EdWorkingPaper No. 25-1291

Controlling For Measurement Error in Evaluation Models When Treatment Group Assignment is Based on Noisy Measures: Evaluation of an Achievement Gap-Closing Initiative

Robert H. Meyer
University of Chicago

Michael S. Christian
Education Analytics

This paper develops new models to evaluate the effects of interventions and intervention-by-site heterogeneity when treatment group assignment is based on a fallible variable and the outcome of interest is determined in part by the corresponding true control variables (measured without error). The specific application concerns a school report card redesign in which school performance is evaluated based on the achievement growth of students in the bottom quartile of prior achievement. We show using Monte Carlo data that the traditional errors-in-variables estimator (EV) produces severely biased estimates of the gap-closing initiative. We develop an augmented EV estimator (AEV) that addresses this bias and is shown to produce highly accurate estimates in Monte Carlo simulations. We also show how AEV can be implemented using regression calibration (RC). Using state data, we find that there are essentially no differences in the average growth in student achievement between students in and not in the lowest quartile. However, the noise-corrected correlation in school growth estimates for the two groups is high (around 0.8), but not perfect. These findings are important given that most (if not all) state accountability systems prioritize reporting of school performance for multiple student sub-groups, including groups with large gaps in prior student achievement.

VERSION: September 2025

Suggested citation: Meyer, Robert H., and Michael S. Christian. (2025). Controlling For Measurement Error in Evaluation Models When Treatment Group Assignment is Based on Noisy Measures: Evaluation of an Achievement Gap-Closing Initiative. (EdWorkingPaper: 25-1291). Retrieved from Annenberg Institute at Brown University: https://doi.org/10.26300/h5e1-v925

Controlling For Measurement Error in Evaluation Models When Treatment Group Assignment is Based on Noisy Measures: Evaluation of an Achievement Gap-Closing Initiative

Robert H. Meyer, NORC at the University of Chicago
Michael S. Christian, Education Analytics

Author Note

This research was supported by the federal Region 10 Comprehensive Center. We are grateful to a state agency for supporting and providing the data used in this project. Thanks also to Sara Hu, who contributed substantially to the empirical part of the paper. Special thanks to Laura Pinsonneault for her support and guidance. For helpful comments, the authors thank participants at the Association for Education Finance and Policy Conference in 2023 and Society for Research in Educational Effectiveness in 2023. The findings, interpretations, and conclusions expressed in this paper are those of the authors and do not necessarily represent the views of NORC at the University of Chicago or Education Analytics. The authors have no known conflicts of interest. Correspondence concerning this paper should be addressed to Robert H. Meyer, NORC at the University of Chicago. Email: Meyer-Robert@norc.org.

Abstract

This paper develops new models to evaluate the effects of interventions and intervention-by-site heterogeneity when treatment group assignment is based on a fallible variable and the outcome of interest is determined in part by the corresponding true control variables (measured without error). The specific application concerns a school report card redesign in which school performance is evaluated based on the achievement growth of students in the bottom quartile of prior achievement. We show using Monte Carlo data that the traditional errors-in-variables estimator (EV) produces severely biased estimates of the gap-closing initiative. We develop an augmented EV estimator (AEV) that addresses this bias and is shown to produce highly accurate estimates in Monte Carlo simulations. We also show how AEV can be implemented using regression calibration (RC). Using state data, we find that there are essentially no differences in the average growth in student achievement between students in and not in the lowest quartile. However, the noise-corrected correlation in school growth estimates for the two groups is high (around 0.8), but not perfect. These findings are important given that most (if not all) state accountability systems prioritize reporting of school performance for multiple student sub-groups, including groups with large gaps in prior student achievement.

Controlling For Measurement Error in Evaluation Models When Treatment Group Assignment is Based on Noisy Measures: Evaluation of a Gap-Closing Initiative

This paper develops new models to evaluate the effects of interventions and intervention-by-site heterogeneity when treatment group assignment is based on a fallible variable (or set of variables) and the outcome of interest is determined in part by the corresponding true control variables (measured without error). The specific application concerns a school report card redesign intended to stimulate actions to close the gap between low and high-performing students. In the redesigned report card, school performance is evaluated in part based on the achievement growth of students in the lowest quartile of prior achievement in each school. The new model extends standard models of student achievement growth and school performance to produce this measure. Since student achievement is inevitably measured with error, the model implements methods to control for measurement error. The distinctive aspect of the model is that true prior achievement and the achievement quartile indicator are tightly connected yet need to be treated differently since the model specification requires measurement error control be applied to measured prior achievement, but not to the bottom quartile indicator which is based on measured prior achievement. As in the research by Lockwood and McCaffrey (2016, 2019), we demonstrate how the standard errors-in-variables (EV) method (Fuller, 1987; Buonaccorsi, 2010) yields strongly biased parameter estimates in this application. We extend this method, which we refer to as augmented errors in variables (AEV), to address this concern. We also show how AEV can be implemented using regression calibration (RC), a method for controlling for measurement error that is widely used in biostatistics and epidemiology research (Carroll et al, 2006; Spiegelman, McDermott, and Rosner, 1997).

Interest in using student growth models to measure school performance has grown steadily since the enactment of the No Child Left Behind Act in 2001/2002, which required states to assess students in math and English language arts in grades 3 to 8 and in high school and to measure and report school performance using this data. Since passage of the Every Student Succeeds Act (ESSA) in 2015, all states have adopted some type of growth model (Data Quality Campaign, 2019; Mills, 2025). Meyer et al (2009), Meyer and Dokumaci (2015), Sanders and Harris (1994), and Betenbenner (2009) document early implementations of state growth models, including value-added and student growth percentile (SGP) models. Wisconsin Department of Public Instruction (2024) provides a detailed explanation of their current state growth model. Growth models generally share the same basic features: they measure the difference between student achievement (or some other student outcome) and predicted student achievement based on one or more prior measures of student achievement and perhaps other control variables (such as student demographic variables).

School growth models share some of the same features and challenges as teacher growth models (Koedel, Mihaly, and Rockoff, 2015; Harris, 2011; Kane, Kerr & Pianta, 2014; Chetty, Friedman & Rockoff, 2014; Guarino, Reckase, and Wooldridge, 2015) but, as argued by Lockwood and McCaffrey (2019), they are less susceptible to the bias caused by potential nonrandom assignment of students to schools (as opposed to teachers). Angrist et al. (2016, 2017, 2024) find encouraging evidence about the validity of school value-added.

We focus in this paper on the potential bias in school value-added models to the assignment of students based on observed measures rather than on unknown (omitted) variables. Indeed, ESSA requires states to measure and report school performance by student subgroups as well as for all students. Our research considers the challenge of measuring growth when subgroups are defined based on measured prior achievement rather than on variables assumed to be measured without error. We address this issue in the context of statistical models that control for measurement error in prior achievement, but our results are applicable to most growth and evaluation models where assignment to a group, program, or treatment is based on control variables measured with error.

Two of the most used methods of measurement error control (MEC) are errors-in-variables (EV) (Fuller, 1987; Buonaccorsi, 2010) and regression calibration (RC) (Carroll et al, 2006; Spiegelman, McDermott, and Rosner, 1997).[1] Both methods require external information on the magnitude of measurement error. EV is commonly used in social science research and is incorporated into many statistical packages. EV corrects for measurement error by correcting the cross-product matrix of the left- and right-hand-side variables. The cross-product matrix is corrected by subtracting the part of the matrix due to measurement error, given external estimates of the variance (and covariance, if applicable). EV has been used in value-added models of student achievement (Meyer et al, 2009; McCaffrey et al, 2004), in models to correct for measurement error in SEL measures (Loeb et al, 2019; Fricke et al, 2021) and in models of student attendance (Meyer, 2025). RC is widely used in biostatistics and epidemiology research and is easy to implement. RC corrects for measurement error by replacing fallible variables with estimates of the true values of these variables. The estimates are derived from a calibration model that predicts the true values of the fallible variables given all variables (other than the dependent variable).

The two methods yield equivalent results in linear regression models where measurement error is additive (classical measurement error) and where the observed fallible variable is a surrogate measure of the true variable (measured without error) (Buonaccorsi, 2010; Carroll et al, 2006). A fallible variable is a surrogate measure of the true variable in a specified model if it does not add predictive power to the model that includes the true variable.[2] Lockwood and McCaffrey (2019), building on Lockwood and McCaffrey (2016), consider how lack of surrogacy results in bias if measurement error is corrected using EV. Their analysis is motivated by a model of teacher value-added where students may be non-randomly assigned to teachers; that is, assigned to teachers partly based on measured prior achievement, a violation of the

[1] See, also, the following: Bennett et al (2017), Freedman et al (2008), Gleser (1992), Guo (2011), and Keogh (2014).

[2] The stronger definition of surrogacy is that the fallible measure contains no information on the distribution of the dependent variable beyond that contained in the true variable and other control variables included in the model. A model and fallible variable that satisfies surrogacy also implies that measurement error is nondifferential (Buonaccorsi, 2010; Carroll et al, 2006; Buzas et al, 2004).

surrogacy assumption. Below, we consider a similar application, where assignment to the gap-closing treatment group in each school is partly based on prior achievement.

The gap-closing initiative has three parts:

1. Students in each school are identified as being in the gap closing target group if they scored in the lowest quartile by school, on the previous state assessment.
2. Schools and districts select and implement policies and practices to improve the achievement of these students.
3. The state department of education publishes in the subsequent school report card data on the growth in student achievement in both the target and non-target groups.

We develop a model and associated estimation method to measure the differential performance of schools and the state overall with respect to achievement growth of the target (bottom quartile) group and non-target (upper quartile) group. As discussed in Lockwood and McCaffrey (2019), models of this type may have other important applications, including, for example: (1) Evaluation of assignment to (selection into) 8th grade Algebra versus regular math based on 7th grade math score, (2) Evaluation of assignment to elementary school reading group or gifted and talented programs given state and/or interim test scores, and (3) Evaluation of a health intervention where assignment to the intervention is based on fallible diagnostic measures rather than random assignment.

In the remainder of the paper, we:

- Present the gap-closing model
- Develop the new method – augmented errors-in-variables (AEV) – for correcting for measurement error. The method addresses the need to correct for measurement error for a specified set of control variables measured with error, as in a standard errors-in-variables model, but not correct for measurement in a second specified set of variables that are correlated with measurement error. This section reviews the standard errors-in-variables (EV) method (Fuller, 1987; Buonaccorsi, 2010).
- Show how to apply the AEV method to the gap-closing model.
- Evaluate the performance of the AEV method using Monte Carlo data.
- Present estimates of the gap-closing model using state data.
- Summarize the major findings of the paper and address possible next steps.

Gap-closing model with correction for measurement error

We first present a simplified gap-closing model with a focus on the overall gap parameter (δ) for a state or district but later extend the model to allow school effects and growth gaps to differ by school. For simplicity, the simplified model allows for only a single pretest variable measured with error (z), the variable that determines assignment to the bottom quartile of prior achievement (I), and a vector of demographic control variables (assumed to be not measured with error). We later extend this model to allow for multiple control variables measured with error.

True score model

$$y = z^{*}\lambda + I\delta + X\beta + e^{*} \tag{1}$$

Model based on fallible variables

$$z = z^{*} + v \tag{2}$$

$$\begin{aligned} y &= z\lambda + I\delta + X\beta + e \\ e &= e^{*} - v\lambda \end{aligned} \tag{3}$$

where:

- y = end-of-year student achievement (for student i in school k, with subscripts suppressed unless required)
- z^{*} = end-of-year student achievement in the prior year (single lag) in the same subject, measured without error
- $z = z^{*} + v$ = measured prior achievement
- v = classical (additive) measurement error in prior achievement, assumed to be homoscedastic. Information on the measurement error properties of v (and all pretest variables) is provided externally by assessment vendors. As is the case with all test scores, a “gold standard" measure of the true score is not available.
- $I = I\{z \leq c\}$ = a zero/one indicator of student prior achievement in the bottom quartile, where c = the cut point for achievement indicated in the bottom quartile[3]
- X = vector of student demographic variables and an intercept (unless expressly excluded if school effects are added)
- e^{*} = random student error, without pretest measurement error)
- $e = e^{*} - v\lambda$ = random student error, with pretest measurement error
- λ, δ, β = conformable parameters to be estimated

The distinctive aspect of the gap closing model is that true prior achievement z^{*} and the achievement quartile indicator I are tightly connected yet need to be treated differently since the model specification requires measurement error control be applied to measured prior achievement z, but not to the bottom quartile indicator which is based on measured prior achievement. We demonstrate below how the EV method yields strongly biased parameter estimates. We extend this method, which we refer to as augmented errors in variables (AEV), to address this concern, and show the connections of the new approach to regression calibration.

Note that although we focus on estimating bottom quartile effects – the overall average effect and separate effects for each school – we are also interested in how these effects change over time, before and after implementation of the gap-closing initiative. Our interest in the change in the gap-closing parameter is driven in part by two concerns. One, the gap-closing coefficient estimated prior to the initiative may be positive because of programs and policies

[3] The extended model allows for different cut points for each school.

targeted at low achievers that existed prior to the initiative. Two, the true relationship between post and prior achievement is not guaranteed to be linear. As in a before-and-after evaluation design, the change in the gap-closing parameter is a better measure of the impact of the initiative than the estimate based on a single year. This paper focuses only on the model based on outcome data from the 2021-22 school year, a point in time very early in the gap-closing initiative.

Methodology

We develop and compare models using the following matrix notation:

$$\begin{aligned} y &= F^{*}\eta + e^{*} \\ F &= F^{*} + V \end{aligned} \tag{4}$$

where F^{*} = matrix of variables included in the model, which includes both variables subject to measurement error and those that are measured without error, V = measurement error (equal to zero for variables not measured with error), and F = measured predictor variables. In the gap-closing model (as described above), where z is the single variable measured with error:[4]

$$\begin{aligned} F^{*} &= \begin{bmatrix} z^{*} & I & X \end{bmatrix} \\ F &= \begin{bmatrix} z & I & X \end{bmatrix} \\ V &= \begin{bmatrix} v & 0 & 0 \end{bmatrix} \\ \eta' &= \begin{bmatrix} \lambda & \delta & \beta \end{bmatrix} \end{aligned} \tag{5}$$

A key attribute of the gap-closing model is that v and I (and, more generally, V and F^{*}) are correlated. We show below that the standard EV estimator is biased and develop an alternative estimator that addresses this bias.

The standard EV estimator (Fuller, 1987; Buonaccorsi, 2010) adjusts the moments of the observed data, y and F, to correct for measurement error, given external information on the variance (and potentially, covariance) of measurement error. We consider the case where measurement error is confined to some or all the model predictors F and measurement error in the dependent variable is random and is absorbed by the equation error e^{*}. To simplify notation, we allow parameter symbols to represent either the true or estimated parameter, depending on the context. The OLS estimator of the model based on variables measured without error is given by:

[4] More generally, we could let $F = \begin{bmatrix} F_1 & F_2 & F_3 \end{bmatrix}$ and $F^{*} = \begin{bmatrix} F_1^{*} & F_2 & F_3 \end{bmatrix}$ where F_1 = the set of variables measured with error and where there is a need to correct for measurement error, F_2 = the set of variables that are correlated with measurement error but there is no need to correct for error, and F_3 = variables not measured with error.

$$\eta^* = L^{*-1} F^{*\prime} y$$
$$= [F^{*\prime} F^*]^{-1} F^{*\prime} y \tag{6}$$

where $L^* = [F^{*\prime} F^*]$.

The standard EV estimator is given by:

$$\eta_{EV} = L_{EV}^{-1} F'y \tag{7}$$

where the cross-product matrix L_{EV} is corrected for measurement error given external information on the variance-covariance of measurement error S_{VV} for *F*:

$$L_{EV} = F'F - S_{VV}$$
$$S_{VV} = E[V'V] \tag{8}$$

$F'y$ is not corrected for measurement error given the apparently reasonable assumption that the error in the posttest *y* and *F* are uncorrelated (but see below). The EV estimator assumes that $E[L_{EV}] = E[F'F - S_{VV}] = F^{*\prime} F^*$. We show below that this assumption is unwarranted in our application.

Define a provisional error correction estimator as:

$$\eta = L^{-1} R \tag{9}$$

where the left-hand side and right-hand side terms (L and *R*) are error corrected. Let

$$ec(M^*) = M - \{E[M] - E[M^*]\} \tag{10}$$

represent an error-correction function that corrects the observed matrix *M* by subtracting the term in brackets, the expected error in *M* as an estimate of the true matrix *M**. The EV estimator assumes that $ec(F^{*\prime} F^*) = F'F - \{E[F'F] - F^{*\prime} F^*\} = F'F - S_{VV}$. In fact, the error-corrected matrix equals:

$$L = ev(F^{*\prime} F^*) = F'F - \left\{E[F'F] - E[F^{*\prime} F^*]\right\}$$
$$= F'F - E\left[(F^* + V)'(F^* + V) - F^{*\prime} F^*\right] \tag{11}$$
$$= F'F - S_{VV} - S_{*V} - S_{V*}$$

where

$$S_{*V} = E[F^{*\prime} V]$$
$$S_{V*} = E[V'F^*] \tag{12}$$

The matrices L_{EV} and L differ if components of V and F^* are correlated, a key feature of the gap-closing model. In this case, S_{*V} and S_{V*} do not equal zero. Similarly, the right-hand side term R is given by:

$$
\begin{aligned}
R &= ev(F^{*\prime} y) = F'y - \left\{ E[F'y] - E[F^{*\prime} y] \right\} \\
&= F'y - E\left\{ [F - F^*]' y \right\} \\
&= F'y - E[V'y]
\end{aligned}
\tag{13}
$$

EV assumes that the term $E[V'y]$ equals zero but in the gap-closing model measurement error v (in matrix V) causes assignment to intervention I (causing y) and thus $E[V'y]$ does not equal zero; measurement error is differential. The matrix R can further be written as:

$$
\begin{aligned}
R &= F'y - \left\{ E[F'(F^*\eta + e^*)] - E[F^{*\prime}(F^*\eta + e^*)] \right\} \\
&= F'y - \left\{ D^* - L^* \right\} \eta
\end{aligned}
\tag{14}
$$

where $D^* = E[F'F^*]$ and we exploit the fact that the error component e^* is uncorrelated with F and F^*. Since F^* is not observed, we replace D^* with the corresponding error-corrected term:

$$
\begin{aligned}
D &= ec(D^*) = ec(F'F^*) = F'F - \left\{ E[F'F] - E[F'F^*] \right\} \\
&= F'F - \left\{ E[(F^* + V)'F^* + V)] - E[(F^* + V)'F^*] \right\} \\
&= F'F - S_{VV} - S_{*V}
\end{aligned}
\tag{15}
$$

which yields:

$$
\begin{aligned}
R &= ev(F^{*\prime} y) \\
&= F'y - [D - L]\eta
\end{aligned}
\tag{16}
$$

As indicated in (13), $R \neq F'y$ (and equivalently, $D \neq L$) as required for the EV estimator.

The provisional error-correction estimator is thus given by:

$$
\begin{aligned}
\eta &= L^{-1} R \\
&= L^{-1}[F'y - (D - L)\eta]
\end{aligned}
\tag{17}
$$

The problem with this formula is that the parameter to be estimated (η) appears on the right side of the solution which means that we need to know the parameter to estimate it. The solution to this problem is to augment the EV approach by solving for the parameter, which yields the augmented EV (AEV) estimator:

$$
\begin{aligned}
\eta &= L^{-1} F'y - L^{-1}(D - L)\eta \\
\eta_{AEV} &= D^{-1} F'y = ev(F'F^*)^{-1} F'y
\end{aligned}
\tag{18}
$$

The AEV estimator thus differs from the traditional EV estimator in that it is necessary to construct the measurement error-corrected term $D = ev(F'F^*)$ rather than the measurement error-corrected terms L_{EV} or L.

There are two options for constructing $D = ev(F'F^*) = F'F - S_{VV} - S_{*V}$. The first option extends EV by subtracting $S_{VV} + S_{*V}$ from the cross-product matrix rather than only S_{VV}. There is no need to correct the right-hand side term. It is straightforward to compute S_{VV} and S_{*V} if there exists a gold standard sample that either represents the same population as in the study sample or can be matched to represent the study sample. In this case the two correction terms can be calculated. One of the advantages of this option is that EV and augmented EV can handle heteroscedastic errors (Fuller, 1987). However, this option is challenging to implement in our application since there is no obvious gold standard measure of student achievement and thus no gold standard sample.[5]

The second option is to estimate D after conditioning on the observed vector F. We develop and implement this option and show how this option is like regression calibration. Conditioning D on the measured variables F yields the following simplified equation:

$$\begin{aligned}\tilde{D} &= ev(F'F^* \mid F) = E[F'F^* \mid F] = F'E[F^* \mid F] \\ &= F'\tilde{F}\end{aligned} \tag{19}$$

where $\tilde{F}$ equals the measurement error corrected measure of F, where the regression of F^* on F (with coefficient matrix γ) must be constructed since F^* is not observed (given no gold standard sample).

$$\begin{aligned}\tilde{F} &\equiv E[F^* \mid F] \equiv F\gamma \\ \gamma &= (F'F)^{-1}F'F^*\end{aligned} \tag{20}$$

Interestingly, $\tilde{F}$ is identical to the RC estimate of F^* and thus the assumptions that justify standard regression calibration also apply here; namely, that measurement error is homoscedastic. We show below how to construct this matrix in the context of the gap-closing model. The AEV estimator is given by:

$$\eta_{AEV} = \tilde{D}^{-1}F'y = (F'\tilde{F})^{-1}F'y \tag{21}$$

The AEV estimator is interesting in several respects. First, it takes the form of an instrumental variables (IV) estimator with F as the vector of instrumental variables, but with the need to construct the student-level error-corrected vector $\tilde{F}$ for each student, used in place of the unobserved vector F^*. A valuable feature of the AEV estimator is that there is no need to construct or modify the right-side term $F'y$ since both F and y are observed variables. Since the AEV estimator takes the form of an IV estimator, approximate standard errors of AEV estimates

[5] On the other hand, it may be possible to simulate a gold standard sample using item response theory, an exercise we have explored in previous research.

can be obtained using the standard IV formula, but with constructed terms substituted in place of computed terms:

$$Var(\eta_{AEV}) \approx ev(F'F^*)^{-1}(F'F)ev(F'F^*)^{-1} = \sigma^2(F'\tilde{F})^{-1}(F'F)(\tilde{F}'F)^{-1} \tag{22}$$

where σ^2 = the variance of the residual. More accurate estimates can be obtained using the bootstrap, but the precision estimates obtained are likely to be highly accurate in large samples (such as those based on statewide data).[6]

Second, the formula for the AEV estimator and the IV formula in (22), as demonstrated in Appendix A, are numerically equivalent to the formulas for regression calibration (RC) and associated least squares variance. This result contradicts the standard conclusion (Carroll et al, 2006, pa. 72; Buonaccorsi, 2010, p. 116) that the EV and RC estimators are equivalent in linear models with external information on measurement error properties. The derivation of the AEV estimator is useful in that it clarifies the conditions under which the EV estimator fails to produce consistent estimates.

Third, a useful property of the AEV estimator is that the AEV model provides parameter estimates that are identical to those provided by the EV estimator when the augmented features of the AEV estimator are not required. This is a convenient result in that the data constructed to implement the AEV estimator can be used to estimate both models that require AEV estimation and models that could be appropriately estimated using EV estimation.

Application of AEV to the Gap-Closing Model

In this section we consider application of the AEV method to the gap-closing model. We first consider a model with a single pretest variable *z*, which is used to define the bottom quartile target indicator *I*, and a vector of variables not measured with error *X*. Appendix B extends the model to include differential school effects for target and non-target group students and extends the model to include additional pretest variables measured with error. Recall, as in (5), that $F^* = \begin{bmatrix} z^* & I & X \end{bmatrix}$ and $F = \begin{bmatrix} z & I & X \end{bmatrix}$. Thus, *D* - *L* equals:

$$\begin{aligned} D - L &= ec(F'F^{*\prime}) - ec(F^{*\prime}F^{*\prime}) \\ &= ev[(F - F^*)'F^*] \\ &= E\left\{\begin{bmatrix} v & 0 & 0 \end{bmatrix}' \begin{bmatrix} z^* & I & W \end{bmatrix}\right\} \\ &= \begin{bmatrix} 0 & E[v'I] & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \end{aligned} \tag{23}$$

[6] As indicated in the next paragraph and Appendix A, the approximate precision of AEV estimates can also be computed using the OLS formula for the variance-covariance matrix where the predictors are the regression calibrated estimates if the true predictors.

which does not equal the zero matrix, as required for EV. This term does not equal zero because the bottom quartile target indicator I is a function of the measured pretest $z = z^* + v$ and thus is correlated with measurement error v. In this model, measurement error v is effectively a causal determinant of achievement.

As indicated above we construct the error-corrected variable $\tilde{F}$ in our application using regression calibration. Since F^* contains a single variable measured with error (z), we need to construct the parameters of the $z^* \mid F = z^* \mid \begin{bmatrix} z & I & X \end{bmatrix}$ regression using external information on the variance of error v, but not information from a gold standard sample. However, although the target indicator I is a proper conditioning variable, it is excluded from the regression since it has no predictive power given that measured prior achievement z is included in the model and given the assumption that measurement error v is homoscedastic.[7] Below, we derive the parameters of the error correction regression of $z^*|z, X$, given by:

$$z^* \mid z, I, X = z^* \mid z, X = z\gamma_1 + X\gamma_2 + r = \tilde{z} + r \tag{24}$$

where $\tilde{z} = z\gamma_1 + X\gamma_2$ equals the calibrated prediction from the regression of the unknown pretest score z^*.

To compute the parameters of the error correction regression, we estimate an auxiliary regression $z|X$, which captures the relationship between measured prior achievement z and X, and external information on the variance of measurement error v, given by σ_v^2. The error variance is typically provided by assessment vendor. We assume that it is measured without error. The auxiliary regression $z|X$ is given by:

$$\begin{aligned} z^* \mid X &= X\pi + u_1 \\ z \mid X &= X\pi + u_2 \\ u_2 &= u_1 + v \end{aligned} \tag{25}$$

The variance of u_1 (σ_1^2) is unknown but is readily computed given the variance of the observed residual u_2 (σ_2^2) and the externally provided variance of pretest measurement error $v(\sigma_v^2)$:

$$\sigma_1^2 = \sigma_2^2 - \sigma_v^2 \tag{26}$$

We obtain formulas for the parameters γ_1 and γ_2 by applying the Frisch-Waugh-Lovell (FWL) theorem. Given FWL, γ_1 (the coefficient of z^* on z, controlling for X) is given by the regression of the residual of the $z^*|X$ regression (u_1) on the residual from the $z|X$ regression (u_2), which equals:

---

[7] The requirement that student-level measurement error is homoscedastic (within schools) is mirrored in the literature on standard regression calibration (Spiegelman, McDermott, and Rosner, 1997; Guo, 2011; Spiegelman, Logan, and Grove, 2011); Sarkar, Mallick, and Carroll, 2014). We show in Appendix B that AEV accommodates heteroscedasticity across schools.

$$\gamma_1 = \sigma_1^2 / (\sigma_1^2 + \sigma_v^2) = \sigma_1^2 / \sigma_2^2 < 1 \tag{27}$$

This parameter is equal to the reliability of pretest *z* given control for *X*. The coefficient γ_2 (the coefficient of z^* on *X*, controlling for *z*) is given by:

$$\gamma_2 = \pi(1-\gamma_1) \tag{28}$$

The residual *r* is equal to:

$$r = (1-\gamma_1)u_1 - \gamma_1 v = u_1 - \gamma_1 u_2 \tag{29}$$

and its variance is given by:

$$\sigma_r^2 = \gamma_1 \sigma_v^2 < \sigma_v^2 \tag{30}$$

Given estimates of the parameters of the auxiliary regression, the error-corrected pretest variable $\tilde{z}$ is given by:

$$\tilde{z} = E[z^* \mid z, X] = z\gamma_1 + X\gamma_2 = z\gamma_1 + X\pi(1-\gamma_1) \tag{31}$$

To implement the AEV estimator of the gap-closing model with measurement error correction, the constructed student-level variable $\tilde{z}$ is used in place of the unobserved pretest variable *z* and the model is estimated using the method of instrumental variables, as indicated in (21). As demonstrated in Appendix A, the model could equivalently be estimated using OLS regression, with the error-corrected variable $\tilde{z}$ used in place of *z.* Appendix B expands the gap-closing model to incorporate differential school effects for low and high achievers and additional variables measured with error. Appendix C presents precision formulas for the extended gap-closing model.

In the next two sections we report on the results of a Monte Carlo evaluation of the simplified model and results of the simplified and extended models using state data.

Monte Carlo Evaluation of the AEV Estimator of the Simplified Model

Table 1 reports the results of a Monte Carlo evaluation of the EV and AEV estimators of the simplified model that includes a pretest variable measured with error *z*, a bottom quartile indicator *I*, a control variable *X*, correlated with the true pretest but not with pretest measurement error, and a constant term.[8] The specified coefficients used to simulate the data on the regressors are: Constant = 0, Pretest = 1, Bottom quartile = 0.25, and Control variable *X* = 1. The simulated test scores and demographic variable have been standardized to have mean zero and standard

[8] We report standard errors obtained from the two-stage least squares estimator in this study rather than boot strap standard errors because standard errors are exceedingly small given the very large size of the simulated (N = 10,000) and actual state (N = 50,869) data sets. Identical numbers were obtained using the OLS formula based on the model with regression calibrated predictors. For example, the z-statistic for the pretest coefficient in the full model with school effects equals 171.4 (see Table 3). Boot strap errors can be calculated by re-estimating the model using multiple data sets constructed by sampling with replacement from the from the full data set (Buonaccorsi, 2010). Use of the boot strap is recommended when samples are small and when the model includes many control variables measured with error.

deviation equal to one. The simulated data set includes 10,000 observations, so all coefficients are estimated with high precision.

Our focus in this section is on assessing the bias in the EV estimator and assessing the degree to which the AEV estimator yields consistent estimates of the true parameters. OLS estimates using the data with and without pretest measurement error are also reported. The specified coefficients used to create the simulated data are, as expected, recovered with almost perfect accuracy in the model of data without measurement error. We compare all other estimates to these "True OLS" estimates. In the OLS model with mismeasured prior achievement, the estimated coefficients on prior achievement and X are biased, as expected. An interesting result is that the coefficient on the bottom quartile indicator, 0.273, is very close to the true parameter value of 0.25. We have verified that this result is consistent with theory given the simple structure of the simplified model, as explained in Appendix A. Since we are ultimately focused on estimating separate school effects for students in and not in the bottom quartile it is necessary to use an estimator that provides consistent estimates of all parameters.

As predicted above, the EV estimator produces biased parameter estimates, and the magnitude of the bias is larger than for OLS. The coefficient for the bottom quartile indicator, 1.150 versus a true value equal to 0.25, is extremely large, more than four times larger than the true parameter. In contrast, the AEV estimator produces estimates for all parameters that are very close to the true values. This analysis confirms the importance of controlling for measurement using an estimator that properly takes account of the endogeneity of the treatment group assignment when assignment is based on noisy measures.

Table 1. Monte Carlo Simulation Results

Variable	**True OLS**	**OLS**	**Errors-in-Variables (EV)**	**Augmented Errors-in-Variables (AEV)**
Constant	**0.002**	**0.004**	**-0.220**	**0.003**
	(0.009)	(0.011)	(0.014)	(0.011)
Pretest z	**1.011**	**0.799**	**1.327**	**1.008**
	(0.010)	(0.012)	(0.023)	(0.015)
Bottom Quartile I	**0.259**	**0.273**	**1.150**	**0.273**
	(0.022)	(0.028)	(0.044)	(0.028)
Control Variable X	**1.001**	**1.152**	**0.961**	**1.008**
	(0.012)	(0.013)	(0.016)	(0.014)
Note: Standard errors are in parentheses. Monte Carlo simulation results are based on sample size 10,000.				

Estimates of the Gap-Closing Model Using State Data

Tables 2 and 3 report estimates of the simplified and extended models (the latter with school effects) for English Language Arts in 8th grade using the EV and AEV estimators. As indicated in Table 2, the EV and AEV estimators yield identical parameter estimates for a model that excludes the bottom quartile variable. This confirms the statement above that the two estimators yield identical estimates when the AEV model is not needed to address inclusion of an endogenous variable, correlated with measurement error. However, both models are misspecified because they exclude the bottom quartile variable. In the Monte Carlo simulation, exclusion of the bottom quartile indicator had a large effect on parameter estimates because the coefficient on the indicator was large. The key result, consistent with the results based on simulated data, is that the EV and AEV estimators of the full model yield very different estimates of the bottom quartile effect. The AEV estimate is 0.0282, whereas the EV estimate is 0.3783, an effect size estimate that is unreasonably large. Given the theoretical analysis and Monte Carlo evidence presented above, we conclude that the EV estimate is severely biased.

Table 3 reports estimate of the preferred model that includes school effects based on the AEV estimator. Column (1) repeats estimates of the model without school effects from Table 2. Column (2) adds school effects, not interacted with the indicators of students in and not in the bottom quartile, and column (3) interacts school effects with the two group indicators. The latter two models also add controls for student demographic characteristics. The two new models yield estimates of the average bottom quartile effect as well as school effects. The latter model also allows comparison of the school effects for students in and not in the bottom quartile, respectively. As indicated in Table 3, the addition of school effects and demographic characteristics to the model does not much affect the parameter estimates. The coefficient on prior achievement is somewhat lower in the expanded models, as expected, since school indicators are somewhat correlated with prior achievement due to sorting of students across schools. Again, the estimated bottom quartile effects are very close to zero in both models.

Table 2. Estimates of the Simplified Gap-Closing Model with State Data: English Language Arts, 8th grade, 2021-22

Variable	**EV1**	**AEV1**	**EV2**	**AEV2**
Constant	**0.0000**	**0.0000**	**-0.0960**	**-0.0072**
	(0.0027)	(0.0027)	(0.0038)	(0.0035)
Pretest	**0.8965**	**0.8965**	**1.0247**	**0.9061**
	(0.0031)	(0.0030)	(0.0051)	(0.0042)
Bottom Quartile Effect			**0.3783**	**0.0282**
			(0.0103)	(0.0086)
Demographic Characteristics				
School Effects				
Bottom Quartile Indicator * School Effects				
Adjusted R-squared	0.71	0.63	0.73	0.63
Number of observations	50520			

Note: Standard errors are in parentheses.

Table 3. Estimates of the Extended Gap-Closing Model with State Data: English Language Arts, 8th grade, 2021-22

Variable	**AEV2**	**AEV3**	**AEV4**
Constant	**-0.0072**		
	(0.0035)		
Pretest	**0.9061**	**0.8488**	**0.8497**
	(0.0042)	(0.0049)	(0.0049)
Bottom Quartile Effect	**0.0282**	**-0.0159**	**-0.0137**
	(0.0086)	(0.0087)	(0.0088)
Demographic Characteristics	No	Yes	Yes
School Effects	No	Yes	No
Bottom Quartile Indicator * School Effects	No	No	Yes
Adjusted R-squared	0.63	0.68	0.69
Number of observations	50520		

Note: Standard errors are in parentheses.

Since this study was motivated by the need to report separate school effects for students in and not in the bottom quartile it is important to know whether these effects differ. Table 4 reports the raw and noise-corrected correlations of these two sets of estimates. All correlations are weighted by school size. The average reliability of the estimates is also reported: 0.65 for the bottom quartile school effects and 0.85 for school effects of students not in the bottom quartile. The higher reliability for the latter group is expected given that the sample size is exactly three times larger than for the bottom quartile group. Given the inevitable noise in estimating school effects, it is not surprising that the raw correlation of the two effects is relatively low: 0.61. The noise-corrected correlation is much higher, however: 0.82. This suggests that there were uniformly modest differences in the effectiveness of schools with their lower and high-achieving students, but this could be due in part to the fact that the 2021-22 school year placed much greater emphasis on reestablishing preferred instructional practices following the pandemic-related disruptions of the past two years than on responding to the gap-closing initiative.

Estimates of the noise-corrected variance of the school effects can (and likely should) be used to produce bivariate shrinkage estimates of the school effects. As in the case of standard growth models (without separate effects by group), shrinkage estimates optimally trade off estimation noise and accuracy in estimating true effects. For the model considered in this paper, given the relatively high noise-corrected correlation, the bivariate shrinkage estimates of the bottom quartile and upper quartile school effects are likely to be very similar except for schools where the raw effects are quite different and where the precision of these estimates is high.

Table 4. Raw and Noise-Corrected Correlation of School Effects for Students in and Not in the Bottom Quartile

	Bottom Quartile Students	**Students Not in the Bottom Quartile**
Noise Variance	0.0182	0.0057
Noise-corrected Variance	0.0341	0.0310
Total Raw Variance	0.0522	0.0367
Reliability	0.6520	0.8451
Sample Size	12,824	37,696
Raw correlation	0.6101	
Covariance	0.0267	
Noise-corrected correlation	0.8195	

Summary and Conclusions

This paper shows that the traditional EV estimator produces severely biased estimates of the gap-closing model. We present an augmented EV estimator (AEV) that addresses this bias and is shown to produce highly accurate estimates in Monte Carlo simulations. This finding is important given that most (if not all) state accountability systems prioritize reporting of school performance for multiple student sub-groups, including groups with large gaps in prior student achievement. More generally, the proposed method addresses the evaluation challenge of estimating treatment effects when treatment group assignment is based on noisy measures. Bivariate (or multivariate) shrinkage estimation is recommended to optimally control for estimation error when there are many sites (each with limited sample size).

Measurement error-corrected estimates of the gap-closing model based on data from a single state and prior to ramping up of the gap-closing initiative, indicate that there are essentially no differences in the average growth in student achievement for students in and not in the bottom quartile. However, the noise-corrected correlation in school growth estimates for the two groups is high (80%) but not perfect. Over time it is important to learn if growth of the two groups of students, overall and by school, diverges in response to reforms and interventions stimulated by the state gap-closing initiative.

We show that there are two options for implementing AEV. Both options result in corrections to the cross-product matrix $D = ev(F'F^*) = F'F - S_{VV} - S_{*V}$. The first option extends EV by subtracting $S_{VV} + S_{*V}$ from the cross-product matrix rather than only S_{VV}. There is no need to correct the right-hand side term. It is straightforward to compute S_{VV} and S_{*V} if there exists a gold standard sample that represents the same population as in the study sample. In this case the two correction terms can be calculated. However, this option is challenging to implement in our application since there is no obvious gold standard measure of student achievement and thus no gold standard sample. On the other hand, it may be possible to simulate a gold standard sample using item response theory, an approach we have previously explored.

The second option is to estimate *D* after conditioning on the observed variables. This approach requires constructing an error-corrected, regression calibrated, estimate of the true predictor, given external information on the distribution of measurement error. The model is implemented as an instrumental variables estimator. We show in Appendix A that the formula for AEV, implemented using this option, is numerically equivalent to the formula for regression calibration (RC). This result contradicts the standard conclusion (Carroll et al, 2006, pa. 72; Buonaccorsi, 2010, p. 116) that the EV and RC estimators are equivalent in linear models with external information on measurement error properties. The derivation of the AEV estimator is useful in that it clarifies the conditions under which the EV estimator fails to produce consistent estimates. One of the advantages of this option is that EV and augmented EV (estimated using the first option) can readily handle heteroscedastic error (Fuller, 1987) whereas the standard regression calibration application requires homoscedasticity of measurement error. Spiegelman, Logan, Grove (2011) consider how to implement RC given heteroscedastic error provided that a gold standard sample exists. They show that standard regression calibration performs quite well unless measurement error is "severe."

The methods developed in this paper were motivated by an application where assignment to the treatment group (the bottom quartile group) is based on a simple assignment rule: students are assigned to the treatment group if they scored in the bottom quartile within their school of measured prior achievement. In fact, the method can handle more general assignment rules. Firstly, the model could be extended to handle multiple treatments, for example, assignment based on different levels of prior achievement. Secondly, assignment rules might be more complicated. Assignment could be based on multiple characteristics, including student demographic characteristics, multiple lags of prior achievement, and random error. This rule is readily handled by the proposed method given that the multiple variables (but not the random error) are also included in the primary model and the calibration regression.[9] Indeed, if measurement error is additive, assignment can be based on nonlinear functions of multiple variables. Note that the assignment mechanism need not be reported, although it could be empirically studied, but, as in all evaluation studies, assignment (or the intent of assignment) must be known.

It is theoretically possible that a state could be interested in the outcomes of students in the true bottom quartile, although a treatment variable would then not exist since students could only be identified probabilistically. In this case, the model would need to be generalized to address measurement error in both prior achievement and a proxy treatment variable. Measurement error control methods would then be applied to both variables.

The bottom quartile indicator introduces a simple form of nonlinearity into the model of student achievement and school performance considered in this paper. One interesting and

[9] It is quite possible that assignment to treatments designed to support low achieving students could be explicitly or implicitly based on multiple lags of prior achievement. In this case, the model discussed in Appendix B that includes multiple predictors measured with error would be the preferred model.

potentially important model extension would be to allow the primary regression to be nonlinear, based on a spline or polynomial model. In this case, the gap-closing effect would be estimated as a shift in student growth on top of the underlying relationship between student achievement and student predictors.

References

Angrist, J., Hull, P., Pathak, P., & Walters, C. (2016). Interpreting tests of school VAM validity. *American Economic Review, 106*(5), 388-392. https://doi.org/10.1257/aer.p20161080

Angrist, J., Hull, P., Pathak, P., & Walters, C. (2017). Leveraging lotteries for school value-added: Testing and estimation. *Quarterly Journal of Economics, 132*(2), 871-919. https://doi.org/10.1093/qje/qjx001

Angrist, J., Hull, P., Pathak, P., & Walters, C. (2024). Credible school value-added with undersubscribed school lotteries. *Review of Economics and Statistics*, *106*(1), 1-19. https://doi.org/10.1162/rest_a_01149

Bennett, DA; Landry, D; Little, J; Minelli, C. (2017). "Systematic review of statistical approaches to quantify, or correct for, measurement error in a continuous exposure in nutritional epidemiology," *BMC Medical Research Methodology,* 17: 146.

Buonaccorsi, J.P. (2010). *Measurement Error: Models, Methods, and Applications*, Chapman Hall/CRC, Boca Raton, FL.

Buzas, J.S.; Stefanski, L.A.; Tosteson, T.D. (2004). Measurement Error, In W. Ahrens & L. Pigeot (Eds.), *Handbook of Epidemiology*, London: Springer.

Carroll, RJ; Ruppert, D; Stefanski, LA; Crainuceanu, CM. (2006). *Measurement Error in Nonlinear Models: A Modern Perspective,* Vol. 2, Chapman Hall/CRC, Boca Raton, FL.

Chetty, R.; Friedman, J.N.; Rockoff, J.E. (2014). Measuring the Impacts of Teachers I: Evaluating Bias in Teacher Value-Added Estimates, *American Economic Review*, Vol. 104, No. 9, pp. 2593-2632.

Data Quality Campaign (2019). *Growth Data: It Matters, and It's Complicated*, Report.

Freedman, LS; Midthune, D; Carroll, RJ; Kipnis, V. (2008). "A comparison of regression calibration, moment reconstruction and imputation for adjusting for covariate measurement error in regression," *Statistics in Medicine*, 27(25), 4195-5216.

Fuller, WA. (1987). *Measurement Error Models,* John Wiley & Sons, Hoboken, NJ.

Gleser, L. (1992). The importance of assessing measurement reliability in multivariate regression, *Journal of the American Statistical Association*, 87, 696-707.

Guarino, C.M.; Reckase, M.D.; Wooldridge, J.M. (2015). Can Value-Added Measures of Teacher Performance be Trusted? *Education Finance and Policy*, Vol. 10, No. 1, pp. 117-156.

Guo, Y; Little, RJ. (2011). Regression analysis with covariates that have heteroscedastic measurement error, *Statistics in Medicine*, 30: 2278-2294.

Harris, D.N. (2011). *Value-Added Measures in Education*, Cambridge: Harvard University Press.

Kane, T.J.; Kerr, K.A.; Pianta, R.C. (2014). *Designing Teacher Evaluation Systems*, San Francisco: Jossey-Bass.

Keogh, RH; White, IR. (2014). A toolkit for measurement error correction, with a focus on nutritional epidemiology, *Statistics in Medicine*, 33, 2137-2155.

Koedel, C., Mihaly, K., Rockoff, J.E. (2015). Value-added modeling: A Review, *Economics of Education Review*, Vol. 47, pp. 180-195.

Lockwood, J.R., and McCaffrey, D.F. (2014). Correcting for Test Score Measurement Error in ANCOVA Models for Estimating Treatment Effects, *Journal of Educational and Behavioral Statistics*, Vol. 39, No. 1, pp. 22-52.

Lockwood, J.R.; McCaffrey, D.F. (2016). Matching and Weighting With Functions of Error-Prone Covariates for Causal Inference, *Journal of the American Statistical Association*, Vol. 111, pp. 1831-1839.

Lockwood, J.R., and McCaffrey, D.F. (2019). Impact Evaluation Using Analysis of Covariance With Error-Prone Covariates That Violate Surrogacy, *Evaluation Review*, Vol. 43, No. 6, pp. 335-369.

Longford, N.T. (2005). *Missing Data and Small-Area Estimation: Modern Analytical Equipment for the Survey Statistician*, New York, NY: Springer-Verlag.

Lovell, M. *(2008). "A Simple Proof of the FWL Theorem". Journal of Economic Education. **39** (1): 88–91.*

McCaffrey, D., Lockwood, J., Koretz, D., Louis, T., & Hamilton, L. (2004), Models for value-added modeling of teacher effects, J*ournal of Educational and Behavioral Statistics, Vol. 29, No. 1, 67.*

Meyer, R.H. (2025). Social Emotional Learning, Student Attendance, and Chronic Absenteeism in Pre- and Post-Pandemic Periods, Working Paper, NORC at the University of Chicago.

Meyer, R.H.; Dokumaci, E.; Geraghty, E.L.; McLaughlin, S.; Morgan, E.A. (2009). *Wisconsin Value-Added Model: A Demonstration Project Final Report*, Madison, I: Wisconsin Center for Education Research.

Meyer, R.H; Dokumaci, E. (2015). Value-Added Models and the Next Generation of Assessments, in Lissitz, Robert W. (ed.), *Value added modeling and growth modeling with particular application to teacher and school effectiveness*, Charlotte, NC: Information Age Publishing, 2015.

Mills, A. (2025). Student growth models: the 50-state landscape. Policy Research in Missouri Education, 7(6). Saint Louis University. www.primecenter.org/policy-briefs-database/growth-models @SLU_PRIME prime@slu.edu @SLU_PRIME

Pane, J.F.; Griffin, B.A.; McCaffrey, D.F.; Karam, R. (2014). Effectiveness of Cognitive Tutor Algebra 1 at Scale, *Educational Evaluation and Policy Analysis*, Vol. 36., No. 2, pp. 127-144.

Rao, J.N.K. (2003). *Small Area Estimation*. New York, NY: Wiley-Interscience.

Sanders, W.; Horn, S. (1994). The Tennessee Value-Added Assessment System (TVAAS): Mixed-model methodology in educational assessment, *Journal of Personnel Evaluation in Education*, Vol. 8, No. 3, pp. 299-311.

Spiegelman, D; McDermott, A; Rosner, B (1997). Regression calibration for correcting measurement error-bias in nutritional epidemiology, *American Journal of Clinical Nutrition*, 65 (suppl):1179S-1186S.

Spiegelman, D; Logan, R.; Grove, D. (2011). Regression Calibration with Heteroscedastic Error Variance, *International Journal of Biostatistics*, 7:1.

Wisconsin Department of Public Instruction (2024). Technical Report on the Wisconsin Value-Added Model: Academic Year 2023-24, State Report prepared by Education Analytics. WI_DPI_School_VA_Technical_Report_2024_Final.pdf

Appendix A. Augmented Errors-in-Variables, Regression Calibration, and Ordinary Least Squares

We show in this paper that the EV and AEV estimators yield different estimates for models, such as the gap-closing model, where the models include variables correlated with error but there is no need to control for measurement error in these variables. The analyses using Monte Carlo generated and actual state data showed that the bias in the EV estimator could be quite large. In this appendix we show that the formulas for the AEV estimator and regression calibration are numerically equivalent. This result contradicts the standard conclusion (Carroll et al, 2006, pa. 72; Buonaccorsi, 2010, p. 116) that the EV and RC estimators are equivalent in linear models with external information on measurement error properties. We also explain why the OLS estimator yields an unbiased estimate of the bottom quartile effect, but biased estimates of all other parameters in the simplified model. This result does not carry over to the extended model with school effects since estimation of school effects requires consistent estimates of all coefficients on control variables.

Repeating equations (21) and (20) for convenience, the AEV estimator derived in the paper is given by:

$$\eta_{AEV} = \tilde{D}^{-1}F'y = (F'\tilde{F})^{-1}F'y \tag{21}$$

where the error-corrected matrix $\tilde{F}$ is given by:

$$\begin{aligned} &\tilde{F} \equiv E[F^* \mid F] \equiv F\gamma \\ &\gamma = (F'F)^{-1}F'F^* \end{aligned} \tag{20}$$

The RC estimator, defined in terms of the error-corrected matrix $\tilde{F}$, is given by:

$$\eta_{RC} = (\tilde{F}'\tilde{F})^{-1}\tilde{F}'y \tag{32}$$

Substituting for $\tilde{F}$ in the RC formula yields:

$$\begin{aligned} \eta_{RC} &= \{[\gamma'F'][F\gamma]\}^{-1}[\gamma'F']y \\ &= \left\{[F^{*\prime}F(F'F)^{-1}F'][F(F'F)^{-1}F'F^*]\right\}^{-1}[F^{*\prime}F(F'F)^{-1}F']y \\ &= \left\{F^{*\prime}F(F'F)^{-1}F'F^*\right\}^{-1}F^{*\prime}F(F'F)^{-1}F'y \end{aligned} \tag{33}$$

Since the cross-product matrices are invertible, this equation can be written in the much simpler form as:

$$\begin{aligned} \eta_{RC} &= (F'F^*)^{-1}(F'F)(F^{*\prime}F)^{-1}F^{*\prime}F(F'F)^{-1}F'y \\ &= (F'F^*)^{-1}F'y \end{aligned} \tag{34}$$

where, since F^* is unobserved it is replaced by $\tilde{F}$, which yields the AEV estimator. It can also be shown that the formulas for the variance-covariance of the AEV estimator, using the IV formula in (22), and the RC estimator, using the standard OLS formula, are equivalent. Although we recommend precision of AEV and RC-based estimates be calculated using the bootstrap, the precision estimates obtained using the formula may be useful for exploratory research.

The OLS estimator (without correcting for measurement error) is given by:

$$\eta_{OLS} = (F'F)^{-1} F'y \tag{35}$$

Substituting for $y = F^*\eta + e^*$ and taking expectations given F and F^* yields:

$$\begin{aligned} \eta_{OLS} &= (F'F)^{-1} F'F^*\eta \\ &= (F'F)^{-1} F'(F - V)\eta \\ &= \eta - B\eta \end{aligned} \tag{36}$$

where the bias multiplier B is given by:

$$B = (F'F)^{-1} F'V) \tag{37}$$

In the context of the simplified gap-closing model, $B\eta$ is given by:

$$B\eta = (F'F)^{-1} F'\begin{bmatrix} v & 0 & 0 \end{bmatrix}\eta = \begin{bmatrix} (F'F)^{-1} F'v & 0 & 0 \end{bmatrix}\begin{bmatrix} \lambda \\ \delta \\ \beta \end{bmatrix} \tag{38}$$

The first term in the right-side matrix equals the coefficient of the regression of $v = z - z^*$ on $F = \begin{bmatrix} z & I & X \end{bmatrix}$ or, equivalently, the difference in the regressions of z on F and z^* on F.

Using (24), the regression equation for z^* on F, and given that z predicts z, we obtain:

$$E[(F'F)^{-1} F'v] = \left\{ \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} - \begin{bmatrix} \gamma_1 \\ 0 \\ \gamma_2 \end{bmatrix} \right\} = \left\{ \begin{matrix} 1-\gamma_1 \\ 0 \\ \gamma_2 \end{matrix} \right\} \tag{39}$$

and

$$E[B]\eta = \begin{bmatrix} 1-\gamma_1 & 0 & 0 \\ 0 & 0 & 0 \\ -\gamma_2 & 0 & 0 \end{bmatrix}\begin{bmatrix} \lambda \\ \delta \\ \beta \end{bmatrix} = \begin{bmatrix} (1-\gamma_1)\lambda \\ 0 \\ -\gamma_2\lambda \end{bmatrix} \tag{40}$$

This confirms the empirical result reported in the text that the OLS estimator yields an unbiased estimate of the bottom quartile effect, but biased estimates of all other parameters in the simplified model. As indicated in the text this result (and equation (24) is based on the assumption that measurement error is homoscedastic within schools. This result does not carry over to the extended model with school effects since estimation of school effects requires consistent estimates of all coefficients on control variables.

Appendix B. Extensions of the Gap-Closing Model

Extension of the Gap-Closing Model to Incorporate Differential School Effects for Low and High Achievers

In this section we extend the gap-closing model to incorporate separate school effects for students in the target group versus students in the non-target group. We write the model in two different but equivalent forms. Form 1, the structural model, writes the model to highlight the separate school effects for the respective low (α_1) and high (α_2) achieving target groups. These models allow for separate school effects by interacting the target group indicator *I* with the vector of school indicators *S*. These effects are centered around the average effects in each group. Form 1 includes an intercept (ζ) but allows for separate target and non-target group effects (ξ_1, ξ_2), normalized to have mean zero on average.[10] The average growth between the two groups is equal to $\delta = \xi_1 - \xi_2$, the same parameter as in the simplified model.

Form 1: Structural model: Centered target/non-target group and school effects

$$y = z^*\lambda + I\xi_1 + (1-I)\xi_2 + X\beta + \zeta + S_1\alpha_1 + S_2\alpha_2 + e^* \quad (41)$$

The vector of school variables for the target and non-target students are given by:

- S_1=diag$(I)S$ and S_2=diag$(1-I)S$, where $S_{1ik} = I_{ik}S_{ik}$ and $S_{2ik} = (1-I_{ik})S_{ik}$.

The primary advantage of Form 1 is that it defines multilevel target and non-group effects and school effects for target and non-target group students that are policy relevant. These parameters can be combined to yield school effects for target and non-target group students that capture differences in the growth of students in the two groups on average and differences across schools. This model specification also provides the framework for applying shrinkage estimation to create optimal estimates of school effects for the two groups, as discussed in Appendix C. The combined school effect parameters are given by:

$$\begin{aligned} \eta_{1k} &= \xi_1 + \alpha_{1k} \\ \eta_{2k} &= \xi_2 + \alpha_{2k} \end{aligned} \quad (42)$$

---

[10] Including an intercept and separate target and non-target coefficients allows the combined school effects for the target and non-target groups to be centered around zero, a common normalization in growth models. To impose this normalization, we require: $P_1\xi_1 + (1-P_1)\xi_2 = 0$; that is, the weighted average of the target and non-target group effects is zero where the weight P_1 = 0.25 = the proportion of students in the bottom quartile. Note that the state allows this proportion to exceed 0.25 (but not be greater than 0.50) for small schools so that the number of students in the bottom quartile is at greater than or equal to 20.

where the school index *k* has been added to clearly distinguish the multilevel parameters. The interacted school effects could in principle be treated as fixed or random effects, but it is straightforward to apply the AEV estimator if the school effects are treated as fixed effects, so we follow this approach in this paper.

Model form 2 is designed to allow straightforward estimation of fixed school effects. The model drops the intercept and group effects variable *I*. These parameters are absorbed into the school effects for the target and non-target group students. Structural parameters are recovered in a second step, as indicated in the Appendix C.

Form 2: Model with fixed school effects and no intercept or target group effect

$$y = z^{*}\lambda + S_1 a_1 + S_2 a_2 + X\beta + e^{*} \tag{43}$$

where:

- $a_1 = \zeta + \xi_1 + \alpha_1$
- $a_2 = \zeta + \xi_2 + \alpha_2$
- X now excludes the intercept

Application of the AEV method to estimate Form 2 of the extended gap-closing model is quite like that discussed above for the simplified model. However, we build the model by considering separate equations for each school and then combine them to create a multilevel model that includes all students. As in (24), the equation to construct the measurement error corrected variable $\tilde{z}_{ik}$ for student *i* in school *k*, is given by:

$$z^{*}_{ik} \mid F_{ik} = z_{ik}\gamma_{1k} + X_{ik}\gamma_{2k} + I_{ik}S_{ik}g_{1k} + (1-I_{ik})S_{ik}g_{2k} + r_{ik} \tag{44}$$

As argued above, although the target indicators *I* and (1 – I) are proper conditioning variables, they can be excluded from the regression since they have no predictive power given that measured prior achievement *z* is included in the model given the assumption that measurement error *v* is homoscedastic. They are replaced by a school fixed effect $S_{ik}\gamma_{3k}$. This is equivalent to setting the coefficients on the school interaction variables to be equal: $\gamma_{3k} = g_{1k} = g_{2k}$. This yields the following measurement error correction equation:

$$z^{*}_{ik} = z_{ik}\gamma_{1k} + X_{ik}\gamma_{2k} + S_{ik}\gamma_{3k} + r_{ik} = \tilde{z}_{ik} + r_{ik} \tag{45}$$

As in (25), the auxiliary regression of $z_{ik} \mid X_{ik}, S_{ik}$ for school *k* is given by:

$$\begin{aligned} z^{*}_{ik} \mid X_{ik}, S_{ik} &= X_{ik}\pi_{1k} + S_{ik}\pi_{2k} + u_{1ik} \\ z_{ik} \mid X_{ik}, S_{ik} &= X_{ik}\pi_{1k} + S_{ik}\pi_{2k} + u_{2ik} \\ u_{2ik} &= u_{1ik} + v_{ik} \end{aligned} \tag{46}$$

Given the inclusion in the model of the school effect π_{2k} it is feasible and useful from an efficiency standpoint to restrict the slope parameter to be the same across schools: $\pi_{1k} = \pi_{1k}$. School-specific estimates of π_k are inevitably noisy since they are based on the number of students in each school, not on the entire sample.[11]

As indicated in (45), the coefficients used to construct the measurement error corrected variable $\tilde{z}_{ik}$ are allowed to vary across schools *k*. The formulas in the extended model, analogous to equations (26) - (31), are given by:

$$\begin{aligned} \sigma_{2k}^2 &= \sigma_{1k}^2 + \sigma_{vk}^2 \\ \gamma_{1k} &= \sigma_{1k}^2 / (\sigma_{1k}^2 + \sigma_{vk}^2) = \sigma_{1k}^2 / \sigma_{2k}^2 < 1 \\ \gamma_{2k} &= \pi_2 (1 - \gamma_{1k}) \\ \gamma_{3k} &= \pi_3 (1 - \gamma_{1k}) \end{aligned} \tag{47}$$

Note, however, that it is reasonable and likely advantageous from an efficiency standpoint to continue to assume that $\sigma_1^2 = Var(u_1)$ is constant (homoscedastic). An estimate of this variance parameter is given by $\sigma_1^2 = \bar{\sigma}_2^2 - \bar{\sigma}_v^2$, where $\bar{\sigma}_2^2 = Var(u_2)$ is now interpreted as the average variance of u_2 and $\bar{\sigma}_v^2 = \sum_i n_k \sigma_{vk}^2 / N$ is the average weighted variance of measurement error over schools and students. The results presented in the text adopt this approach.

One very important generalization of the extended model with school effects (or any multilevel model) is that it is no longer necessary to restrict the model to cases where measurement error is homoscedastic across all students. Measurement error variances in multilevel models can be allowed to vary across schools (*k*), as indicated in (47), but not across students within schools.

Extension of the Gap-Closing Model to Incorporate Additional Variables Measured with Error

Since it is common to include multiple pretest variables in student growth models, either multiple lags in the same subject as the dependent variable or test scores in other subject areas, it is important to extend the gap-closing model to incorporate additional variables measured with error and show how this affects application of the AEV estimation method. The extended model takes the same forms as the two models considered above, but with the modification that the pretest variable *z* now represents multiple pretests, given by:

[11] The practical importance of this restriction can in principle be evaluated using a multilevel model of z_{ik} that allows for variable (random) slopes as well as random or fixed school effects.

Pretest measurement model

$$
\begin{aligned}
z &= z^* + v \\
z &= \begin{bmatrix} z_1 & \dots & z_Q \end{bmatrix} \\
z^* &= \begin{bmatrix} z_1^* & \dots & z_Q^* \end{bmatrix} \\
v &= \begin{bmatrix} v_1 & \dots & v_Q \end{bmatrix}
\end{aligned}
\tag{48}
$$

where z, z^*, v represent measured prior achievement, true prior achievement, and pretest measurement error, respectively, for *Q* pretest variables. Assignment to the gap-closing target indicator continues to be based on a single pretest variable, given by z_1. To simplify the notation, we combine variables *X* and *S* into the variable $W = \begin{bmatrix} X & S \end{bmatrix}$ and write the equation for $z^* \mid W$ as:

$$
z^* = z\gamma_1 + W\gamma_2 + r = \tilde{z} + r \tag{49}
$$

To implement the AEV estimator it is necessary to construct the matrix of predictions of $\tilde{z} = E[z^* \mid z, W]$ given the matrix of measured pretests *z* and *W*. As above, since z^* is not observed we must construct the coefficients of the error correction equations (49), where these equations now represents a system of equations, one for each pretest variable, γ_1 is now a (*QxQ*) matrix, and γ_2 is a (*MxQ*) vector where M = dimension (*W)*. The auxiliary equation of (25) now represents the regression of the matrix *z* given *W,* with the matrix of coefficients equal to π, a (*MxQ*) matrix.

As above, the variance matrix of measurement errors Σ_{vvkk} (replacing σ_{vk}^2 in (47)) is assumed to be constant across students (homoscedastic), but may be heteroscedastic across schools. The variance matrix of vector u_1, given by Ω (replacing σ_1^2) is assumed to be constant across students and is given by the difference between the average variance-covariance of u_2 and the average variance of measurement error, $\Omega = \bar{V} - \bar{\Sigma}_{vv}$, where:

$$
\begin{aligned}
\bar{V} &= u_2' u_2 / N \\
\bar{\Sigma}_{vv} &= \sum_i n_k \Sigma_{vvk} / N \\
\Omega &= \bar{V} - \bar{\Sigma}_{vv} \\
V_k &= \Omega + \bar{\Sigma}_{vvk}
\end{aligned}
\tag{50}
$$

and where the matrices are of dimension (*QxQ*). Application of the FWL Theorem yields the following solutions for the (*QxQ)* matrices γ_{1k} and γ_{2k}, which generalizes (47) to the case of multiple pretests.

$$
\begin{aligned}
V_k &= \Omega + \Sigma_{vvk} \\
\gamma_{1k} &= V_k^{-1}\Omega = (\Omega + \Sigma_{vvk})^{-1}\Omega \\
\gamma_{2k} &= \pi(I - \gamma_{1k})
\end{aligned} \tag{51}
$$

Note that if pretest measurement error is homoscedastic across schools (and students), γ_1, γ_2 take on the same values for all schools and students. Gleser (1992) and Buonaccorsi (2010) refer to (51) as the reliability matrix.

Values of the constructed error corrected variables $\tilde{z}_{qi}$ for pretest variable *q* for student *i* can be substituted into (49) to create the required error-corrected variables. Finally, the AEV estimator is implemented using the instrumental variables estimator of equation (21), where *F* represents the matrix of measured variables and $\tilde{F}$ represents the matrix of error corrected variables, including variables not measured with error. As indicated in Appendix A, the AEV estimator can also be implemented using regression calibration.

Appendix C. Precision Formulas for the Extended Gap-Closing Model

In this appendix we derive the formulas to compute the estimates of the parameters and variance-covariance matrix of Form 1, the model that includes structural parameters representing the intercept, target group indicator, and fixed school effects of the extended gap closing model. To eliminate the collinearity that arises between these variables, we estimate parameters using Form 2. We show how to recover estimates of the preferred (policy relevant) structural parameter from the reduced form parameter estimates of Form 2. The equations for Forms 1 and 2 are reproduced below for convenience.

Form 1: Structural model: Centered target/non-target group and school effects

$$
y = z^*\lambda + I\xi_1 + (1-I)\xi_2 + X\beta + \zeta + S_1\alpha_1 + S_2\alpha_2 + e^* \tag{41}
$$

Form 2: Model with fixed school effects and no intercept or target group effect

$$
y = z^*\lambda + S_1 a_1 + S_2 a_2 + X\beta + e^* \tag{43}
$$

In matrix notation, the reduced form parameters of Form 2 equal:

$$
\begin{aligned}
a_1 &= 1_K(\zeta + \xi_1) + \alpha_1 \\
a_2 &= 1_K(\zeta + \xi_2) + \alpha_2
\end{aligned} \tag{52}
$$

where 1_K = a ($Kx1$) vector of ones and K = the number of schools. Estimates of the reduced form parameters contain estimation errors w_1 and w_2 and are given by:

$$
\begin{aligned}
\hat{a}_1 &= 1_K(\zeta + \xi_1) + \alpha_1 + w_1 \\
\hat{a}_2 &= 1_K(\zeta + \xi_2) + \alpha_2 + w_2
\end{aligned} \tag{53}
$$

We obtain estimates of the structural parameters in the following steps.

Step 1: Compute weighted estimates of the average of the school fixed effects for each group.

$$\bar{\hat{a}}_1 = \phi_1{}'\hat{a}_1 = \left(\zeta \hat{+} \xi_1\right)$$
$$\bar{\hat{a}}_2 = \phi_2{}'\hat{a}_2 = \left(\zeta \hat{+} \xi_2\right) \tag{54}$$

where ϕ_1, ϕ_2 are weight vectors that each sum to one, the caret (^) sign indicates a parameter estimate; and the bar sign indicates an average. The average school effects by group can be interpreted as the implicit estimated intercepts for each group, of the terms $\left(\zeta + \xi_1\right)$and $\left(\zeta + \xi_2\right)$, respectively. We consider two weighting options:

1. Weighting by the number of students in each group by school: $\phi_1 = n_1 / N_1$ and $\phi_2 = n_2 / N_2$, where n_1 and n_2 represent vectors of student enrollment by school for each group, respectively, and N_1 and N_2 represent the total number of students in each group, respectively. If the slope coefficients in the model are estimated with high precision (as is the case in our study), *n* weighting is equivalent to weighting by the precision of the school by group estimates.
2. GLS weighting, where ϕ_1 and ϕ_2 are based on the inverse of the variance of the error in (52) that includes the errors due to estimation error, w_1 and w_2, and the true components α_1 and α_2. This approach effectively treats the school by group effects as random effects.

The two options often yield similar results and do so in our study.

Step 2: Center the estimated school effects for each group around the computed average.

$$\hat{\alpha}_1 = \hat{a}_1 - \bar{\hat{a}}_1$$
$$\hat{\alpha}_2 = \hat{a}_2 - \bar{\hat{a}}_2 \tag{55}$$

Step 3. Compute the overall model intercept ζ , distinct from the mean school effects for each group, ξ_1, ξ_2 . This parameter is defined given the normalization that the N-weighted average of the mean school effects for each group equals zero: $P_1\xi_1 + (1 - P_1)\xi_2 = 0$, where $P_1 = 0.25$, the proportion of students in the target group.

$$\hat{\zeta} = P_1\bar{\hat{a}}_1 + (1 - P_1)\bar{\hat{a}}_2 \tag{56}$$

Step 4. Compute the mean school effects for each group.

$$\hat{\xi}_1 = \bar{\hat{a}}_1 - \hat{\zeta} = (1 - P_1)(\bar{\hat{a}}_1 - \bar{\hat{a}}_2)$$
$$\hat{\xi}_2 = \bar{\hat{a}}_2 - \hat{\zeta} = -P_1(\bar{\hat{a}}_1 + \bar{\hat{a}}_2) \tag{57}$$

Step 5. Compute the target group effect δ

$$\hat{\delta} = \hat{\xi}_1 - \hat{\xi}_2 = (\bar{\hat{a}}_1 - \bar{\hat{a}}_2) \tag{58}$$

Estimates of parameters *a_1* and *a_2* and the associated variance-covariance matrix for these parameters, call this Σ, are all that is needed to estimate the structural parameters $\alpha_1, \alpha_2, \zeta, \xi_1, \xi_2, \delta$ and their associated variance-covariance matrices. The matrix Σ is a subset of the variance-covariance matrix from the complete Form 2 model and can be extracted from the estimates of that model. Note that Σ captures all sources of uncertainty in the estimation of parameters, including estimation error due to imprecision in estimates of the slope parameters λ and β. Below, we present the matrix formulas for computing estimates of all parameters and their associated variances-covariances of estimation errors. We define these matrices as follows (with matrix dimensions listed on the right side of each parameter).

$$\hat{a} \equiv \begin{bmatrix} \hat{a}_1 \\ \hat{a}_2 \end{bmatrix} \quad \text{(2K x 1)}$$

$$\bar{\hat{a}} \equiv \begin{bmatrix} \bar{\hat{a}}_1 \\ \bar{\hat{a}}_2 \end{bmatrix} = A\hat{a} \quad \text{(2 x 1)} \tag{59}$$

$$\text{where A} = \begin{bmatrix} \phi_1' & 0 \\ 0 & \phi_2' \end{bmatrix} \quad \text{(2 x 2K)}$$

$$\hat{\alpha} \equiv \begin{bmatrix} \hat{\alpha}_1 \\ \hat{\alpha}_2 \end{bmatrix} = \hat{a} - B\bar{\hat{a}} = \hat{a} - BA\hat{a} = C\hat{a} \quad \text{(2K x 1)}$$

$$\text{where } B = \begin{bmatrix} 1_K & 0 \\ 0 & 1_K \end{bmatrix} \quad \text{(2K x 2)} \tag{60}$$

$$\text{and } C = I_{2K} - BA \quad \text{(2K x 2K)}$$

$$\hat{\zeta} = P\bar{\hat{a}} = PA\hat{a} \quad \text{(1 x 1)}$$

$$\text{where} \quad P = \begin{bmatrix} P_1 & 1 - P_1 \end{bmatrix} \quad \text{(1 x 2)} \tag{61}$$

$$\hat{\xi}_1 = (1 - P_1)D^- \bar{\hat{a}} = (1 - P_1)D^- A\hat{a} \quad \text{(1 x 1)}$$

$$\hat{\xi}_2 = -P_1 D^+ \bar{\hat{a}} = -P_1 D^+ A\hat{a} \quad \text{(1 x 1)}$$

$$\hat{\delta} = \hat{\xi}_1 - \hat{\xi}_2 = D^- \bar{\hat{a}} = D^- A\hat{a} \quad \text{(1 x 1)} \tag{62}$$

$$\text{where} \quad D^- = \begin{bmatrix} 1 & -1 \end{bmatrix} \quad \text{(1 x 2)}$$

$$\text{and } D^+ = \begin{bmatrix} 1 & 1 \end{bmatrix} \quad \text{(1 x 2)}$$

The variance-covariance matrices of estimation error are given by:

$$
\begin{aligned}
&\Sigma_{\alpha\alpha} = Var[\hat{\alpha} \mid \alpha] = C\Sigma C' \quad \text{(2K x 2K)} \\
&\sigma_{\zeta}^2 = Var[\hat{\zeta}\,[\zeta]] = PA\Sigma A'P' \quad \text{(1 x 1)} \\
&\sigma_{\xi_1}^2 = Var[\hat{\xi}_1 \mid \xi_1] = (1-P_1)^2 D^- A\Sigma A' D^{-\prime} \quad \text{(1 x 1)} \\
&\sigma_{\xi_2}^2 = Var[\hat{\xi}_2 \mid \xi_2] = P_1^2 D^+ A\Sigma A' D^{+\prime} \quad \text{(1 x 1)} \\
&\sigma_{\delta}^2 = Var[\hat{\delta} \mid \delta] = D^- A\Sigma A' D^{-\prime} \quad \text{(1 x 1)}
\end{aligned}
\qquad (63)
$$

Given estimates of α_1, α_2, it is essential to compute bivariate shrinkage estimates of the two parameter vectors since both are typically estimated with error. Table 4 reports that the reliability of the school effects for the two achievement groups are 0.69 and 0.85, respectively. Shrinkage estimation optimally borrows information from the school effects estimated with for the high achievement group to produce estimates of the school effects for the low achievement group that have higher precision. Shrinkage estimates are computed given the weighted variance-covariance of the true (noise-corrected) estimates of α_1, α_2 and the estimated (not noise-corrected) estimates of these parameters as discussed in Longford (2005) and Rao (2003).