Measuring the Power of Learning.®

Research Memorandum

ETS RM–19-15

A Note on the Causal Definition of Teacher Effects

Daniel F. McCaffrey

J. R. Lockwood

December 2019

ETS Research Memorandum Series

A Note on the Causal Definition of Teacher Effects

Daniel F. McCaffrey and J. R. Lockwood
Educational Testing Service, Princeton, New Jersey

December 2019

Corresponding author: D. F. McCaffrey, E-mail: dmccaffrey@ets.org

Suggested citation: McCaffrey, D. F., & Lockwood, J. R. (2019). *A note on the causal definition of teacher effects* (Research Memorandum No. RM-19-15). Princeton, NJ: Educational Testing Service.

Action Editor: Elizabeth Stone

Reviewers: Heather Buzick and Xiang Liu

Abstract

Value-added models are now a mainstay of research on teachers and are used by some states and school systems for teacher evaluation or accountability. Linear analysis of covariance (ANCOVA) models are widely used in the estimation of teacher value added. In this research memorandum, we use the potential outcomes framework to define teacher value-added effects as causal effects of the contributions of teachers and classroom peers, and we establish conditions for ANCOVA models to yield consistent estimates of those effects.

Key words: Value-added models, potential outcomes, peer effects, analysis of covariance

Acknowledgments

The research reported here was supported in part by the Institute of Education Sciences, U.S. Department of Education, through grant R305D140032 to Educational Testing Service (ETS). The opinions expressed are those of the authors and do not represent views of the Institute or the U.S. Department of Education. We thank Heather Buzick, Xiang Liu, and Elizabeth A. Stone for constructive comments on a previous draft.

Value-added (VA) models are now a mainstay of research on teachers and are used by some states and school systems for teacher evaluation or accountability. Linear analysis of covariance (ANCOVA) models are widely used in the estimation of teacher value added, with many examples in the literature (e.g., Clotfelter, Ladd, & Vigdor, 2006; Goldhaber & Hansen, 2010; Goldhaber, Quince, & Theobald, 2018; Harris & Sass, 2011; Isenberg et al., 2013; Sass, Hannaway, Xu, Figlio, & Feng, 2012). The parameters of these models corresponding to *teacher effects* are typically treated as the causal effect the teacher has on student achievement (i.e., the difference in students' achievement when taught by this teacher relative to what it would have been if they had been taught by a different teacher). However, as Raudenbush (2004) noted, these effects are just adjusted mean test scores for students in each class, and those adjusted means may be susceptible to many factors beyond the teacher. For example, the students also contribute to the learning environment in a classroom. Multiple authors (cf. Raudenbush, 2004; Raudenbush & Jean, 2012; Rothstein, 2009, 2010) have acknowledged that the estimated effects from ANCOVA models are better understood as *teacher grouping effect*s—the effects of being in a teacher's class, but not necessarily causal effects that can be attributed to the teacher. Referring to previous work for school effects (Raudenbush & Willms, 1995), Raudenbush (2004) called such grouping effects *Type A* effects. In the context of teacher groupings, Type A effects are the effects of being assigned to a classroom for a fixed student and include the combined effects of the teacher, the peers, and other factors unique to that classroom experience for the individual student. Raudenbush (2004) and Raudenbush and Willms (1995) contrasted these effects with *Type B* effects, which are the effects of each teacher, distinct from any other sources.

While there is a heuristic sense that the teacher effects in linear ANCOVA value-added models are more closely aligned with Type A effects, the effects being estimated from these models have not been rigorously defined, and their interpretation remains debated. Potential outcomes (Holland, 1986; Rubin, 1974) are commonly used to formally define causal effects. In this research memorandum, we define *teacher effects* in terms of students' potential outcomes and establish the link between these effects and the parameters of linear ANCOVA models. We demonstrate that when the causal model allows for interference such that a student's potential outcome depends on the assignments of all students, ANCOVA models do indeed yield Type A rather than Type B effects. Whether Type A effects are appropriate for the uses of VA model estimates is beyond the scope of this research memorandum, but we do provide additional clarity on how to interpret the estimated effects.

The Analysis of Covariance Model for Teacher Effects

We assume that we are studying a group of N students in a particular school during a given target grade and year (e.g., grade 5 in the 2011–2012 school year) and an associated group of K teachers from this school who teach a selected subject (e.g., mathematics) to these students. We consider the problem of estimating the effects of being in a given teacher's class on student achievement in the selected subject. Typically in VA modeling, data include students and teachers from multiple schools. In that case, our model would be replicated for each school. Applications of VA models often include comparing teachers across schools; however, there is some debate about the appropriateness of such comparisons (Raudenbush, 2004). We follow Rothstein (2009) by focusing on estimating teacher VA within schools. Thus the effect associated with each teacher is defined relative to the other teachers in the same school, who also teach the selected subject to the students in the target grade and year. It is also common in VA modeling to ignore distinctions among multiple classes of students taught by the same teacher, so in this research memorandum, we focus on the situation in which each teacher teaches only one group of students. We discuss the implications of this assumption at the end of the memorandum. In the following development, we do not use subscripts for subject, grade, year, or school to simplify the notation.

We assume that in the given school, grade, and year, each of the N students is taught the selected subject by exactly one of the K teachers. Let n_t be the number of students taught by teacher t for $t = 1, \ldots, K$. The $\{n_t\}_{t=1}^{K}$ satisfy $\sum_{t=1}^{K} n_t = N$. The sample sizes N, K, and $\{n_t\}_{t=1}^{K}$ are assumed fixed, so that inferences are conditional on these values. Given that students and teachers can transfer throughout the year, the sample sizes are typically those that exist in the testing data used for analysis. Let Y_i, $i = 1, \ldots, N$, equal the observed achievement test score for student i in a given subject for the given school, grade, and target year. For each student, we also define $\mathbf{X}_i$ equal to a vector of achievement variables from prior school years and $\mathbf{Z}_i$ equal to a vector of other background characteristics, such as demographics or family economic status. We assume that $\mathbf{X}_i$ and $\mathbf{Z}_i$ are centered such that $N^{-1}\sum_{i=1}^{N}\mathbf{X}_i = \mathbf{0}$ and $N^{-1}\sum_{i=1}^{N}\mathbf{Z}_i = \mathbf{0}$. We further assume that for each student, there is a vector of teacher indicator variables, $\mathbf{T}_i$, linking the student to the student's teacher. The vector $\mathbf{T}_i$ has K elements, one for each teacher in the school, grade, and year who teaches the target subject, with the element corresponding to student i's teacher equal to 1, and all others equal to 0.

The ANCOVA model is a commonly used model for estimating teacher effects from longitudinal student achievement data as described in the previous paragraph. The ANCOVA model for the observed target-year scores is given by

$$Y_i = \boldsymbol{\beta}_0'\mathbf{X}_i + \boldsymbol{\beta}_1'\mathbf{Z}_i + \boldsymbol{\omega}'\mathbf{T}_i + \epsilon_i, \quad (1)$$

where $\boldsymbol{\beta}_0$ and $\boldsymbol{\beta}_1$ are vectors of regression coefficients for the achievement scores and other covariates and $\boldsymbol{\omega} = (\omega_1, \ldots, \omega_K)'$ is a vector of means corresponding to teachers, with one element for each teacher in the school, grade, and year. Because the vector $\mathbf{T}_i$ is also of length K, taking on the value of 1 for element of $\boldsymbol{\omega}$ corresponding to the student's teacher, and zero otherwise, the student's test score is a function of the mean associated with the student's teacher. It is assumed that ϵ_i has mean zero and finite variance.[1] Teacher effects equal contrasts of the teacher means $\boldsymbol{\omega}$. For example, a commonly used set of effects are $\theta_t = \omega_t - K^{-1}\sum_{k=1}^{K}\omega_k$ for $t = 1, \ldots, K$, the difference between the mean for students assigned the teacher t and the average mean across teachers.[2]

Given observed data, an estimation algorithm such as ordinary least squares (OLS) can be used to estimate the parameters of the ANCOVA model in Equation 1. This results in estimates $\widehat{\boldsymbol{\omega}}$ of the teacher parameters $\boldsymbol{\omega}$ or corresponding estimates of teacher effects $\widehat{\theta}_t = \widehat{\omega}_t - K^{-1}\sum_{k=1}^{K}\widehat{\omega}_k$. Thus $\boldsymbol{\omega}$ can be defined as the expected value of $\widehat{\boldsymbol{\omega}}$ over the distribution of possible values of the observed data, and similarly for the vector of effects $\boldsymbol{\theta}$. The goal of this research memorandum is to provide an interpretation of $\boldsymbol{\omega}$ and $\boldsymbol{\theta}$ in terms of causal effects.

The Potential Outcomes Model for Teacher Effects

For $i = 1, \ldots, N$, let $\Upsilon_i = t$ if student i is assigned to teacher $t = 1, \ldots, K$. Let $\boldsymbol{\Upsilon} = (\Upsilon_1, \ldots, \Upsilon_N)'$, the vector of teacher assignments for all students. There are A possible assignment vectors, $\boldsymbol{\Upsilon}$, where

$$A = \binom{N}{n_1}\binom{N - n_1}{n_2}\cdots\binom{N - \sum_{k=1}^{K-2} n_k}{n_{K-1}}.$$

Let $\mathcal{T} = \{\boldsymbol{\Upsilon}_1, \ldots, \boldsymbol{\Upsilon}_A\}$. Let $\boldsymbol{\Upsilon}_{obs} \in \mathcal{T}$ denote the observed vector of students' teacher assignments. If students were randomly assigned to teachers, then all elements of $\mathcal{T}$ would have an equal probability of being observed; however, schools might not assign students at random. For example, Clotfelter et al. (2006) found that within some schools, teachers with stronger credentials tend to teach students from more affluent families than teachers with weaker credentials. Because assignments might not be equally likely to be observed, let $\pi_a = Pr(\boldsymbol{\Upsilon}_{\text{obs}} = \boldsymbol{\Upsilon}_a)$ for $a = 1, \ldots, A$. For each $a = 1, \ldots, A$ and $i = 1, \ldots, N$, denote the teacher assignment for student i under the assignment vector $\boldsymbol{\Upsilon}_a$ by Υ_{ai}. For each assignment vector a and teacher t, let $\mathcal{T}_{ta} = \{i \in 1, \ldots, N : \Upsilon_{ai} = t\}$, the set of all students assigned to teacher t under assignment a. As with the ANCOVA model, n_t equals the number of students assigned to teacher t, so that $\mathcal{T}_{ta}$ contains n_t student for all values of a. For each student i and teacher t, let $\mathcal{T}_{it} = \{\boldsymbol{\Upsilon} \in \mathcal{T} : \Upsilon_{ai} = t\}$, the set of all assignments in which student i is assigned to teacher t. For each student i and teacher t, $\mathcal{T}_{it}$ contains $m_t = An_t/N$ assignments. That is, for each student, there are exactly m_t possible assignments in which the student is assigned to teacher t.

Each student i has A potential outcomes $Y_{ia} = Y_i(\boldsymbol{\Upsilon}_a)$ for $\boldsymbol{\Upsilon}_a \in \mathcal{T}$. A student's potential outcome for a given assignment depends on the teacher to whom the student is assigned and the assignments of all the other students. In particular, each potential outcome Y_{ia} depends on the other students assigned to the same teacher, that is, all students in $\mathcal{T}_{ta}$. Thus the model does not make the stable unit treatment value assumption (SUTVA; Rubin, 1974). Instead, it allows for interference (e.g., Hudgens & Halloran, 2008; Tchetgen Tchetgen & VanderWeele, 2012) because students in the same class could influence their peers. We are unaware of other applications in education that develop a causal model for teacher effects using potential outcomes with interference. Although, in their evaluation of kindergarten retention, Hong and Raudenbush (2005, 2006, 2013) used this construction to define students' potential outcomes in the presence of interference, they did not consider teacher effects.

Let $\mu_{ta} = n_t^{-1} \sum_{i \in \mathcal{T}_{ta}} Y_{ia}$, the mean of the potential outcomes for the students assigned to teacher t under assignment a. We define

$$\mu_t = A^{-1} \sum_{a=1}^{A} \mu_{ta}. \tag{2}$$

It is the average of the class means across all the potential classes of size n_t that could be assigned to the teacher. If, for each student, we let $\bar{Y}_{it} = m_t^{-1} \sum_{a \in \mathcal{T}_{it}} Y_{ia}$, the average of the outcome for student i when the student is assigned to teacher t, then $\mu_t = N^{-1} \sum_{i=1}^{N} \bar{Y}_{it}$. It is the population average over students of the students' average outcomes when students are assigned to teacher t. To see this, let $I(u)$ denote the indicator function equal to 1 if the argument u is true, and zero otherwise; then,

$$\begin{aligned}
\mu_t &= A^{-1} \sum_{a=1}^{A} \mu_{ta} = A^{-1} \sum_{a=1}^{A} n_t^{-1} \sum_{i \in \mathcal{T}_{ta}} Y_{ia} \\
&= A^{-1} \sum_{a=1}^{A} n_t^{-1} \sum_{i=1}^{N} Y_{ia} I(\Upsilon_{ia} = t) \\
&= A^{-1} n_t^{-1} \sum_{i=1}^{N} \sum_{a=1}^{A} Y_{ia} I(\Upsilon_{ia} = t) \\
&= A^{-1} n_t^{-1} \sum_{i=1}^{N} \sum_{a \in \mathcal{T}_{it}} Y_{ia} = A^{-1} n_t^{-1} \sum_{i=1}^{N} m_t \bar{Y}_{it} = N^{-1} \sum_{i=1}^{N} \bar{Y}_{it}.
\end{aligned}$$

The last equality holds because $m_t = An_t/N$.

As McCaffrey et al. (2013) and Frölich (2004) have noted, with multiple treatments, causal effects are defined by linear contrasts of the means of the potential values. In particular, we consider the causal effect defined as the difference between the expected outcome for each teacher and the average of these means for all teachers in the school who teach the target-grade students and subject; that is,

$$\phi_t = \mu_t - K^{-1} \sum_{k=1}^{K} \mu_k. \tag{3}$$

If SUTVA holds, then each student has only K potential values $Y_i(1)$ to $Y_i(K)$, one per teacher. In this case, $Y_{ia} = Y_i(\Upsilon_{ai})$, meaning that the potential outcome of student i under assignment a depends only on the teacher t to which the student is assigned. It does not depend on the assignments of other students and specifically does not depend on which other students are also assigned to teacher t under assignment a. Consequently, $\bar{Y}_{it} = Y_i(t)$ and $\mu_t = N^{-1} \sum_{i=1}^{N} Y_i(t)$, the marginal population mean of the potential outcomes for this teacher. When SUTVA holds, the marginal mean for a teacher of the potential outcomes across students when assigned to that teacher equals the marginal mean across assignments. Causal effects, which are contrasts of the marginal means across students, are also contrasts of the marginal means across assignments. Thus the definitions of μ_t and ϕ_t are natural extensions of the marginal mean and causal effect from the canonical case where SUTVA holds to our case where there is interference.

Potential Outcomes and the Cumulative Achievement Model

The previous section provided a model for potential outcomes and causal effects of teachers. However, the potential outcomes and effects are not connected to the linear ANCOVA model. In this section, we develop a model for the potential outcomes in terms of covariates and then use this model to evaluate the effects in the ANCOVA model in the next section. In the VA literature, the ANCOVA model is often motivated by cumulative achievement model (Boardman & Murnane, 1979; Sass, Semykina, & Harris, 2014; Todd

& Wolpin, 2003). In this model, a student's achievement in a given year is a function of student (e.g., effort), family (e.g., help with homework or food), and educational inputs from both the current year and all previous years of schooling, along with a time-invariant endowment characteristic of the student (e.g., innate ability) and a random error. Under this framework, the goal of VA modeling is to estimate the effects of the current-year educational inputs distinct from historical inputs. We follow this approach and model each potential outcome for the student as a function of vectors of current and past student ($\mathbf{P}_i$), family ($\mathbf{F}_i$), and educational ($\mathbf{E}_i$) inputs; the time-invariant endowment ζ_i; random error ϑ_{ia}; and additive measurement error ν_{ia}. Measurement error is included because the input model is for achievement, but the potential outcomes are for test scores, which have measurement error. This yields

$$Y_{ia} = G_a(\mathbf{P}_i, \mathbf{F}_i, \mathbf{E}_{ia}, \zeta_i, \vartheta_{ia}) + \nu_{ia}. \tag{4}$$

We assume the first element of each of the vectors $\mathbf{P}_i$, $\mathbf{F}_i$, and $\mathbf{E}_{ia}$ is for the current-year inputs and that the remaining elements are the historical inputs. The current-year educational inputs include school and classroom inputs. Classroom inputs include contributions from teachers and peers, so the educational input vector depends on the assignment. Peer inputs would be a source of the interference among students. The general form of the cumulative achievement function given by Equation 4 depends on the assignment a and the full history of student, family, and educational inputs. Models in such generality are difficult to use for estimation, so the VA literature often makes additional assumptions. (Sass et al., 2014, provided a thorough review of the common assumptions made for the cumulative achievement model.) We follow this tradition. First, we assume that the achievement function is the same for all assignments so that the only differences across the potential outcomes are educational inputs and the random errors and measurement errors:

$$Y_{ia} = G(\mathbf{P}_i, \mathbf{F}_i, \mathbf{E}_{ia}, \zeta_i, \vartheta_{ia}) + \nu_{ia}. \tag{5}$$

Next, following most authors, we assume additive separability, time-invariant family inputs, and equal effects for individual time-invariant student endowment and the time-invariant family inputs. These assumptions yield

$$Y_{ia} = \boldsymbol{\alpha}'_P \mathbf{P}_i + \boldsymbol{\alpha}'_E \mathbf{E}_{ia} + \psi\chi_i + \vartheta_{ia} + \nu_{ia}, \tag{6}$$

where χ_i is the combined time-invariant family inputs and student endowment. Finally, the VA literature also commonly relies on the assumption that the coefficients on each of the inputs decay geometrically with the gap in years between the current and the historic input and that the model is invariant across school years. These assumptions yield

$$Y_{ia} = \alpha_{P1} P_{i1} + \alpha_{E1} E_{ia1} + \beta X_{i1} + \delta\chi_i + \eta_{ia} + \nu_{ia}, \tag{7}$$

where α_{P1}, P_{1i}, α_{E1}, and E_{ia1} equal the first elements of the corresponding coefficient or input vectors and represent the current-year inputs. The variable X_{i1} equals the student's prior-year (lag-1) achievement in the target subject, δ depends on differences between the current-year and prior-year effects from the family and endowment contributions, and η_{ia} depends on the difference between the random error for the current year and assignment and the random error for the prior year. The lag-1 achievement is measured prior to the current school year, so it cannot depend on the assignment and is constant across all assignments. Similarly the student's inputs for the current year are also assumed equal across the potential vectors of student assignments. These assumptions are all strong and may not hold for some applications. It is beyond the scope of this research memorandum to evaluate the veracity of the assumptions. The goal of this memorandum is to determine, given that the assumptions hold and the resulting model is used to estimate teacher effects, what the potential outcomes model implies about the interpretation of those estimated effects. See Sass et al. (2014) for details on this development and for tests of the various assumptions used in this model. We make the additional assumptions that current-year student inputs are functions of student background characteristics $\mathbf{Z}_i$, or are otherwise random, and that the time-invariant component $\delta\chi_i$ depends on student background characteristics and prior achievement, which results in

$$Y_{ia} = \boldsymbol{\beta}_0'\mathbf{X}_i + \boldsymbol{\beta}_1'\mathbf{Z}_i + \xi_{ia}, \tag{8}$$

where ξ_{ia} depends on the current-year educational inputs, the quantity of interest, and random errors unique to the student and the assignment.

The key feature of this model is that the quantity of interest is the effect of the current-year educational inputs distinct from the student's history and other annual inputs, which are specified by $\boldsymbol{\beta}_0'\mathbf{X}_i + \boldsymbol{\beta}_1'\mathbf{Z}_i$. Hence the teacher effects from the potential outcomes must be defined distinctly from these terms as well.

Teacher Effects and the Analysis of Covariance Model

Given the desire to estimate education inputs distinct from $\boldsymbol{\beta}_0'\mathbf{X}_i + \boldsymbol{\beta}_1'\mathbf{Z}_i$, we write μ_{ta} in terms of the covariate vectors $(\mathbf{X}, \mathbf{Z})$ and the residuals ξ:

$$\begin{aligned} \mu_{ta} &= n_t^{-1} \sum_{i \in \mathcal{T}_{ta}} Y_{ia} \\ &= n_t^{-1} \sum_{i \in \mathcal{T}_{ta}} (\boldsymbol{\beta}_0'\mathbf{X}_i + \boldsymbol{\beta}_1'\mathbf{Z}_i + \xi_{ia}) \\ &= \boldsymbol{\beta}_0'\bar{\mathbf{X}}_{ta} + \boldsymbol{\beta}_1'\bar{\mathbf{Z}}_{ta} + \bar{\xi}_{ta}, \end{aligned} \tag{9}$$

where $\bar{\mathbf{X}}_{ta} = n_t^{-1} \sum_{i \in \mathcal{T}_{ta}} \mathbf{X}_i$, $\bar{\mathbf{Z}}_{ta} = n_t^{-1} \sum_{i \in \mathcal{T}_{ta}} \mathbf{Z}_i$, and $\bar{\xi}_{ta} = n_t^{-1} \sum_{i \in \mathcal{T}_{ta}} \xi_{ia}$. This implies that $\mu_t = \boldsymbol{\beta}_0' \left(A^{-1} \sum_{a=1}^{A} \bar{\mathbf{X}}_{ta} \right) + \boldsymbol{\beta}_1' \left(A^{-1} \sum_{a=1}^{A} \bar{\mathbf{Z}}_{ta} \right) + A^{-1} \sum_{a=1}^{A} \bar{\xi}_{ta}$. Intuitively, both $\left(A^{-1} \sum_{a=1}^{A} \bar{\mathbf{X}}_{ta} \right)$ and $\left(A^{-1} \sum_{a=1}^{A} \bar{\mathbf{Z}}_{ta} \right)$ must be $\mathbf{0}$ because the mean of both $\mathbf{X}_i$ and $\mathbf{Z}_i$ over students is $\mathbf{0}$ and each student is assigned to teacher t the same number of times across the

A possible assignments. Formally,

$$\begin{aligned} A^{-1}\sum_{a=1}^{A}\bar{\mathbf{X}}_{ta} &= A^{-1}\sum_{a=1}^{A}n_t^{-1}\sum_{i=1}^{N}\mathbf{X}_i I(\Upsilon_{ia}=t) \\ &= A^{-1}n_t^{-1}\sum_{i=1}^{N}\sum_{a=1}^{A}\mathbf{X}_i I(\Upsilon_{ia}=t) \\ &= A^{-1}n_t^{-1}\sum_{i=1}^{N}\mathbf{X}_i\sum_{a=1}^{A}I(\Upsilon_{ia}=t) \\ &= A^{-1}n_t^{-1}\sum_{i=1}^{N}\mathbf{X}_i\left(\frac{n_tA}{N}\right) \\ &= N^{-1}\sum_{i=1}^{N}\mathbf{X}_i = 0. \end{aligned}$$

Similar calculations yield that $A^{-1}\sum_{a=1}^{A}\bar{\mathbf{Z}}_{ta} = 0$. Consequently,

$$\mu_t = A^{-1}\sum_{a=1}^{A}\bar{\xi}_{ta}. \tag{10}$$

The comparison of the expression for μ_t in Equation 10 to the original definition of μ_t in Equation 2 is important. The original definition of μ_t in Equation 2 involves averages of potential outcomes without any restrictions on the functional form of those potential outcomes. Alternatively, the expression for μ_t in Equation 10 makes clear that under the assumptions made about the potential outcomes that lead to Equation 8, μ_t depends only on current educational inputs, distinct from the historical inputs for the student captured by $\mathbf{X}_i$ and $\mathbf{Z}_i$.

We now return to the issue of what these developments imply for the interpretation of $\boldsymbol{\omega}$ in Equation 1. Model Equation 1 assumes that the observed scores are a linear function of the covariates, the teacher effects, and an additive mean-zero residual. To connect the potential outcomes to Equation 1, we first write the potential outcome for each student in terms of covariates, a teacher effect for the assignment, and additive residuals by defining $\epsilon_{ia} = \xi_{ia} - \bar{\xi}_{ta}$, so that $\sum_{i\in\mathcal{T}_{ta}}\epsilon_{ia} = 0$ and $\sum_{i=1}^{N}\epsilon_{ia} = 0$. Next, we set $\omega_{ta} = \bar{\xi}_{ta}$, which yields

$Y_{ia} = \boldsymbol{\beta}_0'\mathbf{X}_i + \boldsymbol{\beta}_1'\mathbf{Z}_i + \sum_{t=1}^{K} T_{it}\omega_{ta} + \epsilon_{ia}$, where $T_{it} = 1$ if student i is assigned to teacher t, and zero otherwise. The variables in Equation 1 are the observed values for the realized assignment vector $\boldsymbol{\Upsilon}_{\text{obs}}$, so that $Y_i = Y_{i,\text{obs}}$, $\boldsymbol{\omega} = (\omega_{1,\text{obs}}, \ldots, \omega_{K,\text{obs}})'$, and $\epsilon_i = \epsilon_{i,\text{obs}}$. The covariate vectors $(\mathbf{X}_i, \mathbf{Z}_i)$ were measured prior to the target year, and so for each student, they take on the same value for every value of $\boldsymbol{\Upsilon}_{\text{obs}}$. Consistent estimation of $\boldsymbol{\beta}_0$ and $\boldsymbol{\beta}_1$ requires that ϵ_{ia} be uncorrelated with the elements of $\mathbf{X}_i$ and $\mathbf{Z}_i$ for all students and every potential assignment. This requirement implies that any way in which a student deviates from the classroom mean in any assigned classroom can only depend on the covariates in the model or factors that are uncorrelated with those covariates.

Fitting the ANCOVA model in Equation 1 yields the estimate $\hat{\omega}_t = \bar{Y}_{t,\text{obs}} - \hat{\boldsymbol{\beta}}_0\bar{\mathbf{X}}_{t,\text{obs}} - \hat{\boldsymbol{\beta}}_1\bar{\mathbf{Z}}_{t,\text{obs}}$, where $\bar{Y}_{t,\text{obs}} = n_t^{-1}\sum_{i\in\mathcal{T}_{t,\text{obs}}} Y_i$. For large samples and under the necessary regularity conditions, including the orthogonality of ϵ_{ia} and the elements of $\mathbf{X}_i$ and $\mathbf{Z}_i$, $\hat{\boldsymbol{\beta}}_0$ and $\hat{\boldsymbol{\beta}}_1$ will closely approximate $\boldsymbol{\beta}_0$ and $\boldsymbol{\beta}_1$, so we consider the conditions required for $\tilde{\omega}_t = \bar{Y}_{t,\text{obs}} - \boldsymbol{\beta}_0\bar{\mathbf{X}}_{t,\text{obs}} - \boldsymbol{\beta}_1\bar{\mathbf{Z}}_{t,\text{obs}}$ to be an unbiased estimator of μ_t. By Equation 9, $\tilde{\omega}_t = \bar{\xi}_{ta}$.

Let $\mathbb{E}$ denote the expectation operator over possible assignments where assignment a is given probability π_a, for $a = 1, \ldots, A$. The estimator $\tilde{\omega}_t$ is unbiased if $\mathbb{E}\tilde{\omega}_t = \mu_t$. Because $\mathbb{E}\tilde{\omega}_t = \sum_{a=1}^{A}\tilde{\omega}_{ta}\pi_a$ and $\tilde{\omega}_{ta} = \bar{\xi}_{ta}$, $\tilde{\omega}_t$ is unbiased if $\sum_{a=1}^{A}\bar{\xi}_{ta}\pi_a = \mu_t$, or equivalently, if $\sum_{a=1}^{A}(\bar{\xi}_{ta} - \mu_t)\pi_a = 0$. This orthogonality condition holds if π_a is constant, meaning that assignments are completely random. When assignment probabilities are not constant, as is expected to be true in real applications, the requirement for unbiasedness is much more stringent.

Although $\bar{\xi}_{ta}$ excludes the inputs captured by the quantity $\boldsymbol{\beta}_0'\bar{\mathbf{X}}_{ta} + \boldsymbol{\beta}_1'\bar{\mathbf{Z}}_{ta}$, this does not mean that $\bar{\xi}_{ta}$ does not depend on $\bar{\mathbf{X}}_{ta}$ or $\bar{\mathbf{Z}}_{ta}$. For example, if high-achieving peers (e.g., peers with high average values of $\mathbf{X}$) positively influence their classmates' current-year achievement, then $\bar{\xi}_{ta}$ could depend on $\bar{\mathbf{X}}_{ta}$. In such cases, if assignment probabilities π_a are such that teacher t is more likely to be assigned students with high average values of $\mathbf{X}$ and/or $\mathbf{Z}$, then it is reasonable to expect that the condition $\sum_{a=1}^{A}(\bar{\xi}_{ta} - \mu_t)\pi_a = 0$ required for unbiasedness of the ANCOVA estimator would fail to hold. Unbiased estimation appears to require that either the peer effects or the assignment probabilities be unrelated to the covariates. These are rather strong assumptions that are contradictory to the general concerns in the VA literature that both assignments and peer effects depend on the covariates. Thus, when there is interference, estimation with the ANCOVA models appears unlikely to yield unbiased estimates of the teacher means μ_t as we have defined them. Similarly, estimates of the teacher effects, $\tilde{\theta}_t = \tilde{\omega}_t - K^{-1}\sum_{k=1}^{K}\tilde{\omega}_k$, also are unlikely to have expected value equal to the teacher effects ϕ_t in Equation 3 when there is interference.

Although ANCOVA does not yield unbiased estimators of the teacher means μ_t or teacher effects ϕ_t, it does yield unbiased estimators of alternative estimands defined by weighted means across possible assignments. Specifically, if we define the mean for each teacher as $\mu_t^* = \sum_{a=1}^{A}\bar{\xi}_{ta}\pi_a$ and teacher effects by $\phi_t^* = \mu_t^* - K^{-1}\sum_{k=1}^{K}\mu_k^*$, then by definition, ANCOVA provides unbiased estimators of μ_t^* and ϕ_t^* for $t = 1, \ldots, K$. However, the weighted means conflate the effects of the teacher and the probability of assignment, which involves the contributions of students. Hence the weighted means conflate teacher and peer effects on student achievement. Some students will be more likely to be assigned to a given teacher t than other students and, compared to μ_t, μ_t^*, will be more influenced by those students than by students who have a lower probability of being assigned to teacher t. Thus the "teacher effects" defined by ϕ_t^* and estimated unbiasedly by ANCOVA will depend not only on contributions of the teacher but also on the peer contributions of the student groups that are likely to be assigned to the teacher. Even though the estimates remove the contributions of the prior inputs as measured by $\boldsymbol{\beta}_0\bar{\mathbf{X}} + \boldsymbol{\beta}_1\bar{\mathbf{Z}}$, if there is interference, the estimated teacher effects are unbiased estimates of contributions of the teacher and the average peer effects of the classes that are likely to be assigned to the teacher. This combines teacher and contextual inputs, consistent with the Type A effects discussed in the introduction.

A limitation of this development is the focus on teachers being assigned a single group of students rather than multiple classes. The potential outcomes and ANCOVA models could be extended to handle the case of multiple classes. The ANCOVA model could include classroom-level predictors for each of the multiple classes taught by the teacher. Including such variables might allow for separating peer and teacher effects under specific assumptions, and future work could explore the requirements of those assumptions given the interference among students assigned to the same class and teacher.

References

Boardman, A. E., & Murnane, R. J. (1979). Using panel data to improve estimates of the determinants of educational achievement. *Sociology of Education*, *52*, 113–121. https://doi.org/10.2307/2112449

Clotfelter, C. T., Ladd, H. F., & Vigdor, J. L. (2006). Teacher-student matching and the assessment of teacher effectiveness. *Journal of Human Resources*, *41*, 778–820. https://doi.org/10.3368/jhr.XLI.4.778

Culpepper, S., & Aguinis, H. (2011). Using analysis of covariance (ANCOVA) with fallible covariates. *Psychological Methods*, *16*, 166–178. https://doi.org/10.1037/a0023355

Frölich, M. (2004). Programme evaluation with multiple treatments. *Journal of Economic Surveys*, *18*, 181–224. https://doi.org/10.1111/j.0950-0804.2004.00001.x

Goldhaber, D., & Hansen, M. (2010). Using performance on the job to inform teacher tenure decisions. *American Economic Review*, *100*, 250–255. https://doi.org/10.1257/aer.100.2.250

Goldhaber, D., Quince, V., & Theobald, R. (2018). Has it always been this way? Tracing the evolution of teacher quality gaps in U.S. public schools. *American Educational Research Journal*, *55*, 171–201. https://doi.org/10.3102/0002831217733445

Harris, D. N., & Sass, T. R. (2011). Teacher training, teacher quality and student achievement. *Journal of Public Economics*, *95*, 798–812. https://doi.org/10.1016/j.jpubeco.2010.11.009

Holland, P. W. (1986). Statistics and causal inference. *Journal of the American Statistical Association*, *81*, 945–960. https://doi.org/10.2307/2289064

Hong, G., & Raudenbush, S. W. (2005). Effects of kindergarten retention policy on children's cognitive growth in reading and mathematics. *Educational Evaluation and Policy Analysis*, *27*, 205–224. https://doi.org/10.3102/01623737027003205

Hong, G., & Raudenbush, S. W. (2006). Evaluating kindergarten retention policy: A case study of causal inference for multilevel observational data. *Journal of the American Statistical Association*, *101*, 901–910. https://doi.org/10.1198/016214506000000447

Hong, G., & Raudenbush, S. W. (2013). Heterogeneous agents, social interactions, and causal inference. In S. L. Morgan (Ed.), *Handbook of causal analysis for social research* (pp. 331–352). New York, NY: Springer.

Hudgens, M., & Halloran, M. (2008). Toward causal inference with interference. *Journal of the American Statistical Association*, *103*, 832–842. https://doi.org/10.1198/016214508000000292

Isenberg, E., Max, J., Gleason, P., Potamites, L., Santillano, R., Hock, H., & Hansen, M. (2013). *Access to effective teaching for disadvantaged students* (NCEE Report No. 2014-4001). Washington, DC: Institute of Education Sciences.

Lockwood, J. R., & McCaffrey, D. F. (2014). Correcting for test score measurement error in ANCOVA models for estimating treatment effects. *Journal of Educational and Behavioral Statistics*, *39*, 22–52. https://doi.org/10.3102/1076998613509405

Lockwood, J. R., & McCaffrey, D. F. (2019). Impact evaluation using analysis of covariance with error-prone covariates that violate surrogacy. *Evaluation Review.* Advance online publication. https://doi.org/10.1177/0193841X19877969

Lord, F. M. (1980). *Applications of item response theory to practical testing problems.* Hillsdale, NJ: Erlbaum.

McCaffrey, D. F., Griffin, B. A., Almirall, D., Slaughter, M. E., Ramchand, R., & Burgette, L. F. (2013). A tutorial on propensity score estimation for multiple treatments using generalized boosted models. *Statistics in Medicine*, *32*, 3388–3414. https://doi.org/10.1002/sim.5753

Raudenbush, S. W. (2004). What are value-added models estimating and what does this imply for statistical practice? *Journal of Educational and Behavioral Statistics*, *29*, 121–129. https://doi.org/10.3102/10769986029001121

Raudenbush, S. W., & Jean, M. (2012). *How should educators interpret value-added scores?* (Knowledge Brief No. 1). Stanford, CA: Carnegie Foundation for the Advancement of Teaching.

Raudenbush, S. W., & Willms, J. (1995). The estimation of school effects. *Journal of Educational and Behavioral Statistics*, *20*, 307–335. https://doi.org/10.3102/10769986020004307

Rothstein, J. (2009). Student sorting and bias in value-added estimation: Selection on observables and unobservables. *Education Finance and Policy*, *4*, 537–571. https://doi.org/10.1162/edfp.2009.4.4.537

Rothstein, J. (2010). Teacher quality in educational production: Tracking, decay, and student achievement. *Quarterly Journal of Economics*, *125*, 175–214. https://doi.org/10.1162/qjec.2010.125.1.175

Rubin, D. B. (1974). Estimating causal effects of treatments in randomized and nonrandomized studies. *Journal of Educational Psychology*, *66*, 688–701. https://doi.org/10.1037/h0037350

Sass, T. R., Hannaway, J., Xu, Z., Figlio, D. N., & Feng, L. (2012). Value added of teachers in high-poverty schools and lower poverty schools. *Journal of Urban Economics*, *72*, 104–122. https://doi.org/10.1037/e721792011-001

Sass, T. R., Semykina, A., & Harris, D. N. (2014). Value-added models and the measurement of teacher productivity. *Economics of Education Review*, *38*, 9–23. https://doi.org/10.1016/j.econedurev.2013.10.003

Tchetgen Tchetgen, E., & VanderWeele, T. (2012). On causal inference in the presence of interference. *Statistical Methods in Medical Research*, *21*, 55–75. https://doi.org/10.1177/0962280210386779

Todd, P. E., & Wolpin, K. I. (2003). On the specification and estimation of the production function for cognitive achievement. *Economic Journal*, *113*(485), F3–F33. https://doi.org/10.1111/1468-0297.00097

Notes

[1] Achievement test scores are error-prone measures of latent (unobserved) achievement attributes because they are based on only a limited number of test items (Lord, 1980). Hence we follow Lockwood and McCaffrey (2014) and define the ANCOVA model in terms of latent prior achievement $\mathbf{X}_i$ and the observed other background characteristics in $\mathbf{Z}_i$. In practice, many authors fit models using the observed test scores. Lockwood and McCaffrey (2019) discussed the assumptions required for those models to yield consistent estimators of effects. Alternatively, correction for measurement error could be applied in the estimation of the ANCOVA model (Culpepper & Aguinis, 2011; Lockwood & McCaffrey, 2014).

[2] Often the ANCOVA model in Equation 1 is parameterized as

$$Y_i = \mu + \boldsymbol{\beta}_0'\mathbf{X}_i + \boldsymbol{\beta}_1'\mathbf{Z}_i + \boldsymbol{\theta}'\mathbf{H}_i + \epsilon_i, \tag{11}$$

where $\boldsymbol{\theta}$ and $\mathbf{H}_i$ are each a vector of $K-1$ elements corresponding to teachers 1 to $K-1$. The element of $\mathbf{H}_i$, corresponding to a student's assigned teacher, takes a value of 1; all the other elements of $\mathbf{H}_i$ equal zero for students assigned to teachers 1 to $K-1$; and all the elements of $\mathbf{H}_i$ equal -1 for students assigned to teacher K. In this case, the model is parameterized not in terms of the mean outcome for students assigned to each teacher but in terms of the effects $\theta_t = \omega_t - K^{-1}\sum_{k=1}^{K}\omega_k$ for $t = 1, \ldots, K$ defined earlier.