]}$$

This is equal to ACR$_{stud}$(3).

VI. Bounds for ATE for Years 2 and 3

As shown above, we can estimate $ITT_{tchr}(t)$ and $ACR_{tchr}(t)$ using Sample 1 teachers, $ITT_{sch}(t)$ using Sample 2 teachers, and $ITT_{stud}(t)$ and $ACR_{stud}(t)$ using student test scores in our school data. In addition, for year 1 we call also estimate $ATE_{tchr}(1)$, which also equals $ATE_{sch}(1)$, and $ATE_{stud}(1)$ since in year 1 these are all equal to the corresponding ITT estimands. Unfortunately, we cannot estimate $ATE_{tchr}(2)$, $ATE_{tchr}(3)$, $ATE_{sch}(2)$, $ATE_{sch}(3)$, $ATE_{stud}(2)$ or $ATE_{stud}(3)$. However, under plausible assumptions it is possible to show that ITT estimands are lower bounds on several ATE estimands.

Consider first $ATE_{tchr}(2)$ and $ITT_{tchr}(2)$. Their difference is:

$$ATE_{tchr}(2) - ITT_{tchr}(2) = 2\bar{\delta} + \bar{\gamma}_{1,2} - [\bar{\delta} + p^R(\delta^R+\gamma^R_{1,2}) + p^L\gamma^L_{1,2} + p^M\tau\gamma^M_{1,2}] \qquad \text{(B.103)}$$

$$= \delta^R p^R + \delta^L p^L + \delta^D p^D + \delta^M p^M + \gamma^R_{1,2}p^R + \gamma^L_{1,2}p^L + \gamma^D_{1,2}p^D + \gamma^M_{1,2}p^M - [p^R(\delta^R+\gamma^R_{1,2}) + p^L\gamma^L_{1,2} + p^M\tau\gamma^M_{1,2}]$$

$$= \delta^L p^L + \delta^D p^D + \delta^M p^M + \gamma^D_{1,2}p^D + (1-\tau)\gamma^M_{1,2}p^M$$

$$= \delta^L p^L + (\delta^D+\gamma^D_{1,2})p^D + (\delta^M+\gamma^M_{1,2}(1-\tau))p^M$$

As long as the first year of the program does not have a negative effect on the skills of likers (i.e. $\delta^L \geq 0$) and the second year does not have a negative effect on the skills of dislikers ($\delta^D+\gamma^D_{1,2} \geq 0$) or movers ($\delta^M+\gamma^M_{1,2} \geq 0$), $ITT_{tchr}(2)$ will be a lower bound for $ATE_{tchr}(2)$.

It is less clear that $ITT_{sch}(2)$ is a lower bound for $ATE_{sch}(2)$, because for these treatment effects follow schools over time, as opposed to following teachers over time, and the composition of teachers in APM and non-APM schools can change over time. More specifically:

$$ATE_{sch}(2) - ITT_{sch}(2) \qquad \text{(B.104)}$$

$$= (2\delta^R + \gamma^R_{1,2})p^R + (\delta^L(1+\tau) + \gamma^L_{1,2}\tau)(p^L/\tau) + (\delta^M(1+\tau) + \gamma^M_{1,2}\tau)p^M(\mu/\tau) + \bar{\theta}^{2,L}((1-\tau)/\tau) - \bar{\theta}^{2,D}p^D + \bar{\theta}^{2,M}p^M((\mu/\tau) - 1)$$

$$- [(2\delta^R + \gamma^R_{1,2})p^R + (\delta^L(1+\tau) + \gamma^L_{1,2}\tau)(p^L/\tau) + (\delta^M(1+\tau) + \gamma^M_{1,2}\tau)p^M(\mu/\tau) - (\delta^D p^D(\tau/1-\tau)) + \delta^M\tau p^M((1-\mu)/(1-\tau)))]$$

$$- [\bar{\theta}^{2,L}(p^L/\tau) + \bar{\theta}^{2,M}p^M(\mu/\tau) - (\bar{\theta}^{2,D}(p^D/(1-\tau)) + \bar{\theta}^{2,M}p^M((1-\mu)/(1-\tau)))]$$

$$= \bar{\theta}^{2,L}((1-\tau)/\tau) - \bar{\theta}^{2,D}p^D + \bar{\theta}^{2,M}p^M((\mu/\tau) - 1) + \delta^D p^D(\tau/1-\tau)) + \delta^M\tau p^M((1-\mu)/(1-\tau))$$

$$- [\bar{\theta}^{2,L}(p^L/\tau) + \bar{\theta}^{2,M}p^M(\mu/\tau) - (\bar{\theta}^{2,D}(p^D/(1-\tau)) + \bar{\theta}^{2,M}p^M((1-\mu)/(1-\tau)))]$$

$$= \delta^D p^D(\tau/1-\tau)) + \delta^M\tau p^M((1-\mu)/(1-\tau)) - \bar{\theta}^{2,L}p^L + \bar{\theta}^{2,D}p^D(\tau/(1-\tau)) + \bar{\theta}^{2,M}p^M((\tau-\mu)/(1-\tau)).$$

The two δ terms are ≥ 0 (assuming δ^D and δ^M are ≥ 0), but the sign of the combined effect of the θ terms, which reflect changes in teacher composition, is ambiguous, even though it is reasonable to assume that all of the $\bar{\theta}^2$ terms are > 0. One could argue that this combined effect is not far from zero and, if negative, is smaller in absolute value than the sum of the two δ terms, and so $ITT_{sch}(2)$ is a lower bound for $ATE_{sch(2)}$, but it is possible that the sum of the composition terms is negative and larger in absolute value than the two δ terms. Note, however, that if there are no likers or dislikers then there is no composition effect (since $p^L = p^D = 0$ and $\mu = \tau$) and so $ITT_{sch}(2)$ is a lower bound for $ATE_{sch}(2)$. In particular, $ATE_{sch}(2) - ITT_{sch}(2) = \delta^M\tau p^M$, which is ≥ 0 as long as $\delta^M \geq 0$, which is plausible.

Next, consider $ATE_{stud}(2)$ and $ITT_{stud}(2)$, and more specifically their difference:

$$ATE_{stud}(2) - ITT_{stud}(2) \qquad \text{(B.105)}$$

$$= \sigma\pi\bar{\delta} + \pi[(2\delta^R + \gamma^R_{1,2})p^R + (\delta^L(1+\tau) + \gamma^L_{1,2}\tau)p^L/\tau + ((1+\tau)\delta^M + \tau\gamma^M_{1,2})(\mu/\tau)p^M]$$

$$+ \pi[\bar{\theta}^{2,L}p^L((1-\tau)/\tau) - \bar{\theta}^{2,D}p^D + \bar{\theta}^{2,M}p^M((\mu/\tau) - 1)]$$

$$– [\sigma\pi\bar{\delta} + \pi((2\delta^R + \gamma^R_{1,2})p^R + (\delta^L(1+\tau) + \tau\gamma^L_{1,2})(p^L/\tau) – \delta^D\tau(p^D/(1-\tau)) + ((\delta^M(1+\tau) + \gamma^M_{1,2}\tau)(\mu/\tau) – \delta^M\tau(1-\mu)/(1-\tau))p^M)]$$

$$– \pi[\bar{\theta}^{2,L}(p^L/\tau) – \bar{\theta}^{2,D}(p^D/(1-\tau)) + \bar{\theta}^{2,M}p^M((\mu/\tau) – (1-\mu)/(1-\tau))]]$$

$$= \pi[\delta^D\tau(p^D/(1-\tau)) + \delta^M\tau(1-\mu)/(1-\tau))p^M – \bar{\theta}^{2,L}p^L + \bar{\theta}^{2,D}p^D(\tau/(1-\tau)) + \bar{\theta}^{2,M}p^M((\tau-\mu)/(1-\tau))]$$

This is simply π multiplied by $ATE_{sch}(2) – ITT_{sch}(2)$, and so, as with $ATE_{sch}(2) – ITT_{sch}(2)$, the sign of this expression for the general case is ambiguous. yet again if there are no likers or dislikers then $ITT_{stud}(2)$ is a lower bound for $ATE_{stud}(2)$ as long as $\delta^M \geq 0$.

Next, consider $ATE_{tchr}(3)$ and $ITT_{tchr}(3)$. Their difference is:

$$ATE_{tchr}(3) – ITT_{tchr}(3) \qquad (B.106)$$

$$= 3\bar{\delta} + 3\bar{\gamma}_{1,2} + \bar{\gamma}_{1,2,3} – [\bar{\delta} + p^R(2\delta^R+3\gamma^R_{1,2}+\gamma^R_{1,2,3}) + p^L(2\gamma^L_{1,2}+\gamma^L_{1,2,3}) + p^M(2\tau\gamma^M_{1,2}+\tau^2\gamma^M_{1,2,3})]$$

$$= p^L(2\delta^L+ \gamma^L_{1,2}) + p^D(2\delta^D + 3\gamma^D_{1,2}+\gamma^D_{1,2,3}) + p^M(2\delta^M + (3-2\tau)\gamma^M_{1,2}+(1-\tau^2)\gamma^M_{1,2,3})$$

This is plausibly ≥ 0. The term $p^L(2\delta^L+ \gamma^L_{1,2})$ is ≥ 0 as long as two years of exposure to the program does not reduce the skills of liker teachers. The term $p^D(2\delta^D + 3\gamma^D_{1,2}+\gamma^D_{1,2,3})$ is ≥ 0 as long as three years of exposure to the program does not reduce the skills of disliker teachers relative to the skills they would obtain from one year of exposure to the program. Finally, $p^M(2\delta^M + (3-2\tau)\gamma^M_{1,2}+(1-\tau^2)\gamma^M_{1,2,3}) \geq 0$ as long as three years of exposure to the program does not reduce the skills of mover teachers relative to the skills they would obtain relative to one year of exposure to the program.

In contrast, it is less clear that $ITT_{sch}(3)$ can serve as a lower bound for $ATE_{sch}(3)$. Their difference is:

$$ATE_{sch}(3) – ITT_{sch}(3) \qquad (B.107)$$

$$= (3\delta^R + 3\gamma^R_{1,2} + \gamma^R_{1,2,3})p^R + (\delta^L(2+\tau) + (1+2\tau)\gamma^L_{1,2} + \tau\gamma^L_{1,2,3})p^L/\tau + [(1+2\tau)\delta^M + \tau(2+\tau)\gamma^M_{1,2} + \gamma^M_{1,2,3}\tau^2](\mu/\tau)p^M$$

$$+ \bar{\theta}^{3,L}p^L((1\text{-}\tau)/\tau) - \bar{\theta}^{3,D}p^D + \bar{\theta}^{3,M}p^M((\mu/\tau) - 1)$$

$$- [(3\delta^R + 3\gamma^R_{1,2} + \gamma^R_{1,2,3})p^R + (\delta^L(2+\tau) + \gamma^L_{1,2}(1+2\tau) + \gamma^L_{1,2,3}\tau)(p^L/\tau) - \delta^D\tau(p^D/(1\text{-}\tau))]$$

$$- [(\mu/\tau)(\delta^M(1+2\tau) + \gamma^M_{1,2}\tau(2+\tau) + \gamma^M_{1,2,3}\tau^2) - ((1\text{-}\mu)/(1\text{-}\tau)(\delta^M 2\tau + \gamma^M_{1,2}\tau^2)]p^M$$

$$- [\bar{\theta}^{3,L}(p^L/\tau) - \bar{\theta}^{3,D}(p^D/(1\text{-}\tau)) + \bar{\theta}^{3,M}p^M((\mu/\tau) - (1\text{-}\mu)/(1\text{-}\tau))]$$

$$= \delta^D\tau(p^D/(1\text{-}\tau)) + ((1\text{-}\mu)/(1\text{-}\tau)(\delta^M 2\tau + \gamma^M_{1,2}\tau^2)p^M$$

$$- \bar{\theta}^{3,L}p^L + \bar{\theta}^{3,D}p^D(\tau/(1\text{-}\tau)) + \bar{\theta}^{3,M}p^M((\tau\text{-}\mu)/(1\text{-}\tau))$$

The two δ terms are ≥ 0 (assuming δ^D and δ^M are ≥ 0) as long as two years of exposure to the program does not reduce the skills of movers (as long as $2\delta^M + \gamma^M_{1,2} \geq 0$), but the sign of the combined effect of the θ terms, which again reflects changes in teacher composition, is ambiguous, even though it is reasonable to assume that all of the $\bar{\theta}^2$ terms are > 0. Note, however, that if there are no likers or dislikers then there is no composition effect (since $p^L = p^D = 0$ and $\mu = \tau$) and so $ITT_{sch}(3)$ is a lower bound for $ATE_{sch}(3)$.

Finally, turn to student skills for year three and compare $ITT_{stud}(3)$ with $ATE_{stud}(3)$:

$$ATE_{stud}(3) - ITT_{stud}(3) \qquad (B.108)$$

$$= \sigma^2\pi\bar{\delta} + \sigma\pi[(2\delta^R + \gamma^R_{1,2})p^R + (\delta^L(1+\tau) + \gamma^L_{1,2}\tau)(p^L/\tau) + ((1+\tau)\delta^M + \tau\gamma^M_{1,2})(\mu/\tau)p^M]$$

$$+ \pi[(3\delta^R + 3\gamma^R_{1,2} + \gamma^R_{1,2,3})p^R + (\delta^L(2+\tau) + (1+2\tau)\gamma^L_{1,2} + \tau\gamma^L_{1,2,3})(p^L/\tau) + ((1+2\tau)\delta^M + \tau(2+\tau)\gamma^M_{1,2} + \gamma^M_{1,2,3}\tau^2)(\mu/\tau)p^M]$$

$$+ \sigma\pi[\bar{\theta}^{2,L}((1\text{-}\tau)/\tau) - \bar{\theta}^{2,D}p^D + \bar{\theta}^{2,M}p^M((\mu/\tau) - 1)] + \pi[\bar{\theta}^{3,L}p^L((1\text{-}\tau)/\tau) - \bar{\theta}^{3,D}p^D + \bar{\theta}^{3,M}p^M((\mu/\tau) - 1)]$$

$$- \sigma^2\pi\bar{\delta} + \sigma\pi[(2\delta^R + \gamma^R_{1,2})p^R + (\delta^L(1+\tau) + \tau\gamma^L_{1,2})(p^L/\tau) - \delta^D\tau(p^D/(1\text{-}\tau)) + (\delta^M((1+\tau) + \gamma^M_{1,2}\tau)(\mu/\tau) - \delta^M\tau(1\text{-}\mu)/(1\text{-}\tau))p^M]$$

$$- \pi[(3\delta^R + 3\gamma^R_{1,2} + \gamma^R_{1,2,3})p^R + (\delta^L(2+\tau) + \gamma^L_{1,2}(1+2\tau) + \gamma^L_{1,2,3}\tau)(p^L/\tau) - \delta^D\tau(p^D/(1\text{-}\tau))]$$

$$- \pi[(\mu/\tau)(\delta^M(1+2\tau) + \gamma^M_{1,2}\tau(2+\tau) + \gamma^M_{1,2,3}\tau^2) - ((1\text{-}\mu)/(1\text{-}\tau)(\delta^M 2\tau + \gamma^M_{1,2}\tau^2)]p^M$$

$$- \sigma\pi[\bar{\theta}^{2,L}(p^L/\tau) - \bar{\theta}^{2,D}(p^D/(1\text{-}\tau)) + \bar{\theta}^{2,M}p^M((\mu/\tau) - (1\text{-}\mu)/(1\text{-}\tau))]$$

$$- \pi[\bar{\theta}^{3,L}(p^L/\tau) - \bar{\theta}^{3,D}(p^D/(1\text{-}\tau)) + \bar{\theta}^{3,M}p^M((\mu/\tau) - (1\text{-}\mu)/(1\text{-}\tau))]$$

$$= \sigma\pi[\delta^D p^D(\tau/1\text{-}\tau)) + \delta^M\tau p^M((1\text{-}\mu)/(1\text{-}\tau))] + \pi[\delta^D p^D(\tau/1\text{-}\tau)) + (\delta^M 2\tau + \tau^2\gamma^M_{1,2})p^M((1\text{-}\mu)/(1\text{-}\tau))]$$

$$+ \sigma\pi[\text{-}\bar{\theta}^{2,L}p^L + \bar{\theta}^{2,D}p^D(\tau/(1\text{-}\tau)) + \bar{\theta}^{2,M}p^M((\tau\text{-}\mu)/(1\text{-}\tau))] + \pi[\text{-}\bar{\theta}^{3,L}p^L + \bar{\theta}^{3,D}p^D(\tau/(1\text{-}\tau)) + \bar{\theta}^{3,M}p^M((\tau\text{-}\mu)/(1\text{-}\tau))]$$

The first three δ terms are ≥ 0 (assuming δ^D and δ^M are ≥ 0), and the $\delta^M 2\tau + \tau^2\gamma^M_{1,2}$ term is also ≥ 0 as long as two years of exposure to the program does not reduce the skills of movers (as long as $2\delta^M + \gamma^M_{1,2} \geq 0$). Yet the sign of the combined effect of the θ terms, which again reflects changes in teacher composition, is ambiguous, even though it is reasonable to assume that all of the $\bar{\theta}^2$ terms are > 0. Note, however, that if there are no likers or dislikers then there is no composition effect (since $p^L = p^D = 0$ and $\mu = \tau$) and so $ITT_{stud}(3)$ is a lower bound for $ATE_{stud}(3)$.

Appendix C: Teacher Allocation during the Second Year of Treatment

Under the framework developed in Section 3 and in Appendix B, schools are randomly assigned to treatment and control arms during the first year of treatment. Teachers cannot change schools during that year and, therefore, are also randomly distributed between APM and non-APM schools. Therefore, we expect the proportion of each type of teacher to be equally distributed between treatment arms.

Table C1: Teacher Distribution during Year One

School Treatment Arm	Proportion of each type of teachers				
	Likers	Movers	Remainers	Dislikers	Total
Treatment	p^L	p^M	p^R	p^D	1
Control	p^L	p^M	p^R	p^D	1

As the first year end, some teachers change schools according to their preferences. Remainers stay in their school regardless of the treatment status. Likers that started in control schools will all move to APM schools. Likers that started in APM schools will remain in that type of school, although only a fraction of them will stay in the same school (we defined that proportion as σ), and the rest moving to a different treated school. Dislikers that started in treated schools will all move to control schools, while dislikers that started in control schools will remain in thar treatment arm, with a fraction remaining in their original school (defined as ν) and the rest moving to a different control school. All movers will change schools independently of their original placement. Irrespectively of where they started, a fraction will move to treated schools (defined as μ) and the rest will go to control schools.

Table C1: Teacher Relocation between Years One and Two

	Movement decision	Likers	Movers	Remainers	Dislikers	Row Sum
Assigned to tratment	Moves to treated	$p^L(1-\sigma)$	$p^M\mu$	0	0	$p^L(1-\sigma)+ p^M\mu$
	Moves to control	0	$p^M(1-\mu)$	0	p^D	$p^M(1-\mu)+ p^D$
	Stays in same school	$p^L\sigma$	0	p^R	0	$p^L\sigma+ p^R$
Assigned to control	Moves to treated	p^L	$p^M\mu$	0	0	$p^L + p^M\mu$
	Moves to control	0	$p^M(1-\mu)$	0	$p^D(1-\nu)$	$p^M(1-\mu)+ p^D(1-\nu)$
	Stays in same school	0	0	p^R	$p^D\nu$	$p^R + p^D\nu$